LEUPHANA
UNIVERSITY OF LÜNEBURG

Social Actor or Technology? Experimental Studies on the Perception of Chatbots Versus Humans and Their Implications for Anthropomorphic Chatbot Design

Von der Fakultät Management und Technologie
der Leuphana Universität Lüneburg zur Erlangung des Grades

Doktor der Wirtschafts-, Sozial- und Politikwissenschaften
– Dr. rer. pol. –

genehmigte Dissertation von

Lennart Seitz

geboren am 13.05.1992 in Essen

Eingereicht am: 19.12.2023

Mündliche Verteidigung (Disputation) am: 27.05.2024

Erstbetreuerin: Prof. Dr. Sigrid Bekmeier-Feuerhahn, Leuphana Universität Lüneburg
Erstgutachterin: Prof. Dr. Sigrid Bekmeier-Feuerhahn, Leuphana Universität Lüneburg
Zweitgutachter: Prof. Dr. David D. Loschelder, Leuphana Universität Lüneburg
Drittgutachter: Prof. Dr. Claas C. Germelmann, Universität Bayreuth

Die einzelnen Beiträge des kumulativen Dissertationsvorhabens sind oder werden ggf. inkl. des Rahmenpapiers wie folgt veröffentlicht:

Seitz, L., Bekmeier-Feuerhahn, S., & Gohil, K. (2022). Can we trust a chatbot like a physician? A qualitative study on understanding the emergence of trust toward diagnostic chatbots. *International Journal of Human-Computer Studies*, *165*, 102848. https://doi.org/10.1016/j.ijhcs.2022.102848

Seitz, L. (2024). Artificial empathy in healthcare chatbots: Does it feel authentic? *Computers in Human Behavior: Artificial Humans*, *2*(1), 100067. https://doi.org/10.1016/j.chbah.2024.100067

Seitz, L., & Bekmeier-Feuerhahn, S. (2023). Bots have to be fast: The detrimental effects of response delays in service chatbots and the moderating role of anthropomorphism. Manuscript submitted for publication.

DANKSAGUNG

In dem Moment, in dem ich diese Worte verfasse, befinde ich mich von der Fertigstellung meiner Dissertation nicht mehr weit entfernt. Ich sitze gerade fast allein auf dem Flur, die Sonne scheint und es ist kurz vor Feierabend – genau die richtige Zeit, den Prozess des eigenen, noch nicht ganz abgeschlossenen Promotionsprozesses zu reflektieren. Ich könnte nun lang und breit mit mir selbst diskutieren, was wohl ein adäquater Einstieg oder überhaupt ein angemessener Inhalt für diese Danksagung ist. Aber ich beginne einfach mit dem ersten Gedanken, der mir durch den Kopf ging.

Als ich im ersten Semester des Bachelors voller Aufregung in einer der Auftaktveranstaltungen saß, zeigte uns ein Professor eine Abbildung mit einer Art "Akademischen Pyramide". Auf der untersten Ebene war der Bachelor abgebildet, dessen angebliches Ziel es sei, bestehendes akademisches Wissen aufzunehmen, zu verstehen und wiedergeben zu können. Das sollte doch zu schaffen sein, dachte ich mir damals. Auf der zweiten Ebene stand dann der Master. Dessen Ziel sei es, akademisches Wissen analysieren, kritisch reflektieren und auf andere Kontexte übertragen zu können. Das schien schon eine wesentlich abstraktere und nur schwer zu bewältigende Herausforderung zu sein – wie soll ich denn analysieren und hinterfragen, was Expert*innen postulieren? Aber die Pyramiden-Spitze sollte dies noch toppen: Ziel des Doktorats sei es, eigenes akademisches Wissen zu generieren und etwas völlig Neues zu erforschen. Für mich erschien dies unmöglich. Schließlich machen das doch nur echte Wissenschaftler*innen. Aber das war ja auch überhaupt nicht das Ziel meines etwa zehn Jahre jüngeren Ichs, das sich damals so sicher war, nach dem Bachelor sofort in die Praxis einsteigen zu wollen.

Nun sitze ich noch immer hier in meinem Büro in der Uni. Der Flur ist noch leerer. Die Sonne geht langsam unter. Und ich stehe am Ende meines Doktorats: Ich habe – so hoffe ich – mit meiner Forschung nun doch ein kleines bisschen mehr Wissen geschaffen und für die Nachwelt in Publikationen, Konferenzbeiträgen und dieser Dissertation festgehalten. Der Weg dorthin

war zwar lang und hin und wieder auch von Selbstzweifel und Frustration gekennzeichnet, im Rückblick aber doch nur mäßig steinig – schließlich hatte ich immer große Freude an Wissenschaft und Forschung und eine stets hohe intrinsische Motivation. So kam es äußerst selten vor, dass ein Rohdatensatz nach fertiger Datenerhebung länger als fünf Minuten unangetastet blieb. Zu groß war die Neugier, ob sich die aufgestellten Hypothesen tatsächlich bestätigen lassen. All das neigt sich jedoch in gefühlt winzig kleinen, aber dennoch rasend schnellen Schritten dem Ende zu. Da weder das Mammutprojekt "Dissertation" noch der Weg dorthin trotz der notwendigen hohen Eigenständigkeit gänzlich isoliert im stillen Kämmerlein beschritten wird, ist es an der Zeit, denjenigen Menschen zu danken, die mir durch ihr mal mehr, mal weniger aktives Zutun direkt oder indirekt zum erfolgreichen Erlangen des höchsten akademischen Grades verholfen haben.

Chronologisch startend, möchte ich weit vor dem Start der Promotion beginnen – denn das Doktorat ist nur die Spitze der "Akademischen Pyramide". An dieser Stelle möchte ich daher als erstes meiner Mutter Katja Seitz danken, die mich durch alle Ebenen der Pyramide begleitet hat. Egal ob durch emotionalen, finanziellen oder administrativen Support – du hast immer hinter mir gestanden, mir den Rücken freigehalten, an mich geglaubt, mich ermutigt und mir auch so gut es ging inhaltlich versucht zu helfen. Unvergessen sind die vielen Wochenenden und Feierabende, wo du deine Freizeit für das Abfragen von ganzen Skripten geopfert hast. Vom Vergleich zwischen Buchführung nach HGB oder IFRS über die Position der Medulla Oblangata im Hirn bis hin zur Theorie der Kognitiven Dissonanz: Auch wenn du es gern bestreitest und meinen Erfolg allein auf mich attribuierst, so hast du doch einen wesentlichen Anteil daran. Mit sich zuspitzender Pyramide wurde jedoch dein Involvement in meine Themen immer geringer und andere Menschen traten ins Rampenlicht. Hier sei als nächstes meine liebe Freundin Luise Kirsten zu nennen, die in der Pyramidenmitte zunächst meine Kumpanin beim Durchforsten diverser Skripte und später dann zu meiner festen Partnerin wurde. Auch dir

möchte ich herzlich dafür danken, dass du mir den nicht zu unterschätzenden Rückhalt im Privaten geleistet hast. Natürlich habe ich auch deinen inhaltlichen Beitrag zu dieser Dissertation nicht vergessen: Die Durchführung einiger Pre-Tests und Studien wären ohne den Zugang zu deinen Marktforschungspanels sicherlich wesentlich zeit- und kostenintensiver geworden. Sich weiter Richtung Spitze und der eigentlichen Promotion hinbewegend möchte ich an erster Stelle meiner Betreuerin Sigrid Bekmeier-Feuerhahn danken. Noch bevor ich mich definitiv zu einer Promotion entschieden habe, eröffnete sie mir die Möglichkeit, an diesem hochaktuellen Forschungsprojekt im Bereich der Mensch-Chatbot-Interaktion mitzuwirken und im Rahmen dessen meine Dissertation anzufertigen. Tatsächlich bin ich mir nicht sicher, ob ich ohne diese Chance überhaupt eine Promotion begonnen hätte oder ob es mich nicht doch, wie eigentlich schon nach dem Bachelor anvisiert, in die Praxis verschlagen hätte. Darüber hinaus möchte ich dir ganz besonders für dein hohes Vertrauen seit Tag eins danken und die damit verbundene Freiheit, die du mir in Forschung und Lehre eingeräumt hast. Wer mich kennt, der weiß um meine Selbstständigkeit und mein starkes Bedürfnis, meinen Ideen und meiner Kreativität freien Lauf lassen zu können. Gedankt sei dir auch für die schöne Zeit am Lehrstuhl, deinen uneingeschränkten Einsatz für mich und die interessanten Diskussionen sowohl im Rahmen als auch abseits der Forschung und Wissenschaft.

Natürlich gibt es viele weitere Menschen, die in dieser Danksagung einen Platz verdient hätten. Damit dieser Text jedoch nicht länger wird als die eigentliche Dissertation, beschränke ich mich auf ein paar ausgewählte Personen. Insbesondere möchte ich hier auch meinem Zweitbetreuer David Loschelder danken, der nicht nur durch seine fachliche und methodische Kompetenz glänzt, sondern der auch die gefühlt einzige Person ist, die mich im Kartfahren schlagen kann. Danke auch für deine Ermutigung hin zur Promotion und die Betreuung meiner Masterarbeit – irgendwann wird auch diese bestimmt publiziert. Ein großer Dank gilt auch Claas Christian Germelmann, der bereitwillig die Rolle des externen Drittgutachters übernommen hat.

Weiterhin möchte ich Vera Barther danken, die über alles und jeden an der Leuphana Universität Bescheid weiß und auf wirklich jede Frage eine Antwort hat. Ein weiterer Dank geht an Cornelius Neuring, der in Zeiten der ewigen Corona-Lockdowns zeitweise mein einziger sozialer Kontakt am Campus war und mit dem jede Mittagspause zwischen Paper und Studie auszuufern drohte. Ich erinnere mich gern mit einem Schmunzeln an unsere Lunch-Dates, bei denen wir uns das Mensaessen aus Pappschachteln im Besprechungsraum haben schmecken lassen. Zu guter Letzt möchte ich den studentischen Hilfskräften danken, die mich tatkräftig bei der Recherche, der Erstellung von Stimuli für die Studien und vielen anderen Tätigkeiten unterstützt haben. Namentlich sind dies Anne Diedrich, Julia Woronkow, Nico Schwarz, Mia Witte und Mareike Üffing.

Danke!

Lennart

TABLE OF CONTENTS

LIST OF FIGURES

PART I: Framework Paper

PART II: Paper 1

PART III: Paper 2

PART IV: Paper 3

LIST OF TABLES

PART I: Framework Paper

PART II: Paper 1

PART III: Paper 2

LIST OF ABBREVIATIONS[1]

PART I: Framework Paper

AGI ..... Artificial general intelligence
AI ..... Artificial intelligence
CASA ..... Computers are Social Actors
EVT ..... Expectancy Violations Theory
NLP ..... Natural language processing

PART II: Paper 1

ABT ..... Affect-based trust
CA ..... Conversational agent
CBT ..... Cognition-based trust
UVM ..... Uncanny Valley of Mind

PART III: Paper 2

BE ..... Behavioral-empathetic
CC ..... Control condition
EM ..... Empathetic
SY ..... Sympathetic

PART IV: Paper 3

NS ..... Non-social
SD ..... Social delays
SO ..... Social
TAM ..... Technology Acceptance Model
UTAUT ..... Unified Theory of Acceptance and Use of Technology

[1] Abbreviations used across multiple papers are listed only once under the paper where they have been used first.

PART I: FRAMEWORK PAPER

1 Introduction

The emergence of the digital era and new technologies has significantly transformed the dynamics of human communication. This transformation is marked by a shift from traditional human-human communication (e.g., face-to-face) to computer-mediated communication (e.g., chatting via messengers like WhatsApp) and, more notably, human-bot communication (e.g., talking to artificial voice assistants like Amazon's "Alexa"). One of the new technologies that will significantly shape the transition from interpersonal to human-bot communication are so-called "chatbots". Chatbots can be defined as autonomous virtual agents that are designed to communicate with humans in natural language via chat (Araujo, 2018; Mariani et al., 2023). Therefore, they either rely on simple conversation scripts (i.e., users click through a pre-scripted conversation), or they apply artificial intelligence (AI)-based techniques like natural language processing (NLP) and machine learning to analyze and respond to user input (Adamopoulou & Moussiades, 2020). The application areas and purposes of chatbots are manifold ranging from the automation of simple customer services (Crolic et al., 2022) to highly complex tasks like the provision of psychotherapy (Fitzpatrick et al., 2017). Thanks to advancements in AI and NLP, sophisticated chatbots can even mimic a close friend or romantic relationship partner (Pentina et al., 2023; Skjuve et al., 2021) or show signs of an artificial general intelligence (AGI) as became evident by the launch of "ChatGPT" and "GPT-4" by OpenAI (Bubeck et al., 2023).

Even though chatbots have only become prevalent with recent technological advancements, their history dates back over five decades (Shum et al., 2018; Silva & Canedo, 2022). In 1966, the computer scientist Joseph Weizenbaum introduced the software system "ELIZA" that was intended to demonstrate a computer's ability to engage in natural conversations with humans. Therefore, the system employed simple pattern matching and conversation scripts that enabled "ELIZA" to simulate rudimentary aspects of Carl Roger's

client-centered psychotherapy (Weizenbaum, 1966). As Weizenbaum reported in an interview, some people being engaged in intimate conversations with "ELIZA" even asked the experimenter to leave the room (see Goldman, 2017). Other participants, however, believed vehemently that "ELIZA" truly understands their messages and problems. "ELIZA" was thus supposed to be able to pass the so-called "Turing test" that was introduced by the British mathematician, computer scientist, and AI pioneer Alan Turing in 1950 (Turing, 1950). Put simply, the "Turing test" captures the extent to which a computer can imitate a human interaction partner in a chat. For successfully passing the test, the responses from the computer must be indistinguishable from that of a human being. Returning to Weizenbaum's experiments, his observations fostered Weizenbaum's skepticism and criticism towards AI. In his book "*Computer Power and Human Reason. From Judgment to Calculation*" that was published ten years after the launch of "ELIZA", Weizenbaum warns of the potential for computers to dehumanize and undermine human autonomy and intellect (Weizenbaum, 1976).

Although AI pioneers like Joseph Weizenbaum and Alan Turing were computer scientists, the research field of human-bot interaction (or human-computer interaction in general) is distinguished by its strong interdisciplinary character (Diaper, 1989; Dix, 2010; Ebert et al., 2012). Weizenbaum's early experiments demonstrated the significance of addressing not only the technical aspects of bots but also the subjective interpretations and social needs of living, conscious human users in the development of software systems. Consequently, researchers from the fields of psychology and social sciences are equally essential for studying and understanding the dynamics of interactions between humans and artificial agents like chatbots (Ebert et al., 2012). As a doctoral thesis within the realm of social sciences, this dissertation follows the previously outlined anthropocentric approach and places particular emphasis on the humans and their perceptions of chatbots and the corresponding interactions. Specifically, it examines the extent to which humans perceive chatbots as social

actors, i.e., in how far schemas, concepts, heuristics, and expectations from interpersonal communication are applicable to human-chatbot interactions. As Weizenbaum already illustrated, chatbots are highly responsive and much more social than previous generations of data processing systems as they simulate human-like conversations (Mariani et al., 2023). In addition, many chatbots – but also other automated agents like voice assistants or fully embodied robots – are intentionally humanized through the implementation of various social cues which are also known as "anthropomorphic design elements" (Blut et al., 2021; Feine et al., 2019). For instance, chatbots frequently show personifying elements such as a name or an avatar, as well as verbal social cues like the use of emojis, emotional expressions, or keeping small talk (Feine et al., 2019). Voice assistants like Apple's "Siri" or Amazon's "Alexa" react to calling their name, respond with either a female- or male-sounding voice, and they might tell jokes or respond with irony to specific requests. Physically existing robots like the humanoid companion robot "Pepper" enable an even more immersive humanization. For example, "Pepper" has a human-like body, gesturing arms, and facial expressions that can adopt to the emotional state of the person it interacts with, thanks to emotion recognition software (Glaser, 2016; Spezialetti et al., 2020). To give a better impression on how social cues are employed in practice, Figure 1 illustrates three examples of chatbots and robots that are humanized by different anthropomorphic design elements.

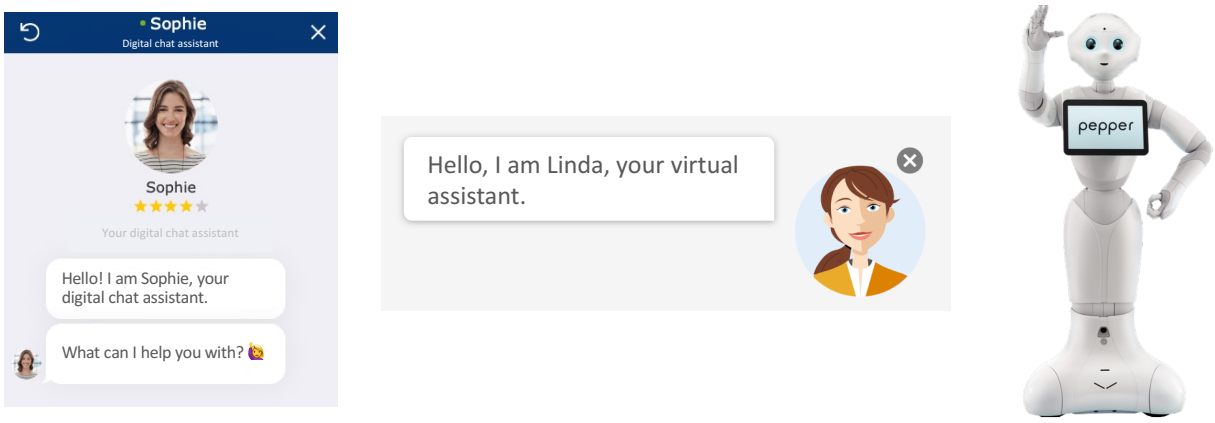


Figure 1. Humanized chatbots and robots.

Notes: Chatbot "Sophie" (Check24), chatbot "Linda" (Sparkasse Bodensee), and social companion robot "Pepper" (from left to right). Texts were translated from German to English. Original sources are cited in the image sources section.

This thesis is structured as follows: first, the subsequent sections of the introduction start with highlighting the practical as well as scientific relevance of examining human-chatbot interactions. In congruence with the topic, a particular focus is set on the social nature of bots and their humanization. Afterwards, it introduces some fundamental theories and approaches for studying social interactions between humans and chatbots and that are vital for all papers in this thesis and its overall contribution. This section contains two sub-sections, one arguing in favor of, and the other arguing against humanization. To substantiate this discussion, a literature overview of related research is given and summarized in a comprehensive table. The introduction closes with outlining the specific research gap and research question. Second, the framework paper provides an overview of the three papers included in this thesis. Instead of going in too much detail, the overview anticipates some key findings and highlights how the individual papers are related to each other and in what way they contribute to answer the overall research question. In addition, I shortly introduce some papers that are not included in this thesis but have emerged during conducting research of this thesis (e.g., conference papers and side projects). Third, the framework paper closes with a comprehensive discussion that is divided

into several sub-sections discussing on the thesis' overall theoretical contributions, managerial implications, limitations and potential future research directions, ethical considerations, and a conclusion. And fourth, following the framework paper, the three papers are included in their original form, i.e., how they have been published or submitted with respect to their content.[2] However, to somehow reach uniformity, the format has been aligned as much as possible (e.g., headings and font, but not citations and references that highly vary across journals).

1.1 Practical Relevance

In times where the impact of scientific research on both society and application is becoming increasingly important, this dissertation starts with outlining the practical relevance of studying human-chatbot interactions before elaborating further on the theoretical foundation and previous research. The undeniably most disruptive innovation emphasizing the topic's relevance is the launch of "ChatGPT" in November 2022. "ChatGPT", where "GPT" stands for "Generative Pre-trained Transformer", is a large language model that described itself as "designed for generating human-like text responses in natural language conversations" when I asked it on October 26, 2023. In contrast to many other chatbots, "ChatGPT" and its successor "GPT-4" show signs of AGI and a very high flexibility (Bubeck et al., 2023). For instance, they can provide recipes for pasta, write poems, explain Einstein's "Theory of Relativity", search for patterns in quantitative data, or even generate codes for smartphone applications within seconds. Unsurprisingly, "ChatGPT" set a record for the fastest-growing user base as it reached one-hundred million active users just two months after its launch – it took the social media platforms "TikTok" more than nine month, and "Instagram" even more than two and a half years to reach that number (Chow, 2023; Hu, 2023; Peres et al., 2023). The extensive applicability and capability of "ChatGPT" have led professionals to anticipate a threat to

[2] The content of manuscripts that have not been published upon submission of this thesis might deviate from final versions due to potential requirements of the peer-review process.

numerous jobs (Briggs & Kodnani, 2023), and the utility of essays to assess students has been questioned (Peres et al., 2023). Universities have thus reacted by introducing guidelines on how to account for and integrate AI-driven systems in teaching (Leuphana Universität, 2023), or they have taken even more extreme measures, such as eliminating bachelor's theses (Zenthöfer, 2023). A recent editorial in the *International Journal of Research in Marketing* kicks of the discussion on the opportunities and threats of generative AI in research, teaching, and practice and introduces corresponding implications (Peres et al., 2023).

However, even before "ChatGPT", automated virtual agents have been considered an emerging technology that will significantly change our everyday life. Microsoft's CEO Satya Nadella already predicted in 2016 that bots will be the new apps (della Cava, 2016). While "ChatGPT" and "GPT-4" are examples for (more or less) simple forms of an AGI, many chatbots are designed for very specific application areas like online retailing (Chung et al., 2020), banking (Trivedi, 2019), customer services (Crolic et al., 2022), healthcare provision (Laranjo et al., 2018), or even mimicking a close friend or romantic relationship partner (Pentina et al., 2023; Skjuve et al., 2021). Given the various application areas and the increasing digitalization, the global chatbot market size is expected to grow from $6 billion in 2023 to more than $26 billion in 2030 with an average yearly growth rate of approximately 23.9% (Grand View Research, 2023; Statista, 2023). Chatbots are thus considered one of the key technologies that will shape the anticipated "5th Industrial Revolution" in service delivery (Huang & Rust, 2018; Noble et al., 2022; Wirtz et al., 2018). Implementing chatbots successfully can thus provide companies with competitive advantages as they can enhance service efficiency, reduce costs, enable standardization of service delivery, and support and relieve human agents in accomplishing tasks (Huang & Rust, 2018; Sheehan et al., 2020).

Despite these opportunities, companies are faced with the lacking acceptance of chatbots by customers. For instance, research has shown that humans still prefer to interact with

human agents (Zhang et al., 2021) and that service evaluation is worse when the service provider is a bot and not a human, even when the provided service is identical (Castelo et al., 2023). The underlying reasons for the tendency to reject chatbots are manifold including lower expectations towards their technical capabilities (e.g., regarding their efficacy and flexibility; Crolic et al., 2022; Yu et al., 2022) and a lack of human warmth (Borau et al., 2021; Gelbrich et al., 2021). To tackle this problem, many companies humanize their chatbots to make the interaction feel more familiar and natural (see Figure 1). Given that chatbots will increasingly complement or even replace employees, it is vital for companies to understand if, how, and when their attempts at humanization are truly beneficial or if there might be backfiring effects and boundary conditions. Only with such an understanding, it will be possible to design and implement powerful chatbots that satisfy customers' needs and are thus more likely to be accepted.

1.2 Scientific Relevance

Studying human-bot interactions with a particular focus on the sociality of bots has also increasingly gained attention in the scientific community. To illustrate this, I conducted a convenience literature search on the "Web of Science" database for articles including relevant keywords in the title, e.g., "chatbot" and "human-like" (see Figure 2).

TI=(("chatbot" OR "robot*" OR "virtual assistant*" OR "conversational agent*" OR "virtual agent*" OR "voice assistant*") AND ("humanlike*" OR "human-like*" OR "humanness" OR "anthropomorph*" OR "social*"))*

Figure 2. The Boolean operator used for the literature search.

Without further filtering or elaborating, "Web of Science" found 2,198 articles, 1,224 of which (55.7%) being published 2020 or later, i.e., during the preparation period of this dissertation. Congruently, a recent systematic literature review on conversational agents including 554 articles shows that only 178 (32.1%) have been published before, and 376

(67.9%) after 2020 (Mariani et al., 2023). Shifting from the publication to the journal level, there have been recent calls for papers in leading business and marketing journals on human-bot interactions, e.g., *Psychology & Marketing* ("Virtual Conversational Agents: Consumer-Machine Relationships in the Age of Artificial Intelligence", 2022), *Journal of Business Research* ("Unanticipated and Unintended Consequences of Service Robots in the Frontline", 2023), and *Journal of Service Research* ("Human-Robot Interactions in Service", 2023). On top of that, the academic publisher "Elsevier" recently launched a journal called *Computers in Human Behavior: Artificial Humans* dealing exceptionally with the social nature of bots. To summarize, there is a lot going on in academic research in studying the dynamics of human-bot interactions in general and with a focus on their sociality in particular. As the development of intelligent bots is still in its infants while they are anticipated to take over a significant role in society, there is still a lot to learn (Blut et al., 2021; Han et al., 2023; Uysal et al., 2022).

1.3 Theoretical Foundation and Literature Review

1.3.1 The Bright Side: Social Bots as Companions

Despite being highly topical, scientific research on the social nature of artificial agents already began in the early to mid-1990's when computers became increasingly prevalent. Although Weizenbaum already noticed human's social reactions towards computers in 1966, it took academia about three decades to build a widely accepted and citable theory. In their prominent book "*The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*" (1996), Byron Reeves and Clifford Nass from Stanford University argue that human's interactions with computers are fundamentally social in nature. In numerous experiments, they found empirical evidence that humans apply politeness and reciprocity norms as well as gender stereotypes to computers (Nass & Moon, 2000; Reeves & Nass, 1996). These early findings culminated in the "Social Response Theory" and the closely related "Computers are Social Actors" (CASA) paradigm (Reeves & Nass, 1996) which are still

referenced frequently by researchers laying their theoretical foundation on humans' social reactions towards artificial agents (Mariani et al., 2023). These theoretical approaches argue that human brain structures developed in a time where social cues and human-like behavior were uniquely human. Today, however, artificial agents like computers, chatbots, or robots frequently mimic humans in their appearance or performance of tasks hence facilitating the adoption of social rules (Konya-Baumbach et al., 2023; Nass & Moon, 2000; Reeves & Nass, 1996).

However, social responses towards non-human entities are not limited to technical agents like computers and bots but also apply to other inanimate objects like cars (Aggarwal & McGill, 2007; Chandler & Schwartz, 2010; Landwehr et al., 2011), brands (Puzakova et al., 2013; Sharma & Rahman, 2022), or products (Velasco et al., 2021). Responding socially towards inanimate entities results from the perception or attribution of human-like qualities to a non-human agent. In psychology, this phenomenon is known as "anthropomorphism". According to Epley et al. (2007), anthropomorphism "describes the tendency to imbue the real or imagined behavior of nonhuman agents with humanlike characteristics, motivations, intentions, or emotions". This innate tendency can already be observed in young children, e.g., when they talk to their stuffed animals, feed them, or put them to sleep. In adults, anthropomorphism might manifest in believing that plants enjoy a touch, in begging an old car to start, in seeing human faces in clouds, or in assuming bad intentions in a printer that continues to work unreliably (Epley et al., 2007; Seitz & Bekmeier-Feuerhahn, 2023a). In their well-established "Three Factor Theory of Anthropomorphism", Epley et al. (2007) attribute human's tendency to anthropomorphize to *one* cognitive and *two* motivational factors. Starting with the cognitive one ("elicited agent knowledge"), humans are more likely to anthropomorphize when the accessibility and applicability of anthropocentric knowledge is high. This might be the case when the application of such knowledge seems appropriate, e.g., when the object has some

human-like qualities or shows human-like behavior. Proceeding with the first motivational factor ("sociality motivation"), the theory argues that anthropomorphism is an intuitive strategy to fulfill humans' inherent desire for social connectedness and embeddedness. Congruently, people have a higher tendency to anthropomorphize objects when they feel lonely and socially isolated (Epley et al., 2008). The second motivational factor ("effectance motivation") attributes anthropomorphism to humans' motivation to be able to explain and understand their environment and the behavior of the agents they are interacting with. In this sense, anthropomorphism is a strategy to enhance the perceived ability to explain and predict an agent's behavior which can help in reducing uncertainty and making sense of an agent's actions.

Even though anthropomorphism is closely related to "Social Response Theory" (or CASA paradigm), they are not the same. First, anthropomorphism is a broader concept which does not only apply to interactions with responsive technology but various kinds of inanimate (e.g., cars) and animate (e.g., pets) agents. Second, while "Social Response Theory" focuses on social responses that are triggered by perceiving social cues in an agent, anthropomorphism describes humans' general tendency to attribute human-like qualities to non-human agents (e.g., having emotions, intentions, or personality traits). "Social Response Theory" hence focuses the behavioral dimension (i.e., social behavior resulting from a stimulus-response mechanism) while anthropomorphism can be considered a cognitive bias that must not necessarily be triggered by external stimuli (although it can). And third, "Social Response Theory" argues that social responses towards computers are mindless while anthropomorphism can be either mindless or mindful. In substantiating their perspective, the authors of "Social Response Theory" and the CASA paradigm reported that the participants in their experiments either failed to recognize their social responses, denied them, or acknowledged the irrationality of responding socially to computers. On the other hand, anthropomorphism can also be mindful,

e.g., when humans explicitly belief that a plant can feel joy or that their dog has intentions (Epley et al., 2007; Kim & Sundar, 2012).

Regardless of the preferred theoretical lens, both anthropomorphism as well as "Social Response Theory" can be used to explain humans' social responses towards bots, especially when they are humanized by certain anthropomorphic design elements (see Figure 1). A chatbot that is humanized (vs. non-humanized) is hence more likely to elicit the activation of human-like schemas and the application of social rules and heuristics (Blut et al., 2021; Crolic et al., 2022; Epley et al., 2007; Konya-Baumbach et al., 2023; Nass & Moon, 2000; Reeves & Nass, 1996). Hence, humanized (vs. non-humanized) chatbots have a higher chance to be consciously or unconsciously perceived and treated as social actors (see Figure 3).

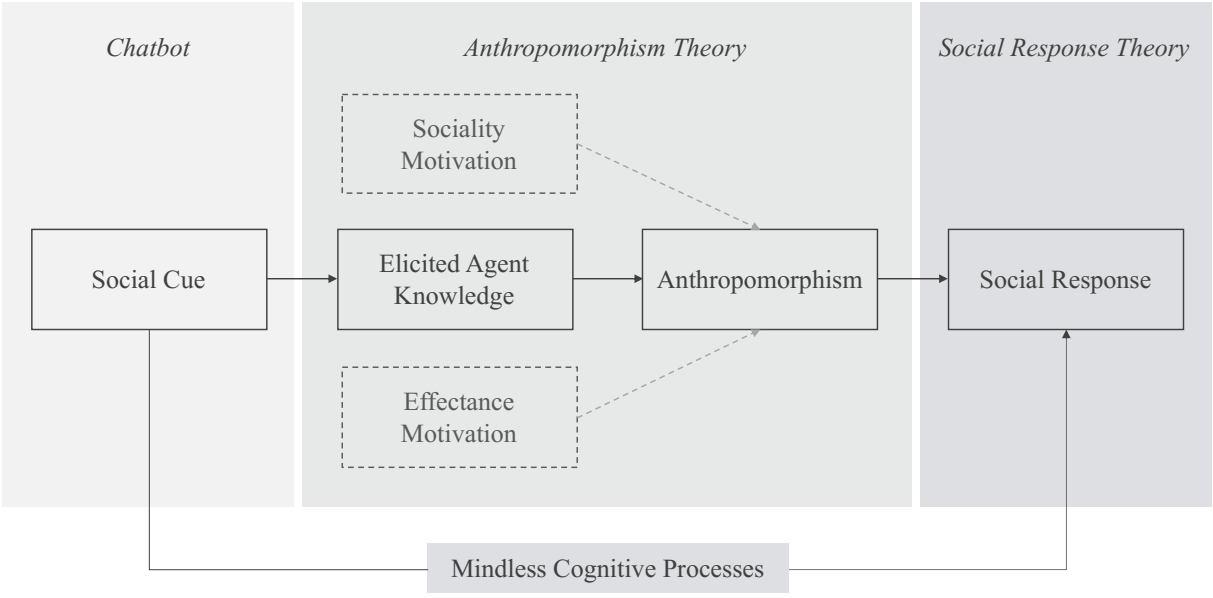


Figure 3. Integrative framework of "Anthropomorphism Theory" and "Social Response Theory" in explaining social responses towards chatbots. (Own illustration)

Moving towards the outcomes of humanizing artificial agents like chatbots and anthropomorphism, there is a lot of research revealing several positive effects. A recent meta-analysis including 108 samples and 3404 effect sizes concludes that anthropomorphism in physical robots, chatbots, and AI leads to improved outcomes such as higher using intentions

(Blut et al., 2021). Diving deeper into the underlying mechanisms, the authors argue that humanized bots and anthropomorphism serve the two motivational needs illustrated in the "Three Factor Theory of Anthropomorphism": sociality motivation (i.e., humans' desire for social connectedness) and effectance motivation (i.e., humans' desire to interact with their environment successfully). Humanized bots are more likely to be perceived as social actors leading to enhanced feelings of social presence (Go & Sundar, 2019; Konya-Baumbach et al., 2023; van Doorn et al., 2017) and human warmth (Borau et al., 2021; Christoforakos et al., 2021; Gelbrich et al., 2021). This can facilitate relationship- and trust-building enabling humans to fulfil their social-emotional and relational needs (Wirtz et al., 2018). Research from other marketing domains like branding reveals similar effects: an anthropomorphized brand can create a more intense emotional attachment resulting in stronger customer-brand relationships and more positive brand attitudes and evaluations (Rauschnabel & Ahuvia, 2014; Seitz & Bekmeier-Feuerhahn, 2023a; Sharma & Rahman, 2022; Velasco et al., 2021; Yuan & Dennis, 2019). Also, the elicitation of social scripts and expectations by human-like cues can reduce uncertainty and enhance perceived familiarity and predictability. This might make human-bot interactions feel easier leading to a more positive evaluation of the bot (Blut et al., 2021; Duffy, 2003; Nass & Moon, 2000). For a better overview, Table 1 summarizes some of the existing evidence on the positive effects of humanization and anthropomorphism on theoretical and practical meaningful outcome variables like trust, satisfaction, and using intentions.

1.3.2 The Dark Side: Social Bots as Mindful and Threatening Actors

According to the previous section, artificial agents like chatbots should be humanized to benefit from the presented positive effects. However, in recent years, several papers have been published dealing with potential negative downstream consequences. The potentially most prominent theory that is referred to when arguing against the humanization of bots is the "Uncanny Valley" that was originally introduced by Masahiro Mori in 1970 (Mori, 1970; Mori

et al., 2012). The "Uncanny Valley" suggests a non-linear relationship between human-likeness and affinity. Specifically, it posits that as inanimate agents become more human-like in their appearance, there is a point at which the response of a human observer turns from positive to negative before becoming positive again once the agent becomes barely indistinguishable from a human (see Figure 4).

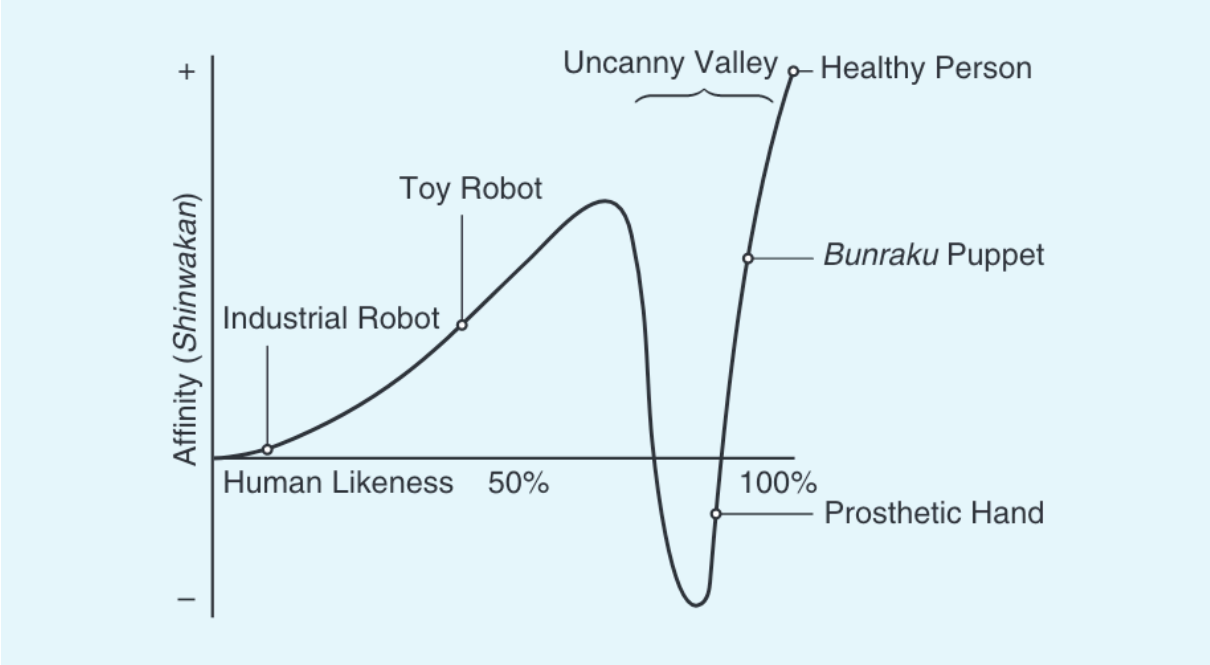


Figure 4. The "Uncanny Valley". (Mori et al., 2012, p. 99)

The assumption is that the dip results from a conflict between the human observer's expectation of human-like behavior and the actual (imperfect) capabilities of the agent to fulfill these expectations. These perceived imperfections and oddities are believed to make the agent appear eerie and creepy. In explaining this effect, the authors argue from an evolutionary theory's perspective postulating that humans have a natural tendency to avoid corpses (i.e., subjects that appear to be living humans at first glance but turn out to be dead). Although there is some critique and ambivalence on the "Uncanny Valley" and its line of argumentation (see Bartneck et al., 2007; Burleigh et al., 2013; Kätsyri et al., 2015), there is empirical evidence for negative emotions and feelings of eeriness elicited by humanized bots. In contrast to the original

paper on the "Uncanny Valley", authors of more recent papers argue that the underlying mechanism stems from a perceived threat to human identity undermining our sense of human uniqueness and distinctiveness (Ferrari et al., 2016). For instance, research has found that a bot-induced perceived threat to human identity can elicit compensatory behavior like enhanced consumption (Mende et al., 2019) or putting higher value on human unique attributes like social creativity (Cha et al., 2020). Also, highly human-like avatars in a virtual reality setting (Stein & Ohler, 2017) as well as robots (Złotowski et al., 2017) were found to be perceived eerier and more threatening when participants believed them to have an autonomous mind (vs. follow a pre-defined script). Moving beyond eeriness perceptions, mind attribution (e.g., the believe in a bot's capacity to have agency or even experiential capabilities; Epley et al., 2007; Gray et al., 2007) can also backfire for other reasons. First, mind attribution can elicit unrealistic high expectations towards a bot's capabilities. Recent research has shown that a humanized (vs. non-humanized) service chatbot enhances customers' expectations regarding its problem-solving capabilities leading to expectancy violations and frustration in case of service failure (Crolic et al., 2022). Second, mind attribution is associated with the capacity for social and moral judgment (Gray et al., 2007; Pitardi et al., 2022). In service environments bearing a high social risk (e.g., embarrassing service encounters), humanization and the resulting perceived social presence can lead to discomfort and have adverse effects (Holthöwer & van Doorn, 2022; Kim et al., 2022). Also, assuming the capacity for moral judgment can backfire in cases of misbehavior since observers are more likely to attribute responsibility to mindful agents (Kwak et al., 2015; Puzakova et al., 2013). In addition to the positive effects, Table 1 also presents selected studies on negative effects of humanization and anthropomorphism.

Paper	Method	Agent type	Manipulation of human-likeness	Effect valence	Key findings
Qiu & Benbasat (2009)	Laboratory experiment	Digital assistant	Yes	Positive	A humanized (vs. non-humanized) product recommendation agent increases trusting beliefs, perceived enjoyment and using intentions. The effect is mediated by an increase in perceived social presence.
de Visser, Monfort, McKendrick, Smith, McKnight, Krueger & Parasuraman (2016)	Laboratory experiments	Digital assistant	Yes	Positive	A humanized (vs. non-humanized) digital assistant is associated with higher resistance to breakdowns in trust.
Ferrari, Paladino & Jetten (2016)	Laboratory and online experiments	Robot	Yes	Negative	The more human-like a robot appears, the higher is the perceived damage to humans. The effect is mediated by an increase in perceived threat to human identity.
Kim, Chen & Zhang (2016)	Online experiments	Digital assistant	Yes	Negative	Receiving help from a humanized (vs. non-humanized) digital assistant reduces perceived enjoyment in playing computer games. The effect is mediated by a loss in perceived autonomy.
Stein & Ohler (2017)	Laboratory experiment	Virtual avatar	Yes	Negative	Computer-controlled virtual avatars elicit higher levels of eeriness when they are believed to act autonomously (vs. scripted).
Araujo (2018)	Online experiment	Chatbot	Yes	Positive	A humanized (vs. non-humanized) chatbot enhances emotional connection with the company. The effect is mediated by an increase in perceived social presence.
Go & Sundar (2019)	Online experiment	Chatbot	Yes	Positive	Human-like design elements (vs. no such elements) in a chatbot can compensate for impersonal communication and low levels of message interactivity.
Mende, Scott, van Doorn, Grewal & Shanks (2019)	Laboratory and online experiments	Robot	Yes	Negative	Consumers display compensatory consumption behavior after having interacted with a humanoid robot (vs. non-humanoid robot and vs. human). The effect is serially mediated by an increase in perceived eeriness and identity threat.
Borau, Otterbring, Laporte & Wamba (2021)	Online experiments	Chatbot and robot	Yes	Positive	Female robots and chatbots are preferred over male equivalents. The effect is serially mediated by an increased perceived warmth making the bot feel more human.
Choi, Mattila & Bolton (2021)	Online experiments	Robot	Yes	Ambivalent	A humanoid (vs. non-humanoid) robot is more strongly associated with human warmth. This leads to lower (higher) satisfaction in case of service process failure when the robot does not apply service recovery strategies (does apply such strategies).
Gelbrich, Hagel & Orsingher (2021)	Online experiments	Digital assistant	Yes	Positive	A digital assistant providing emotional support (vs. no emotional support) enhances behavioral persistence in using technology-mediated services. The effect is serially mediated by an increase in perceived warmth and satisfaction.

Han (2021)	Online survey	Chatbot	No	Positive	Chatbot anthropomorphism in e-commerce enhances purchase intentions. The effect is mediated by an increased perceived social presence and perceived enjoyment.
Barney, Hancock, Jones, Kazandjian & Collier (2022)	Online experiments	Digital assistant	Yes	Positive	A humanized (vs. non-humanized) shopper assistant increases purchase intentions. The effect is serially mediated by an increase in perceived immersion and attitudes towards the app.
Crolic, Thomaz, Hadi & Stephen (2022)	Online experiments and field study	Chatbot	Yes	Negative	Chatbot anthropomorphism (either manipulated or measured) has negative downstream consequences on customer satisfaction, company evaluation, and purchase intentions when users are in an angry emotional state. The effects are mediated by expectancy violations caused by inflated pre-encounter expectations of chatbot efficacy.
Holthöwer & van Doorn (2022)	Laboratory and online experiments, field study	Robot	Yes	Negative	A humanized (vs. non-humanized) robot reduces the intention to acquire a medicine that is considered embarrassing to purchase. The effect is serially mediated by an increase in perceived social presence and social judgment.
Lv, Yang, Qin, Cao & Xu (2022)	Online experiments	Digital assistant	Yes	Positive	Empathetic (vs. non-empathetic) responses in an AI assistant increase using intentions. The effect is serially mediated by a reduction in perceived psychological distance and trust. The effect is larger for voice-based (vs. text-only) assistants.
Uysal, Alavi & Bezençon (2022)	Surveys, field experiment and field study	Voice assistant	No	Negative	Voice assistant anthropomorphism causes an identity threat resulting in negative downstream consequences on consumer empowerment and well-being. The effect particularly occurs in long-time relationships.
Han, Deng & Fan (2023)	Online experiments	Robot	Yes	Ambivalent	Consumers with a competitive mindset respond less favorably to humanized (vs. non-humanized) robots. The opposite is true for consumers with a collaborative mindset. The effect is mediated by an increased perceived psychological closeness towards humanized robots.
Konya-Baumbach, Biller & von Janda (2023)	Online experiments	Chatbot	Yes	Positive	A humanized (vs. non-humanized) chatbot enhances trust, purchase intention, word of mouth, and satisfaction with the shopping experience. The effects are mediated by an increase in perceived social presence.

Table 1. Selected studies on the effects of humanizing bots and anthropomorphism.

1.3.3 Research Gap and Research Question

Table 1 suggests that humanizing bots and anthropomorphism is a quite well-researched phenomenon. However, with taking a more nuanced perspective, there are some shortcomings and open questions to address. First, there is ambivalence regarding the effect valence that is

not only evident between studies, but also within (e.g., Choi et al., 2021; Han et al., 2023). Research has found that the effect direction of humanization and anthropomorphism might depend on contextual factors (e.g., service task; Seeger et al., 2021), individual factors (e.g., value orientation or need for human interaction; Han et al., 2023; Sheehan et al., 2020), and bot-related factors (e.g., type of social cue; Blut et al., 2021). Thus, there is still much uncertainty regarding what kind of humanization is beneficial (vs. harmful) for whom and under what circumstances. Second, research has only begun to examine potential backfiring effects of humanization in recent years. Except from some early work on the "Uncanny Valley" and perceived identity threat (e.g., Ferrari et al., 2016; Stein & Ohler, 2017), many papers have been published during the preparation period of this thesis (e.g., Choi et al., 2021; Crolic et al., 2022; Han et al., 2023; Holthöwer & van Doorn, 2022). In addition, research on backfiring effects for non-embodied chatbots is scarce (e.g., Crolic et al., 2022). Third, many of the existing research focuses on physical robots. However, since chatbots are less human-like than fully embodied robots (e.g., they neither move nor have a voice), it is questionable in how far findings from interactions with social robots are applicable to interactions with chatbots (Blut et al., 2021). For instance, as humans usually clearly identify non-embodied chatbots as software systems, it is unclear if they can reach sufficient levels of human-likeness to fall into the "Uncanny Valley" (Skjuve et al., 2019; Yanxia et al., 2023). Also, it is to be studied if humans truly expect the fulfillment of social needs and relationship-building from software-based and mostly goal-oriented chatbots. Although there are some examples for social companion chatbots (e.g., "Replika"), the major purpose of most chatbots is to enhance the efficiency of performing a specific task. Congruently, research has shown that the fulfilment of utilitarian needs is central in interactions with self-service technologies like chatbots (Blut et al., 2016). And fourth, many of the existing research on both positive and negative effects assumes that social cues in bots elicit human-like schemas having corresponding consequences for the perception of and

expectation towards the bot. Taking the example of empathy, the social cue of empathetic expressions in chatbots might enhance feelings of emotional support and warmth having positive consequences on trust development, just like in interpersonal interactions (Cheng et al., 2022; Gelbrich et al., 2021; Pelau et al., 2021). On the other hand, empathetic expressions might also enhance feelings of social presence and mind attribution that can have adverse effects when humans fear to be socially judged, just like in interpersonal interactions (Holthöwer & van Doorn, 2022; Kim et al., 2022; Pitardi et al., 2022). Although these findings make significant contributions to our theoretical understanding of the social perception of bots, there is barely research examining in how far chatbots are *not* perceived as social actors but technological entities with specific traits and characteristics. In a broader sense, only the "Uncanny Valley" and related research on identity threats state that perceiving uniquely human-like qualities in bots (e.g., a mind or autonomy) might feel eerie. Even though mechanistic stereotypes towards chatbots are known in research and conventional wisdom (e.g., their incapability to feel emotion or being faster and less flexible than human agents), the role of the different schemas humans have about bots vs. humans has only been considered by a few academic papers (e.g., Meng & Dai, 2021; Yu et al., 2022).

To sum up, more research is required to understand whether incorporating social cues to non-embodied chatbots truly yields significant benefits or if the potential of humanization is limited. In three empirical papers including a total of ten studies, this thesis aims at extending our current knowledge on the social perception of chatbots and potential boundary conditions and backfiring effects of humanization. The first paper examines in two studies how patients develop trust towards healthcare chatbots by applying qualitative methods. A particular focus is set on understanding differences between the trust development process towards chatbots vs. human agents (i.e., physicians). Building on these findings, the second paper takes a detailed perspective on the chances and risks of implementing expressions of empathy to diagnostic

chatbots. In three experimental studies, it examines if experiential expressions of empathy by which the chatbot pretends to be able to feel *with* or *for* the patient backfire by feeling inauthentic. The third paper goes beyond the healthcare context and examines whether incorporating service chatbots with human-like response delays has negative downstream consequences. In five experimental studies, it examines if response delays have adverse effects on using intentions and company evaluation as response delays are hypothesized to violate the expectation of receiving a fast service from a chatbot. From a meta-perspective, the findings from these papers contribute to a better theoretical understanding what differentiates the perception and evaluation of a chatbot from that of a human agent. Besides its theoretical contributions, this thesis can also help practitioners in deciding for or against implementing specific social cues to their chatbots.

2 Paper Overview

2.1 Can We Develop Trust in Chatbots as We Do in Physicians?

The first paper originated from a research project conducted in cooperation with a medical service provider, focusing on examining critical factors that drive the development of trust towards diagnostic chatbots. In this initial research, we focused on trust for several reasons: first, trust is a central concept in interpersonal relations (Lewis & Weigert, 1985; Luhmann, 1979; Rempel et al., 1985). As there was limited prior research when the project started, delving into the extent to which trust is applicable in interactions with chatbots allowed us to gain a comprehensive initial understanding of the social perception of chatbots. Second, trust is particularly vital in healthcare and doctor-patient relationships, given the sensitivity, high risk, and vulnerability of patients (Buchanan, 1988; Hillen et al., 2017; Pearson & Raeke, 2000). And third, the novelty and complexity of diagnostic chatbots might entail a lack of trust potentially undermining their adoption (Benbasat & Wang, 2005; Christoforakos et al., 2021; Gefen et al., 2008). Since humans are generally reluctant to adopt chatbots (Araujo, 2018;

Castelo et al., 2023; Zhang et al., 2021), establishing trust becomes essential for increasing their prevalence.

The main goal of this paper was to examine if humans develop trust in chatbots in a similar manner like they do in interpersonal relationships. Drawing on the assumption that chatbots have elements of both technological tools and social actors, the paper starts with a literature review on theories and approaches on interpersonal trust in general, trust towards physicians in particular, and technology trust. In this regard, it outlines peculiarities as well as similarities and differences between these theories and approaches. To answer the broad and open research question, we first conducted a laboratory experiment (Study 1) in which participants had to take the perspective of a patient suffering from symptoms that were described in a scenario. Afterwards, they either interacted with a diagnostic chatbot only or with an additional physician after they had received the preliminary assessment from the chatbot. Data was collected by semi-structured pre- and post-interaction interviews focusing on the process and drivers of trust development. The interview manuscripts were analyzed and coded both inductively and deductively following the "Summarizing Content Analysis" approach (Mayring, 2000; Mayring, 2014). As a follow-up, we verified the coding system in a larger online survey (Study 2) during the COVID-19 pandemic in which participants interacted with a chatbot that was able to assess an individual's risk of a Corona infection.

We identified several internal factors (software-related) as well as external factors (user- and environment-related) influencing the trust-building process which are described in detail in the manuscript. Anticipating some of the key findings that significantly guided the research direction for the following papers and that contribute to the overall story of this thesis, we found that trusting diagnostic chatbots is driven cognitively (i.e., for good reasons) while trusting physicians is also affect-based (i.e., driven by emotion). In this regard, a significant finding that motivated us to dive deeper into potential boundaries and backfiring effects of social cues was

that participants barely indicated to expect the chatbot to appear or to communicate like a human. Specifically, there was evidence that incorporating chatbots with social cues (e.g., empathetic expressions) could even elicit distrust as they appear to be fake. These findings confirmed humans' awareness of the inanimateness of bots and that social responses or anthropomorphism could rather be mindless processes (Nass & Moon, 2000; Reeves & Nass, 1996). Hence, social cues that enter our consciousness due to their explicit conflict with mechanistic stereotypes and expectations towards bots could trigger cognitive processing and have negative downstream consequences when they appear to be not real or unnecessary.

2.2 Does Artificial Empathy in Chatbots Feel Authentic?

The second paper followed up on the findings from the first paper and examined in how far the interpersonal concept of empathy is applicable to interactions with healthcare chatbots. Building on "Social Response Theory", researchers frequently argue that incorporating chatbots with expressions of empathy can enhance perceived warmth resulting in higher trust and using intentions (Christoforakos et al., 2021; Gelbrich et al., 2021; Lv et al., 2023; Pelau et al., 2021). However, the findings from the first paper pointed to the opposite direction as participants reported that empathy is not required in chatbots because it would feel fake. To gain a deeper understanding what could precisely make empathy in chatbots feel fake, I conducted a literature research on the concept of empathy, empathy in bots, and core differences between humans and bots. In essence, the literature research revealed it could be the expression of experiences (i.e., when a bot pretends being able to feel *with* or *for* a human) that seems fake since bots are poorly associated with experiential capabilities, as argued by "Mind Perception Theory" (Gray et al., 2007; Gray & Wegner, 2012). If so, expressions of behavioral empathy (i.e., empathetic helping) by which chatbots provide instrumental rather than emotional support might be more appropriate in modeling artificial empathy.

The overall goal of the second paper was to examine if experiential expressions of empathy (vs. behavioral empathy) backfire by feeling fake resulting in a reduction of the chatbot's perceived authenticity, trust, and ultimately using intentions. I conducted two experimental online studies in which participants took the position of a sick person before interacting with a diagnostic chatbot programmed for the purpose of this paper. Participants were randomly assigned to one of four different diagnostic chatbots whose communication styles were designed in congruence with established empathy theories and related research (*empathetic*: feeling with; *sympathetic*: feeling for; *behavioral-empathetic*: empathetic helping; *non-empathetic*: control condition).

Results from parallel mediation analyses revealed that the positive effect of empathy on trust and using intentions mediated by an increased perceived warmth is attenuated by a simultaneous loss in perceived authenticity for the chatbots utilizing experiential expressions of empathy. The effect occurred independently of whether the bot showed personifying elements (Study 2) or not (Study 1). A third study did not replicate the backfiring effect when participants watched a chat between a patient and a human physician, i.e., experiential expressions were only inauthentic in bots, not in humans. The second paper thus verifies the findings from the first paper by applying quantitative methods. More precisely, it shows that integrating uniquely human-like attributes to chatbots (i.e., experiential expressions of empathy) can backfire by feeling inauthentic. Incorporating these findings into the broader context of this thesis, the second paper demonstrates that humans apply distinct schemas and expectations to their interactions with chatbots by delineating the perception of artificial from interpersonal empathy.

2.3 Should Chatbots Respond as Slow as Humans Just to Be More Human?

The third paper examined if a social cue that reduces a chatbot's efficiency (namely "dynamic response delays") backfires by harming the expectation of receiving a fast and

convenient service. As we have learned from the second paper, humans might enter conversations with chatbots with computer-like schemas, expecting cold but immediate responses (Meng & Dai, 2021). Since increasing the efficiency is a decisive benefit of service chatbots and vital for a technology's perceived usefulness (Blut et al., 2016; Venkatesh et al., 2012), response delays like we know them from interpersonal chats ("*person is typing…*") could lead to expectancy violations resulting in a negative evaluation of the chatbot and the service provider (Crolic et al., 2022). However, if this line of argumentation sustains, the backfiring effect should not occur when users apply human-like schemas to the interaction (i.e., when they expect it to behave like a human).

The goal of the second paper was therefore to examine if humanization backfires when the incorporated social cue is inconsistent with computer-like schemas and usefulness expectations. To test the hypotheses, we conducted five experimental studies in which participants either watched pre-recorded videos of an interaction between a customer and a service agent (either a chatbot or a human agent) or interacted with service chatbots programmed for this paper. Participants were given the task to take the perspective of a person searching for a city trip (Study 1–3 and 5) or seeking for train tickets vs. a medical assessment (Study 4). The main manipulation in all studies was the chatbot's response delay that was either static and very short (about one second) or dynamic depending on the messages' length (like in interpersonal chats, i.e., typing-in "Ok" needs less time than "Ok, see you tomorrow"). In all experiments, participants were randomly assigned to one of the conditions.

Results from simple mediation and moderated mediation analyses revealed that dynamic response delays in chatbots reduce using intentions and attenuate service provider evaluation and that the underlying mechanism stems from violated usefulness expectations. However, the effect was attenuated when participants applied human-like schemas in the interaction, i.e., when they tended to anthropomorphize chatbots (Study 1 and 3), when they

believed the agent to be a human (Study 2), or when the service task was computer- vs. human-like (Study 4). Considering the big picture of this thesis, the third paper demonstrates that users might enter chatbot conversations with either computer-like or human-like schemas that significantly shape users' expectations towards the chatbot's behavior and the perception of social cues. It therefore extents the second paper in two ways: first, it demonstrates that humans not only expect chatbots to lack emotions but also to respond instantly due to the elicitation of computer-like schemas and stereotypical associations. And second, it showcases that the extent to apply computer-like schemas might depend on individual traits, characteristics of the agent, and contextual factors.

2.4 Related Research and Papers Not Included in This Thesis

In conducting research for this thesis, we published four additional papers and two abstracts that are not included but worth mentioning due to their strong relation to this research. Two of the published papers (Seitz et al., 2020; Seitz & Bekmeier-Feuerhahn, 2021) and one abstract (Seitz & Bekmeier-Feuerhahn, 2023b) have a direct relation to this thesis as they represent peer-reviewed conference papers on each of the three thesis's manuscripts. More detailed information on these papers is given on the fact sheet preceding the corresponding paper. Another conference paper presented at the *European Conference on Information Systems 2021* (VHB: B) was produced in the realm of the first paper and considers the differences in the emergence of distrust vs. trust towards diagnostic chatbots (Seitz et al., 2021). The original project was initiated and led by a student assistant and extended in our research project helping us to better understand barriers in adopting diagnostic chatbots. Motivated by a recent call for action in Blut et al. (2021) to extent the considered outcome variables of anthropomorphism in bots, we further conducted research on the impact of social cues in a chatbot on customers' willingness to pay for a product. A first study was presented at the *European Marketing Academy 2022* (VHB: D; Seitz & Bekmeier-Feuerhahn, 2022) and extended in a seminar and

a bachelor's thesis. We found slight but ambivalent evidence that social cues in product recommendation chatbots might have the potential to enhance customers' willingness to pay, particularly for hedonic (vs. utilitarian products). However, this research is not further considered since (1) it does not contribute to better understand potential adverse effect of humanization and (2) further research is needed to resolve the ambivalence and to shed light into the underlying mechanisms. A last paper published in the journal *transfer – Zeitschrift für Kommunikation und Markenmanagement* (Seitz & Bekmeier-Feuerhahn, 2023a) extends anthropomorphism to the branding context. It provides a literature-based overview of the theoretical foundation of anthropomorphic brand design and illustrates advantages and disadvantages of an anthropomorphic brand strategy using numerous practical examples. Related to this thesis, the paper closes with an outlook on the role of anthropomorphism in the era of new technologies and digitalization with an emphasis on potential chances and risks.

3 Discussion

3.1 Summary

A summary of the papers' key findings and contributions to this thesis is provided in Table 2. Taking a very generalist meta-perspective, this thesis found both qualitative and quantitative evidence that humans do not generally apply human-like schemas and social heuristics to their interactions with chatbots. Precisely, it found that humans might expect a chatbot to act technical and computer-like which can result in adverse effects when the chatbot's behavior contradicts these schemas and expectations. Theoretical contributions, managerial implications, limitations and potential future research directions, as well as ethical considerations of these findings are discussed in the subsequent sections.

Paper No.	Research question	Methods	Key contributions to this thesis
1	How does trust towards diagnostic chatbots emerge and what are the differences compared to trust in physicians?	Qualitative (laboratory experiment and online survey)	The emergence of trust is influenced by software-, user-, and environment-related factors. A significant difference is that trust in chatbots emerges for rational reasons (cognitively) while trust in physicians is also driven by emotion (affectively). Participants rather expected a trustworthy chatbot to be objective and reliable rather than human-like. In contrast, uniquely human-like design elements like empathetic expressions might even reduce trust since they appear fake. These findings motivated this thesis to dive deeper into different mental schemas and expectations humans have towards chatbots vs. humans.
2	Is the concept of empathy equally applicable to interactions with healthcare chatbots or does it feel inauthentic?	Quantitative (three online experiments)	The paper shows that experiential expressions of empathy (feeling *with* or *for* another) feel inauthentic. The results demonstrate that human-unique attributes (e.g., the capacity to empathize or sympathize with others) are perceived different in chatbots vs. humans. Artificial empathy might rather be conceptualized by the provision of instrumental support as it interferes less with mechanistic stereotypes towards chatbots.
3	Do computer-like schemas in chatbot interactions elicit the expectation for prompt service and do social cues backfire when contradicting this expectation?	Quantitative (five online experiments)	The social cue of "dynamic response delays" making a chatbot's responses more human-like but slower backfires by violating usefulness expectations. As the effect is attenuated when users apply human-like schemas, the paper reveals that computer-like schemas lead users to expect chatbots to prioritize speed and utility over perfect human-likeness.

Table 2. Summary of key findings and contributions.

3.2 Theoretical Contributions

This thesis contributes to a better understanding of humans' interactions with chatbots. More precisely, it demonstrates that humans might perceive their interactions with chatbots different from interpersonal chat-mediated interactions. In this realm, it illustrates limits and boundaries of humanizing chatbots.

Reflecting "Social Response Theory" and anthropomorphism. The "Social Response Theory" (Nass & Moon, 2000) as well as anthropomorphism (Epley et al., 2007) lay the overall theoretical foundation for this thesis and the papers included. The dissertation's overall goal was to enhance our understanding if and to what extent these theories apply to chatbot interactions. Specifically, it challenged the assumption that humans treat chatbots as social actors by examining potential perceptual and evaluative differences. First, the results supported that social responses to chatbots are predominantly mindless processes confirming "Social Response Theory". Specifically, participants in Papers 1 and 3 reported that they were aware of the inanimateness and the technical nature of chatbots. Results also revealed that fundamental interpersonal concepts like trust, warmth, and authenticity play a role in interactions with chatbots and their evaluation. For instance, expressions of empathy can facilitate trust towards chatbots by enhancing perceived warmth, just like in interpersonal interactions. However, all papers debunk the assumption that interactions with chatbots are *fully* social in nature or follow the *exact* same rules and expectations like in interpersonal interactions. Particularly Papers 2 and 3, which manipulated social cues in the chatbots, found different reactions to verbal and non-verbal cues when shown by a chatbot vs. human. Also, if and in how far humans apply social heuristics and expectations was found to depend on individual or contextual factors (e.g., an individual's predisposition to anthropomorphize chatbots). Given this ambivalence, scholars should take a more differentiated perspective on the social perception of chatbots. Research frequently neglects the complexity and multidimensionality of many interpersonal concepts and social heuristics, or it adopts them one-to-one to chatbot interactions without considering that artificial versions of social or psychological concepts might deviate from their interpersonal equivalents. In other words, perceiving a sense of empathy and warmth could be equally important in interactions with chatbots, however, an appropriate conceptualization of artificial empathy might differ from

interpersonal empathy. Scientific research has just begun to take a more nuanced perspective on the determinants and moderators of the social perception of chatbots (e.g., Crolic et al., 2022; Han et al., 2023), and this thesis takes a significant step forward in advancing our understanding in this emerging field of research. To conclude, this thesis found that "Social Response Theory" and anthropomorphism are crucial to understand the nature of human-bot interactions, however, it also highlights their limits and the non-social nature of chatbots.

Computer-like schemas and mechanistic stereotypes. Unlike the majority of previous research, the present thesis did not focus on identifying potential negative drawbacks caused by perceiving chatbots as social entities (e.g., mind attribution or social presence; Holthöwer & van Doorn, 2022; Pitardi et al., 2022). Instead, it continuously examined to what extent bots are perceived as technology rather than as social actors as there was ambivalent and limited evidence for "Social Response Theory" and anthropomorphism in Paper 1. Therefore, it refers to theories on schemas and mental models (Fiske & Linville, 1980; Kroeber-Riel & Gröppel-Klein, 2019; Rouse & Morris, 1986), arguing that the cognitive structures and the knowledge we possess about chatbots differ from the schemas we have about humans. Congruently, results across papers showed that humans have unique stereotypical associations with chatbots that shape their expectations towards the agent's attributes, behavior, and performance. These stereotypes include a chatbot to be objective, data-driven, unemotional, and incapable of moral judgments or believes. This aligns with "Mind Perception Theory" and related research positing that humans attribute a moderate level of agency and cognition but barely experiential capabilities to bots (Gray et al., 2007; Waytz & Norton, 2014). This could explain the cognitive nature of the trust-building process found in Paper 1, i.e., humans seek for good arguments and quality indices to evaluate a healthcare chatbot's trustworthiness. In contrast, interpersonal trust concepts like integrity or benevolence had a secondary role given that chatbots are poorly associated with emotion, a free will, or moral beliefs. Similarly,

empathetic expressions that require experiential capabilities were found to be perceived ungenuine as they conflict with mechanistic stereotypes towards chatbots. Hence, when examining the social nature of chatbots, researchers should not only consider the consequences of perceiving a mind in a bot (e.g., Gray & Wegner, 2012; Lee et al., 2020; Stein & Ohler, 2017) but also the implications and consequences of *not* attributing complex mindfulness to chatbots (Pitardi et al., 2022).

Another stereotype towards chatbots that does not refer to their lack of mindfulness is related to their speed. Unlike human agents, who have limited cognitive capabilities and might need to allocate resources to solve a request, data-driven chatbots are expected to be consistently available and capable of immediately addressing requests (Meng & Dai, 2021; Schanke et al., 2021; Yu et al., 2022). By demonstrating the adverse effects of response delays, which are considered normal in interpersonal chats, this thesis found further evidence that humans might enter chatbot interactions with different schemas and expectations. This assumption was substantiated by experimental manipulations and moderation analyses showing that the perception of social cues depends on whether humans enter a chatbot conversation applying computer- vs. human-like schemas. Also, it has been shown that the utilitarian value a chatbot provides is strongly associated with its evaluation, further emphasizing its technical nature and the relevance of considering technology acceptance models (e.g., Davis, 1989; Venkatesh et al., 2012). Subsuming, all papers in this thesis provide qualitative and quantitative evidence that humans tend to apply computer-like schemas including corresponding expectations to their interactions with chatbots.

Introducing the concept of authenticity to human-bot interactions. The previous sections described the thesis's findings and contributions on mechanistic stereotypes humans might have towards chatbots. In diving deeper into the potential consequences for humanization and the perception of social cues which interfere with these stereotypes, Paper 2 introduces the

concept of "authenticity" to human-bot interaction. The concept of authenticity is well researched in domains like psychology (e.g., Wood et al., 2008) or marketing (e.g., Morhart et al., 2015), however, it has not yet been examined in the perception of bots. Authenticity as a trait defines the extent to which an individual or an object is perceived to be genuine, original, and true to its own nature (Heidegger, 1996; Wood et al., 2008). Paper 2 shows that a healthcare chatbot pretending to feel *with* or *for* a human is perceived inauthentic having a negative effect on trust and using intentions. This finding contributes to existing knowledge in two ways: first, it provides empirical evidence that social cues can backfire when they appear not credible and fake. Introducing this new underlying mechanism might help researchers in explaining null findings or adverse effects of human-like design elements in chatbots. And second, it also shows that interpersonal concepts like the perceived authenticity or credibility are important in the evaluation of a chatbot's trustworthiness. This finding again highlights that basic concepts of interpersonal interactions are applicable to our perception of chatbot interactions, however, what is perceived (in)authentic is different for chatbots vs. humans. The next paragraph will further elaborate on this aspect.

Boundary conditions and limits for humanization. This last paragraph on theoretical contributions can be considered an intersection between theory and practice. In general, the findings of the thesis suggest that equipping chatbots with anthropomorphic design elements is beneficial only up to the point where humans can successfully meet their basic need for social connectedness. Any cues going beyond this point might become excessive eliciting feelings of annoyance, inauthenticity, or even creepiness. While initially positively perceived on a subconscious level, if a bot's pretense becomes apparent or its social cues starkly contradict mechanistic stereotypes, attempts at humanization can backfire. Humans enter conversations with chatbots applying computer-like schemas and a violation with associated expectations might have consequences on the chatbot's perception and evaluation. This aligns with

"Expectancy Violations Theory" (EVT; Burgoon, 1993) which has its origin in communication studies and argues that a violation of a priori expectations towards a communicator's behavior creates arousal and cognitive processing. This might result in a negative evaluation of the communication partner (i.e., the chatbot) in case that expectations have not been met. While a slight deviation from expectations can yield favorable consequences (e.g., when a chatbot adheres to politeness norms), an excessive and too obvious deviation is prone to result in backfiring effects. This argument is also substantiated by research from marketing which shows that a slight schema incongruency can elicit attention and positive attitudes while a too strong incongruency has adverse effects (Kroeber-Riel & Gröppel-Klein, 2019).

Given that users tend to enter chatbot interactions with computer-like schemas, it is vital to consider technology acceptance models in designing (social) chatbots. Congruently, there is evidence in this thesis that the fulfillment of central dimensions of technology acceptance like perceived usefulness is more important than human-likeness regarding chatbot or service provider evaluation. Also, there is only very limited evidence across the ten studies that social cues have any significant direct positive effect on relevant outcome dimensions like trust or using intentions. The knowledge about the technical nature of chatbots and its perception as a software tool might thus attenuate the positive effect of humanization. Alternatively, these and the null findings from related research could be attributed to a wear-out effect (Croes & Antheunis, 2021). The increasing exposure to chatbots enhance familiarity and the accuracy of schemas. While social cues might facilitate positive attitudes in initial conversations, the effect could be attenuated the more a user is interacting with chatbots. Also, the boundaries and limits for humanization depend on individual characteristics, i.e., someone's tendency to anthropomorphize. Users who apply (do not apply) human-like schemas to their chatbot interactions might show more (less) favorable reactions to social cues as they expect (do not

expect) chatbots to act human-like. Also, these users might be less (more) prone to perceive social cues annoying or inauthentic.

3.3 Managerial Implications

The theoretical findings and contributions of this thesis provide some potentially insightful and promising implications for practitioners.

Functionality beats human-likeness. The three papers barely found evidence for positive effects of social cues on relevant outcome dimensions like trust, using intentions, or service provider evaluation. Neither did the participants explicitly express a desire for a chatbot to be more human-like (Paper 1), nor did this thesis find experimental evidence for direct positive effects of social cues on the mentioned outcomes. Although expressions of empathy were found to enhance trusting intentions indirectly through perceived warmth, the effect was too small to produce a significant main effect. Additionally, the effect was attenuated by inauthenticity perceptions (Paper 2). Also, there were no considerable positive effects for other verbal, visual, and personifying social cues, e.g., a cartoon-like or photo-realistic avatar and giving the chatbot a name (Papers 2 and 3). Instead, a chatbot's functionality (e.g., its reliability, underlying data base, outcomes, and perceived usefulness) was consistently identified as the primary driver for a positive evaluation. These findings suggest that managers and software designers should ensure a chatbot's capability to enhance efficiency and outcomes before optimizing its human-likeness. In this regard, managers should particularly refrain from utilizing humanization for compensating performance shortcomings. A highly human-like appearance can enhance efficacy expectations that can result in even more frustration when the bot fails (Crolic et al., 2022). A well-balanced calibration between a chatbot's appearance and its actual performance is therefore crucial, i.e., a simple bot having a high risk to fail should not be overly humanized. Managers should also be aware that a poorly performing chatbot that takes the role of a service representative can harm the service experience resulting in low

customer satisfaction and company evaluations. However, it is questionable in how far companies will develop their own chatbot structures in future. Given the enormous power and high adaptability of large language models such as "ChatGPT" or "Aleph Alpha", it is conceivable that companies might opt to leverage such well-established systems rather than developing their own chatbots. In this case, a general standard for the technical performance of chatbots could be ensured, prompting companies to invest their resources in customizing and potentially humanizing their chatbots.

Social cues should be selected carefully. The thesis also found that not all social cues are equally promising and effective in enhancing a chatbot's human-likeness and facilitating using intentions. Social cues that clearly conflict with schemas of a chatbot (e.g., emotional expressions or needing time to respond) might be interpreted as gimmicks, ungenuine, implausible, or even annoying. Managers and software designers are encouraged to consider their customers' expectations towards the chatbot and align the social cues accordingly. This might require a critical thinking outside the box and a deep analysis on how a sense of artificial humanness can be created without risking backfiring effects. For instance, software designers are advised to adapt social cues and interpersonal concepts to the characteristics and stereotypes associated with chatbots, rather than simply transferring them without any adjustments. Taking the example of empathy, it might be more appropriate to design a computer-like version in which the chatbot communicates its intend to provide instrumental support ("I am here to help you in solving your service request") instead of emotional support ("I am so sorry that you had problems with your service request").

Also, practitioners should evaluate the appropriateness of specific social cues on the individual or target group level. For example, social cues like humor in a service chatbot that is designated for counseling young customers on hedonic products may yield favorable outcomes. In this scenario, these social cues may align with both the products and the target

group, transparently conveying to users that the bot is deliberately adopting its behavior to harmonize with the specific nature of the service context. In contrast, incorporating schema incongruent social cues to chatbots utilized for computer-like tasks or employed for sensitive services (e.g., banking) could have adverse effects since they might appear inappropriate. Practitioners should also consider the characteristics of their target groups in designing social chatbots. For instance, a human-like chatbot design might be helpful and more effective for target groups who are not familiar with bots or who tend to anthropomorphize them.

More is not always better. To profit from potential positive effects of humanization, it might not be necessary to maximize a chatbot's human-likeness. This thesis did not find evidence that adding further social cues (e.g., response delays in addition to other verbal and visual social cues) enhances the perceived human-likeness of a chatbot (Paper 3). Instead, as discussed before, exceeding the point of appropriate human-likeness can result in adverse effects. Incorporating chatbots with a few distinctive, apprehensible, and appropriate social cues may suffice to establish a sense of humanness and social presence. Furthermore, it is advisable that a chatbot transparently discloses its technical nature and the associated limitations. This could help to foster trust by presenting the chatbot as transparent and honest (Paper 1), and to cultivate realistic a priori expectations towards its capabilities and communicative skills. In case of service failures, such transparency may contribute to prevent customer dissatisfaction and negative company evaluations (Crolic et al., 2022; Mozafari et al., 2022). Combining moderate and appropriate social cues with a disclosing statement could offer a promising compromise in elevating a chatbot's human-like while mitigating the risk of adverse effects.

3.4 Limitations and Future Research Directions

Like any other theses and papers, this dissertation can only explore a limited, carefully selected facet of an expansive research domain, adding a few mosaic pieces to the grand

tapestry of knowledge in the field of human-bot interaction. Also, it has some limitations, boundaries, and shortcomings that are to be reflected critically. Both the limited perspective and potential shortcomings provide promising avenues for future research. The discussion of these limitations and future research directions starts with an internal perspective before taking the greater picture into account.

Methodological shortcomings. One prominent limitation of this thesis is the missing field evidence. All studies in this thesis were based on experiments or surveys conducted in the lab or online decreasing external validity. However, this limitation applies to the majority of research in this field given that chatbots have only emerged within the last years. Many studies even use only hypothetical scenarios or screenshot and video vignettes instead of real interactions (Castelo et al., 2023). In contrast, seven out of ten studies in this thesis either used real chatbots provided by a medical service provider (Paper 1) or chatbots specifically programmed for the purpose of the studies (Papers 2 and 3). Across papers and studies, this thesis tried to enhance the scenarios' realism, participants' engagement, and external validity. Nevertheless, there is no doubt about the necessity of field evidence in the domain of human-bot interaction to substantiate the experimental findings with real data. Another related factor reducing the external validity of the presented findings and similar research is that participants in most experiments are exposed to an interaction with a fictitious bot from a fictitious and unknown service provider. It is thus unclear to what extent the findings apply to interactions with chatbots from potentially well-known brands, or to what degree the perception of a bot changes with repeated interactions. There is some evidence in related research finding that the social perception of chatbots and voice assistants and the respective consequences are influenced by the relationship length (Croes & Antheunis, 2021; Uysal et al., 2022). It could be insightful to conduct future field or experimental studies using bots from well-established brands as the evaluation of bots also depend on their providers, e.g., regarding the evaluation

of their trustworthiness (Paper 1). Considering potential spill-over effects from the provider to the bot, it could also be promising to examine the perception of a personality match between a brand and its bot. A high fit might protect the chatbot from inauthenticity perceptions, even if it shows uniquely human-like characteristics. Lastly, the chatbots in the presented studies have not been very mindful or complex. The medical chatbots utilized in Paper 1 have been prototypes and the ones in Papers 2 and 3 followed quite simple decision trees. Also considering that individuals have been aware to participate in studies on chatbot perception, the salience of their technical nature might have been more prominent compared to when customers interact with a sophisticated chatbot in a real service situation. Furthermore, the social cues utilized in the studies were relatively simple visual, verbal, and non-verbal cues. In the context of anthropomorphism theory, these shortcomings might have made computer-like schemas more accessible thus diminishing anthropomorphism, particularly for users who have already possessed considerable knowledge about chatbots (Epley et al., 2007). Future research is needed to examine the generalizability of the present findings in situations where chatbots provide an overall more human-like service experience, especially considering the rapid technological advancements.

Ambivalences in this thesis. Considering the greater picture across papers, there is some ambiguity in the findings from Papers 2 and 3. Paper 3 argues that users might be more likely to apply human-like schemas to a chatbot interaction when the chatbot performs a human-like service task thus expecting it to communicate or act like a human. Indeed, the paper finds that the backfiring effect of response delays is attenuated when the bot performs a medical assessment (human-like) vs. train ticket booking (computer-like). However, the chatbots in Paper 2 performed a quite similar medical assessment but still their expressions of empathy were considered inauthentic. This provides evidence that although the task was human-like, humans applied computer-like schemas making experiential expressions feel not real in

chatbots. I outline two potential explanations which might approach this ambivalence: first, it could be diverging service efficiency expectations in computer- vs. human-like tasks rather than a general application of human-like schemas which account for the moderating effect found in Paper 3. And second, the negative effect of experiential expressions on perceived authenticity could be even larger in computer-like tasks, i.e., the effect might also be moderated by the service task's human-likeness. As there is no empirical data, I can only speculate on this. To shed light into the role of human-like schemas in authenticity perceptions, future research could combine the findings from Papers 2 and 3 and examine if an individual's tendency to anthropomorphize moderates the effect of experiential expressions (or other human-unique attributes) on perceived authenticity. Given that authenticity in bots has barely been studied yet, scholars are encouraged to examine its role in more detail. For instance, previous research considered a bot's authenticity by the conversation's perceived naturalness and realness (Morrissey & Kirakowski, 2013; Wünderlich & Paluch, 2017). Separating the conversation's authenticity (i.e., the extent to which a chatbot conversation has a natural flow) from the chatbot's authenticity (i.e., the extent the chatbot's appearance is in congruence with its technical nature) could help in understanding how to improve the human-likeness of a chatbot interaction without incorporating fake appearing social cues.

Generalizability of the findings to other interaction contexts. In interpreting the present findings, it is important to consider that all studies have been conducted using text-based and outcome-oriented service chatbots. However, a diverse array of bots (e.g., physical robots, virtual avatars, or voice assistants) extends beyond singular purposes, actively accompanying users in their everyday lives. For instance, Apple's "Siri" exemplifies a voice-based personal assistant capable of adapting to users and executing a wide spectrum of tasks. It is hence more likely that users develop a relationship with such assistants and perceive them different from simple service chatbots. The chatbot "Replika" is an even better example, as it

is explicitly intended to foster friendships or romantic connections with its users (Pentina et al., 2023; Skjuve et al., 2021). Indeed, many users report to feel bonded and emotionally connected to their "Replika" supporting "Social Response Theory" (Nass & Moon, 2000) and anthropomorphism (Epley et al., 2007). Emotional expressions or mimicking other human-unique behavior could feel more authentic when users have already accepted the bot as a social companion or even integrated it to their self-concept. The results presented in this paper might thus only apply to simple service chatbots, but not to sophisticated personal assistants. Also, it is questionable whether the present findings are applicable to interactions with physical robots, three-dimensional avatars in virtual reality settings, or voice assistants. All these bots enable a more immersive humanization since they are capable of movement and/or speech. The implementation of human-unique attributes may be more appropriate for bots that are inherently predisposed to activate human-like schemas. There is supporting evidence as empathy was found to be more effective when the bot had a voice (vs. text-only) (Lv et al., 2022). Moreover, studies revealed that humans tend to anthropomorphize robots more when their verbal expressions were complemented by congruent gestures (Salem et al., 2013). Finally, as emphasized in the introduction and the theoretical foundation, researchers posit that text-based chatbots lack sufficient human-likeness to fall into the "Uncanny Valley" (Skjuve et al., 2019). The findings from this thesis should hence always be interpreted within the specific context in which the studies were conducted.

Examining anthropomorphic design elements and their effects more granularly. The discussion on what it means to be human is a philosophical debate potentially filling dozens of doctoral theses. Without further elaboration, the papers included here could only delve into specific interpersonal concepts and social cues – namely trust, warmth, empathy, response delays, and some minor cues that held a subordinate position. In contrast, the opportunities for humanization and their potential outcomes seem to be virtually limitless. Some researchers

provide taxonomies for social cues in conversational agents (e.g., Feine et al., 2019), while others systematized and examined the outcomes of humanization and anthropomorphism in meta-analyses (e.g., Blut et al., 2021; Yanxia et al., 2023). However, there is still a high need for further research examining which social cues have which kind of consequences under which circumstances and for which kind of target group. Creating a more granular perspective on the chances and risks of humanizing bots could help both theorists and practitioners in understanding better their (non-)social nature. I therefore encourage researchers to conduct further systematic literature reviews and meta-analyses considering the multidimensionality of interpersonal concepts and social cues (e.g., emotional intelligence or visual vs. auditive cues), their effects on relevant mediators (e.g., relational aspects or perceived mindfulness), central outcome dimensions (e.g., consumer well-being or using intentions), and potential moderators (e.g., user demographics or service context). This implies that future studies should account for the complexity and potential interactions of all these factors to have a sufficient data base. Also, while there is a general meta-analysis on anthropomorphism in bots and AI (Blut et al., 2021), there is – to the best of my knowledge – no systematic overview on potentially negative consequences that might result from humanization and anthropomorphism.

Regarding the present findings in particular, future research should identify further social cues that are considered usual and distinctive in interpersonal interactions, but inauthentic or annoying in bots. Also, social cues that feel fake or that exceed an appropriate level of human-likeness could be interpreted as a persuasion technique or the intent of a company to convince customers of buying a specific product (Gröppel-Klein et al., 2018; Seitz & Bekmeier-Feuerhahn, 2023a). This could particularly hold true for individuals being skeptical towards new technologies or holding negative attitudes towards robots and chatbots.

Technological advancements and environmental changes. The rapid technological advancements in AI and chatbots within recent years constitute a potentially profound factor

guiding future research directions. With the launch of "ChatGPT" in 2022, the chatbot landscape underwent a significant disruption, as chatbots suddenly gained the ability to accomplish tasks that were barely imaginable at the inception of this thesis and the conceptualization of the three papers. The role of AI and chatbots in society has changed ever since and it is hard to predict the potential advancements for the next years. According to the "Hype Cycle for Artificial Intelligence" which is regularly published by the market research and consulting firm "Gartner", we are currently approaching the peak of inflated expectations towards generative AI, AGI, and smart robots (Perri, 2023). However, the graph also predicts that it will take at least five to ten years until these technologies reach the plateau of productivity (see Figure 5). It is hence likely that we will witness further profound disruptions and the emergence of novel application areas for these technologies in the foreseeable future.

In light of this dynamic and highly innovative environment, the potential capabilities of future chatbots could greatly surpass those of current systems. I will illustrate the significance of these developments for the creation of academic knowledge and theorizing by providing two examples: first, Paper 1 published in September 2022 stated that "it is foreseeable that future technologies will be more sophisticated and have capabilities that cannot be anticipated today" (p. 11). Two months after the publication date, "ChatGPT" was introduced. Second, Skjuve et al. (2019) posited that "text-based chatbots still have a long way to go before they become sufficiently humanlike for an uncanny effect to be relevant". With the emergence of generative AI and AGI, this point no longer seems too distant.

The two examples demonstrate the rapid obsolescence that research on chatbots and human-bot interaction in general may face. This could also change the schemas humans have about chatbots dramatically. I will elaborate two potential paths and their implications for humanizing bots: in the first scenario, human's schemas about bots will become more human-

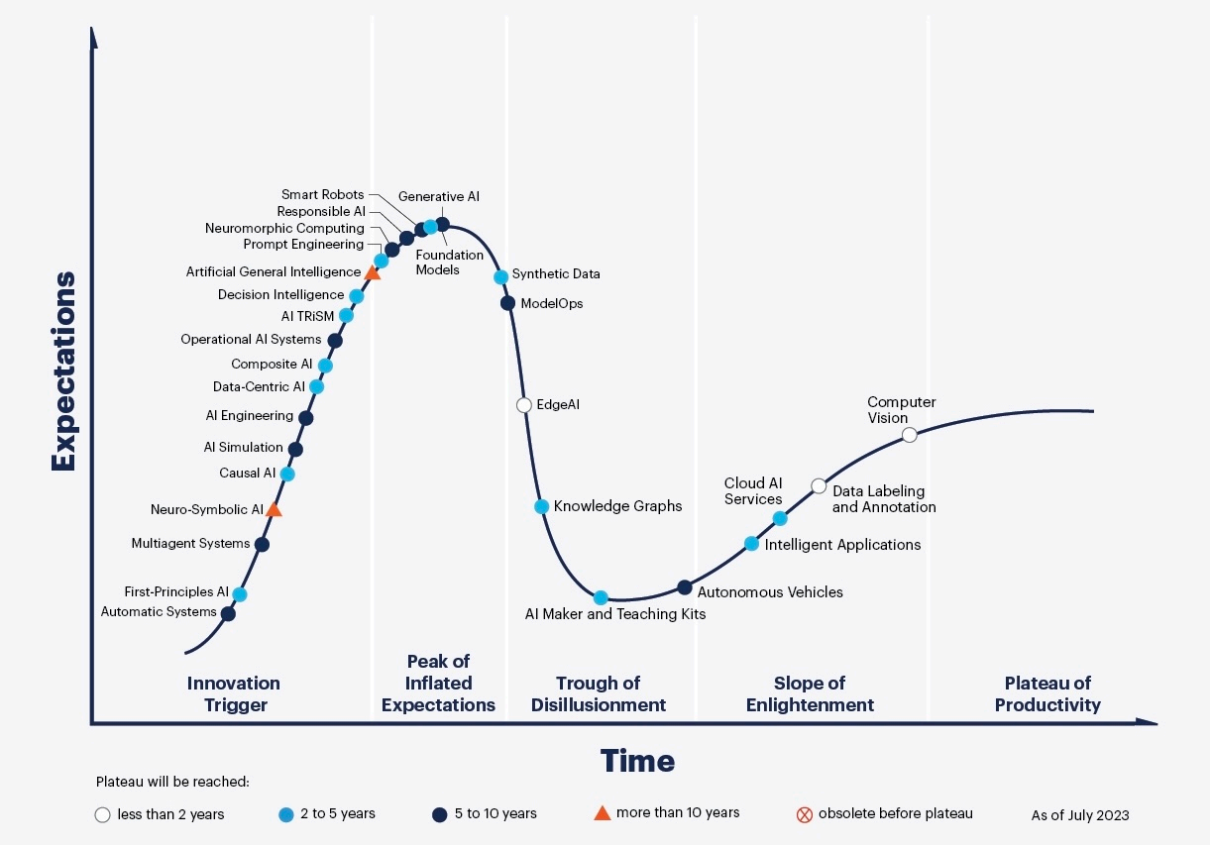


Figure 5. The "Hype Cycle for Artificial Intelligence". (Perri, 2023)

like. Given the increasing performance of AI in accomplishing complex tasks and in mimicking human behavior, the attribution of specific capabilities that have been considered uniquely human could expand to bots. This may include mind perception which would have a fundamental impact on future research and the applicability of the present findings. Also, the (perceived) efficacy of chatbots in comparison to humans in performing specific tasks could undergo dramatic changes. While chatbots were previously regarded less capable or flexible in handling requests (Crolic et al., 2022; Yu et al., 2022), advanced AI-driven systems might increasingly outperform humans in various tasks (Huang & Rust, 2018). And lastly, future generations will potentially grow-up having daily interactions with highly capable bots from birth. Assuming that these interactions will become nearly indistinguishable from interhuman interactions, future generations may develop more human-like schemas about bots. All these factors might move chatbots closer to humans facilitating their perception as social actors. In

the second scenario, however, human's schemas about bots will become more computer-like. The "Hype Cycle for Artificial Intelligence" (see Figure 5) suggests that we are close to the peak of inflated expectations towards generative AI, AGI, and smart robots. This hype might also foster unrealistic expectations regarding the potential of bots to become very close to human-like. With deflating expectations, the technical nature and potential limitations of bots may become more salient creating the awareness that even the most advanced bots are still algorithms. Also, humans might become more experienced and knowledgeable about bots given their increasing prevalence. As a result, schemas could become more accurate (Gambino et al., 2020; Gnewuch et al., 2022; Rouse & Morris, 1986). Social responses to computers and the tendency to anthropomorphize them can be considered cognitive biases that are more likely to occur when users have only little knowledge about them (Epley et al., 2007; Nass & Moon, 2000). All these arguments, on the other hand, suggest that chatbots may be perceived as technology rather than social actors in future. As I can only speculate on these developments, future research is needed on how the social perception of chatbots may change over time. In doing so, scholars could focus on conducting longitudinal studies on the individual level (i.e., within-subject designs) or the societal level. Also, future meta-analyses on the social perception of bots could include the publication date of a study as a moderator to account for potential time effects (Blut et al., 2021).

3.5 Ethical Considerations

The emergence of powerful AI, chatbots, and robots is believed to hold the potential to fundamentally transform social interactions, the workplace, service delivery, teaching, and many other aspects of life (Araujo, 2018; Huang & Rust, 2018; Larivière et al., 2017; Noble et al., 2022; Peres et al., 2023). Anticipating this huge impact and the barely predictable developments, the physicist Stephen Hawking stated that AI will be "either the best, or the worst thing, ever to happen to humanity" (Hern, 2016). This requires a thoughtful examination

of the ethical dilemmas associated with the use of such technologies. However, as delving into the ethical dimensions of AI-induced challenges is a distinct philosophical and highly extensive debate, this thesis will focus on the ethical issues that are associated with (overly) humanizing bots.[3]

Despite the illustrated backfiring effects and potential limits, humanizing chatbots is accompanied by several ethical challenges as well. First, with AI and bots becoming more capable and human-like, it is to be evaluated what it really means to be human. "Mind Perception Theory" (Gray et al., 2007) and anthropomorphism research (Waytz & Norton, 2014) contend that the true essence of humanity, distinguishing humans from machines, lies in their minds, particularly in their capacity to experience emotions. However, while bots are not expected to possess genuine experiential capabilities soon (Wirtz et al., 2018), they could potentially simulate agency and experiences through advanced AI, including sophisticated NLP and emotion recognition software (Miner et al., 2016). In the future, bots could easily pass the "Turing test", as demonstrated by the voice assistant "Google Duplex". It showcased its capability to mimic a human in phone calls, leaving the conversation partner unaware of its technical nature (Leviathan & Matias, 2018). To avoid confusion and misuse, even highly human-like bots should disclose their true nature at the beginning of the conversation. The state of California can be considered a pioneer in this domain as the 2019 introduced "Bolstering Online Transparency Act" requires bots to disclose their identity to humans (Stricke, 2020). Regardless of the legal context, chatbot disclosure ensures transparency and fosters realistic expectations towards the interaction partner (Crolic et al., 2022; Mozafari et al., 2022). This is also important considering that humans tend to exhibit greater trust in human-like agents because of higher competence and mind attribution (Waytz et al., 2014). To prevent damage and harm, humanization should not be employed as a strategy to compensate for technical

[3] For an overview on ethical guidelines on AI in general see Hagendorff (2020).

shortcomings, i.e., the degree of human-likeness should align with the bot's actual capabilities to avoid eliciting inappropriately high levels of trust (Glikson & Woolley, 2020). Furthermore, humanization should not be utilized to manipulate users in an unfavorable sense. For instance, certain anthropomorphic design elements could potentially be utilized to exploit cognitive biases or encourage behavior that is disadvantageous for the user (e.g., persuading customers to purchase unnecessary products or dissuading subscription cancellations). In contrast, such elements could be leveraged to cultivate beneficial behaviors and enhance customer well-being (e.g., encouraging the purchase of sustainable products or motivating health-promoting behavior).

Beyond the issues of misleading and manipulating humans, there is an additional ethical concern regarding the potential harm to social relations and society. Interactions with highly human-like bots might jeopardize interpersonal relationships, particularly for vulnerable groups who face social anxiety. Companion chatbots like "Replika" that build up friendships or fall in love with humans are no longer science fiction but reality (Pentina et al., 2023; Skjuve et al., 2021). Rather than binding users tightly and encouraging frequent interaction, the potential of such systems could be leveraged to alleviate the fears of socially isolated individuals. Especially chatbots can offer a secure and non-judgmental space for practicing social interactions (Olson, 2018). The danger of social withdrawal not only affects vulnerable groups but also all the other potential users. AI-based chatbots can adapt perfectly to users and their needs through the collection of personal data and powerful algorithms. This poses the risk for chatbots to become superior and more enjoyable interaction companions. For instance, they might please and confirm their users rather than asking unpleasant questions and risking conflicts. Furthermore, interactions with submissive and user-pleasing bots could result in a brutalization of language and interpersonal treatment. Bots will adapt to their users uncritically and follow their commands, regardless of the tone of the language. To mitigate potential threats, Amazon

introduced the "Magic Word" feature to their smart speakers that offers positive reinforcement when kids use the word "please" when asking questions (Amazon, 2021).

When humanizing bots, managers and software designers must choose specific anthropomorphic design elements (e.g., the avatar, the name, or the voice). This choice might be influenced by stereotypes and social biases, i.e., practitioners try to choose the most fitting social cues to maximize naturalness and realism. For instance, the default voice of personal assistants like Amazon's "Alexa" and Apple's "Siri" is set to a female tone potentially since women are stronger associated with service-oriented roles. In light of the previously outlined submissiveness and servility of bots, there is a risk of reinforcing and amplifying gender stereotypes (West et al., 2019). However, social biases do not only affect software designers in selecting social cues. In addition, users are susceptible to apply them in their interactions as well (Fossa & Sucameli, 2022; Nass & Moon, 2000). Continuing with the example of gender stereotypes, the perception of a "male" or "female" social cue in an agent can trigger associated stereotypes, even when the cue appears subtle (e.g., the color of the lips or fashion accessories). Research has shown that user responses to a bot with a female gender have been more positive when operating in domains where perceived warmth is crucial (i.e., healthcare; Borau et al., 2021) but worse when it assumed a role strongly associated with males (i.e., mechanic; McDonnell & Baxter, 2019). Although these findings and conventional wisdom might motivate to align the chatbot design with prevalent stereotypes, practitioners should consider the risk of amplifying gender biases. Instead, bots could be intentionally designed to counteract and dismantle these stereotypes, such as employing a strong, female product recommendation chatbot on a do-it-yourself store's website. Furthermore, there is a risk that female bots, in particular, may be portrayed in a sexist manner. For instance, "Replika" is occasionally promoted on social media with images featuring revealing attire, thereby contributing to the perpetuation of sexism (see Figure 6).

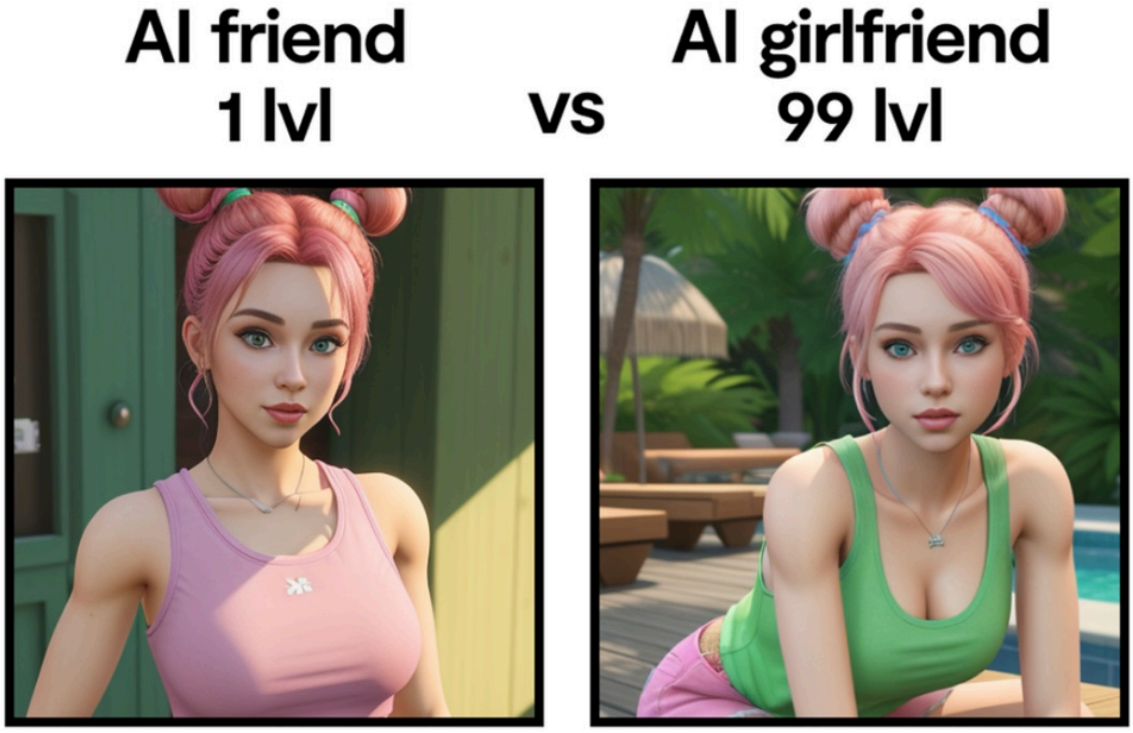


Figure 6. Social media advertisement of "Replika".

Note: Original source is cited in the image sources section.

Given that the algorithms of bots are crafted by human programmers, various biases and stereotypes beyond gender find their way into the development process. Even though commonly associated with objectivity and rationality (see Paper 1), bots and algorithms inherently mirror the values and biases present in society. AI-based algorithms which autonomously learn from their users are even more prone to adopting biases. A striking example is Microsoft's Twitter chatbot "Tay" which, within a single day, adopted racist and antisemitic statements from users, leading to its immediate removal (Beuth, 2016). To mitigate the adoption of biases and avert potential risks, it is essential to a) cultivate diverse development teams and b) incorporate security mechanisms within algorithms which can identify offensive content and prevent inappropriate responses by the bot.

Finally, many companies may humanize their bots with the intention to mimic interpersonal conversations as closely as possible or even to leave users unaware of the bot's

technical nature. Even though this strategy can facilitate using intentions and the transition to automated services, there is also the risk of undermining the distinctive role that humans play in service delivery. For instance, many researchers argue that bots should always be seen as a complement to human employees, especially in services that require true empathy and social relatedness (e.g., Huang & Rust, 2018; Powell, 2019; Waizenegger, 2020). Hence, bots should rather take over routine or analytical tasks providing humans more time for the fulfillment of relational tasks and caretaking. Regarding the utilization of healthcare chatbots, which have been employed in at least one study in all papers of this thesis, researchers assert that utilizing these bots for medical assessments comes with huge ethical challenges (Brown & Halpern, 2021; Parviainen & Rantala, 2022). For instance, given that healthcare chatbots are frequently used for self-diagnosis, a safe and successful use implies that the patient (1) provides all necessary information to the chatbot and (2) can understand the diagnosis and information provided by the chatbot. Also, patients might question a physician's diagnosis if it differs from the assessment of a chatbot, as the latter may be considered non-biased and objective (Paper 1). Conversely, physicians may also lean toward relying on a chatbot's recommendations for the same reasons which is called "automation bias" (Sujan et al., 2019). And ultimately, while physicians are mindful individuals with moral beliefs and values, it remains obscure to what extent chatbots can ever be aware of the potential consequences of their behavior or capable of taking responsibility for their actions (Paper 1). Given these ethical challenges, chatbots should be considered tools that support and collaborate with humans rather than substituting them.

3.6 Concluding Remarks

Given the increasing prevalence of chatbots mimicking humans, this thesis examined whether and to what extent chatbots are perceived as social actors. While existing theories posit that interactions with computers and bots are inherently social, there is ambivalence on the positive and negative consequences of employing anthropomorphic design elements to

chatbots. Specifically, there is limited research on the consequences of *not* perceiving chatbots as social actors and their implications for anthropomorphic design. Addressing this research gap, three empirical papers have found both qualitative and quantitative evidence suggesting that humans may enter chatbot interactions with computer-like, rather than human-like schemas, which can elicit unique expectations towards a chatbot's behavior. A violation of these expectations can result in adverse effects, i.e., when an anthropomorphic design element clearly contradicts computer-like schemas. However, if and to what extent humans apply computer- vs. human-like schemas may depend on individual or context-related factors. To sum up, this thesis contributes to a more nuanced perspective on the limits and boundaries of humanizing chatbots. Future research and a continuous evaluation of the applicability of the present and previous findings is necessary considering the rapid technological advancements and environmental changes. Regardless of future developments, the opportunities of humanizing bots should always be evaluated with consideration of ethical aspects and human uniqueness. Chatbots should serve the ultimate purpose of assisting humans, not replacing, or threatening them.

References

Adamopoulou, E., & Moussiades, L. (2020). An overview of chatbot technology. In I. Maglogiannis, L. Iliadis, & E. Pimenidis (Eds.), *Artificial intelligence applications and innovations* (Vol. 584, pp. 373–383). Springer. https://doi.org/10.1007/978-3-030-49186-4_31

Aggarwal, P., & McGill, A. L. (2007). Is that car smiling at me? Schema congruity as a basis for evaluating anthropomorphized products. *Journal of Consumer Research*, *34*(4), 468–479. https://doi.org/10.1086/518544

Amazon (2021, June 30). *Amazon announces all-new Alexa experiences built for kids in the UK*. https://www.aboutamazon.co.uk/news/innovation/amazon-announces-all-new-alexa-experiences-built-for-kids-in-the-uk [Retrieved December 11, 2023]

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

Barney, C., Hancock, T., Esmark Jones, C. L., Kazandjian, B., & Collier, J. E. (2022). Ideally human-ish: How anthropomorphized do you have to be in shopper-facing retail technology? *Journal of Retailing*, *98*(4), 685–705. https://doi.org/10.1016/j.jretai.2022.04.001

Bartneck, C., Kanda, T., Ishiguro, H., & Hagita, N. (2007). Is the uncanny valley an uncanny cliff? *RO-MAN 2007—The 16th IEEE International Symposium on Robot and Human Interactive Communication*, South Korea, 368–373. https://doi.org/10.1109/ROMAN.2007.4415111

Benbasat, I., & Wang, W. (2005). Trust in and adoption of online recommendation agents. *Journal of the Association for Information Systems*, *6*(3), 72–101. https://doi.org/10.17705/1jais.00065

Beuth, P. (2016, March 24). *Twitter-Nutzer machen Chatbot zur Rassistin*. Zeit Online. https://www.zeit.de/digital/internet/2016-03/microsoft-tay-chatbot-twitter-rassistisch [Retrieved December 12, 2023]

Blut, M., Wang, C., & Schoefer, K. (2016). Factors influencing the acceptance of self-service technologies: A meta-analysis. *Journal of Service Research*, *19*(4), 396–416. https://doi.org/10.1177/1094670516662352

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

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

Briggs, J., & Kodnani, D. (2023, March 26). *The potentially large effects of artificial intelligence on economic growth*. Goldman Sachs. https://www.key4biz.it/wp-content/uploads/2023/03/Global-Economics-Analyst_-The-Potentially-Large-Effects-of-Artificial-Intelligence-on-Economic-Growth-Briggs_Kodnani.pdf [Retrieved December 14, 2023]

Brown, J. E. H., & Halpern, J. (2021). AI chatbots cannot replace human interactions in the pursuit of more inclusive mental healthcare. *SSM—Mental Health*, *1*, Article 100017. https://doi.org/10.1016/j.ssmmh.2021.100017

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., & Zhang, Y. (2023). *Sparks of artificial general intelligence: Early experiments with GPT-4*. arXiv. https://doi.org/10.48550/arXiv.2303.12712

Buchanan, A. (1988). Principal/agent theory and decision making in health care. *Bioethics*, *2*(4), 317–333. https://doi.org/10.1111/j.1467-8519.1988.tb00057.x

Burgoon, J. K. (1993). Interpersonal expectations, expectancy violations, and emotional communication. *Journal of Language and Social Psychology*, *12*(1–2), 30–48. https://doi.org/10.1177/0261927X93121003

Burleigh, T. J., Schoenherr, J. R., & Lacroix, G. L. (2013). Does the uncanny valley exist? An empirical test of the relationship between eeriness and the human likeness of digitally created faces. *Computers in Human Behavior*, *29*(3), 759–771. https://doi.org/10.1016/j.chb.2012.11.021

Castelo, N., Boegershausen, J., Hildebrand, C., & Henkel, A. P. (2023). Understanding and improving consumer reactions to service bots. *Journal of Consumer Research*, *50*(4), 848–863. https://doi.org/10.1093/jcr/ucad023

Cha, Y.-J., Baek, S., Ahn, G., Lee, H., Lee, B., Shin, J., & Jang, D. (2020). Compensating for the loss of human distinctiveness: The use of social creativity under human–machine comparisons. *Computers in Human Behavior*, *103*, 80–90. https://doi.org/10.1016/j.chb.2019.08.027

Chandler, J., & Schwarz, N. (2010). Use does not wear ragged the fabric of friendship: Thinking of objects as alive makes people less willing to replace them. *Journal of Consumer Psychology*, *20*(2), 138–145. https://doi.org/10.1016/j.jcps.2009.12.008

Cheng, X., Zhang, X., Cohen, J., & Mou, J. (2022). Human vs. AI: Understanding the impact of anthropomorphism on consumer response to chatbots from the perspective of trust

and relationship norms. *Information Processing & Management*, *59*(3), Article 102940. https://doi.org/10.1016/j.ipm.2022.102940

Choi, S., Mattila, A. S., & Bolton, L. E. (2021). To err is human(-oid): How do consumers react to robot service failure and recovery? *Journal of Service Research*, *24*(3), 354–371. https://doi.org/10.1177/1094670520978798

Chow, A. R. (2023, February 8). *How ChatGPT managed to grow faster than TikTok or Instagram*. Time Magazine. https://time.com/6253615/chatgpt-fastest-growing/ [Retrieved December 14, 2023]

Christoforakos, L., Gallucci, A., Surmava-Große, T., Ullrich, D., & Diefenbach, S. (2021). Can robots earn our trust the same way humans do? A systematic exploration of competence, warmth, and anthropomorphism as determinants of trust development in HRI. *Frontiers in Robotics and AI*, *8*, Article 640444. https://doi.org/10.3389/frobt.2021.640444

Chung, M., Ko, E., Joung, H., & Kim, S. J. (2020). Chatbot e-service and customer satisfaction regarding luxury brands. *Journal of Business Research*, *117*, 587–595. https://doi.org/10.1016/j.jbusres.2018.10.004

Croes, E. A. J., & Antheunis, M. L. (2021). Can we be friends with Mitsuku? A longitudinal study on the process of relationship formation between humans and a social chatbot. *Journal of Social and Personal Relationships*, *38*(1), 279–300. https://doi.org/10.1177/0265407520959463

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. https://doi.org/10.1177/00222429211045687

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319–340. https://doi.org/10.2307/249008

de Visser, E. J., Monfort, S. S., McKendrick, R., Smith, M. A. B., McKnight, P. E., Krueger, F., & Parasuraman, R. (2016). Almost human: Anthropomorphism increases trust resilience in cognitive agents. *Journal of Experimental Psychology: Applied*, *22*(3), 331–349. https://doi.org/10.1037/xap0000092

della Cava, M. (2016, March 30). *Microsoft CEO Nadella: 'Bots are the new apps'*. USA Today. https://eu.usatoday.com/story/tech/news/2016/03/30/microsof-ceo-nadella-bots-new-apps/82431672/# [Retrieved December 14, 2023]

Diaper, D. (1989). The discipline of HCI. *Interacting with Computers*, *1*(1), 3–5. https://doi.org/10.1016/0953-5438(89)90002-7

Dix, A. (2010). Human–computer interaction: A stable discipline, a nascent science, and the growth of the long tail. *Interacting with Computers*, *22*(1), 13–27. https://doi.org/10.1016/j.intcom.2009.11.007

Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems*, *42*(3–4), 177–190. https://doi.org/10.1016/S0921-8890(02)00374-3

Ebert, A., Gershon, N. D., & van der Veer, G. C. (2012). Human-computer interaction: Introduction and overview. *KI—Künstliche Intelligenz*, *26*(2), 121–126. https://doi.org/10.1007/s13218-012-0174-7

Epley, N., Waytz, A., Akalis, S., & Cacioppo, J. T. (2008). When we need a human: Motivational determinants of anthropomorphism. *Social Cognition*, *26*(2), 143–155. https://doi.org/10.1521/soco.2008.26.2.143

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

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

Ferrari, F., Paladino, M. P., & Jetten, J. (2016). Blurring human–machine distinctions: Anthropomorphic appearance in social robots as a threat to human distinctiveness. *International Journal of Social Robotics*, *8*(2), 287–302. https://doi.org/10.1007/s12369-016-0338-y

Fiske, S. T., & Linville, P. W. (1980). What does the schema concept buy us? *Personality and Social Psychology Bulletin*, *6*(4), 543–557. https://doi.org/10.1177/014616728064006

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR Mental Health*, *4*(2), Article e19. https://doi.org/10.2196/mental.7785

Fossa, F., & Sucameli, I. (2022). Gender bias and conversational agents: An ethical perspective on social robotics. *Science and Engineering Ethics*, *28*(3), Article 23. https://doi.org/10.1007/s11948-022-00376-3

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., Benbasat, I., & Pavlou, P. (2008). A research agenda for trust in online environments. *Journal of Management Information Systems*, *24*(4), 275–286. https://doi.org/10.2753/MIS0742-1222240411

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

Glaser, A. (2016, June 7). *Pepper, the emotional robot, learns how to feel like an American*. WIRED. https://www.wired.com/2016/06/pepper-emotional-robot-learns-feel-like-american/ [Retrieved December 14, 2023]

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

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*, 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

Goldman, E. (2017, September 30). *Before Siri and Alexa, there was ELIZA* [Video]. YouTube. https://www.youtube.com/watch?v=RMK9AphfLco [Retrieved November 3, 2023]

Grand View Research (2023). *Chatbot market size, share, trends & growth report, 2030*. https://www.grandviewresearch.com/industry-analysis/chatbot-market#:~:text=The%20global%20chatbot%20market%20is,USD%2027%2C297.2%20million%20by%202030 [Retrieved September 18, 2023]

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*(5812), 619. https://doi.org/10.1126/science.1134475

Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, *125*(1), 125–130. https://doi.org/10.1016/j.cognition.2012.06.007

Gröppel-Klein, A., Pfeifer, K., & Helfgen, J. (2018). Mit Personifizierungen wirkungsvoll in der Kommunikation emotionalisieren. In T. Langner, F.-R. Esch, & M. Bruhn (Eds.), *Handbuch Techniken der Kommunikation* (2nd ed., pp. 381–398). Springer Gabler. https://doi.org/10.1007/978-3-658-04653-8

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*, *30*(1), 99–120. https://doi.org/10.1007/s11023-020-09517-8

Han, M. C. (2021). The impact of anthropomorphism on consumers' purchase decision in chatbot commerce. *Journal of Internet Commerce*, *20*(1), 46–65. https://doi.org/10.1080/15332861.2020.1863022

Han, B., Deng, X., & Fan, H. (2023). Partners or opponents? How mindset shapes consumers' attitude toward anthropomorphic artificial intelligence service robots. *Journal of Service Research*, *26*(3), 441–458. https://doi.org/10.1177/10946705231169674

Heidegger, M. (1996). *Being and time*. State University of New York Press.

Hern, A. (2016, October 19). *Stephen Hawking: AI will be 'either best or worst thing' for humanity*. The Guardian. https://www.theguardian.com/science/2016/oct/19/stephen-hawking-ai-best-or-worst-thing-for-humanity-cambridge#:~:text=Professor%20Stephen%20Hawking%20has%20warned,future%20of%20our%20civilisation%20and) [Retrieved December 11, 2023]

Hillen, M. A., Postma, R.-M., Verdam, M. G. E., & Smets, E. M. A. (2017). Development and validation of an abbreviated version of the Trust in Oncologist Scale—The Trust

in Oncologist Scale–short form (TiOS-SF). *Supportive Care in Cancer*, *25*(3), 855–861. https://doi.org/10.1007/s00520-016-3473-y

Holthöwer, J., & van Doorn, J. (2022). Robots do not judge: Service robots can alleviate embarrassment in service encounters. *Journal of the Academy of Marketing Science*, *51*, 767–784. https://doi.org/10.1007/s11747-022-00862-x

Hu, K. (2023, February 2). *ChatGPT sets record for fastest-growing user base—analyst note*. Reuters. https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/ [Retrieved 14.12.2023]

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

Kätsyri, J., Förger, K., Mäkäräinen, M., & Takala, T. (2015). A review of empirical evidence on different uncanny valley hypotheses: Support for perceptual mismatch as one road to the valley of eeriness. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00390

Kim, S., Chen, R. P., & Zhang, K. (2016). Anthropomorphized helpers undermine autonomy and enjoyment in computer games. *Journal of Consumer Research*, *43*(2), 282–302. https://doi.org/10.1093/jcr/ucw016

Kim, T. W., Jiang, L., Duhachek, A., Lee, H., & Garvey, A. (2022). Do you mind if I ask you a personal question? How AI service agents alter consumer self-disclosure. *Journal of Service Research*, *25*(4), 649–666. https://doi.org/10.1177/10946705221120232

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

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*, Article 107513. https://doi.org/10.1016/j.chb.2022.107513

Kroeber-Riel, W., & Gröppel-Klein, A. (2019). *Konsumentenverhalten* (11th ed.). Vahlen.

Kwak, H., Puzakova, M., & Rocereto, J. F. (2015). Better not smile at the price: The differential role of brand anthropomorphization on perceived price fairness. *Journal of Marketing*, *79*(4), 56–76. https://doi.org/10.1509/jm.13.0410

Landwehr, J. R., McGill, A. L., & Herrmann, A. (2011). It's got the look: The effect of friendly and aggressive "facial" expressions on product liking and sales. *Journal of Marketing*, *75*(3), 132–146. https://doi.org/10.1509/jmkg.75.3.132

Laranjo, L., Dunn, A. G., Tong, H. L., Kocaballi, A. B., Chen, J., Bashir, R., Surian, D., Gallego, B., Magrabi, F., Lau, A. Y. S., & Coiera, E. (2018). Conversational agents in healthcare: A systematic review. *Journal of the American Medical Informatics Association*, *25*(9), 1248–1258. https://doi.org/10.1093/jamia/ocy072

Larivière, B., Bowen, D., Andreassen, T. W., Kunz, W., Sirianni, N. J., Voss, C., Wünderlich, N. V., & De Keyser, A. (2017). "Service Encounter 2.0": An investigation into the roles of technology, employees and customers. *Journal of Business Research*, *79*, 238–246. https://doi.org/10.1016/j.jbusres.2017.03.008

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

Leuphana Universität (2023, November 28). *Bedingungen und Empfehlungen für die Nutzung von KI-basierten Anwendungen in Lehre und Prüfungen*.

https://www.leuphana.de/lehre/organisation/pruefungsorganisation-fuer-lehrende/ki-empfehlungen.html [Retrieved December 14, 2023]

Leviathan, Y., & Matias, Y. (2018, May 8). *Google Duplex: An AI system for accomplishing real-world tasks over the phone.* Google Research. https://blog.research.google/2018/05/duplex-ai-system-for-natural-conversation.html [Retrieved December 12, 2023]

Lewis, J. D., & Weigert, A. (1985). Trust as a social reality. *Social Forces*, *63*(4), 967–985. https://doi.org/10.2307/2578601

Luhmann, N. (1979). *Trust and Power*. Wiley.

Lv, X., Yang, Y., Qin, D., Cao, X., & Xu, H. (2022). Artificial intelligence service recovery: The role of empathic response in hospitality customers' continuous usage intention. *Computers in Human Behavior*, *126*, Article 106993. https://doi.org/10.1016/j.chb.2021.106993

Mariani, M. M., Hashemi, N., & Wirtz, J. (2023). Artificial intelligence empowered conversational agents: A systematic literature review and research agenda. *Journal of Business Research*, *161*, Article 113838. https://doi.org/10.1016/j.jbusres.2023.113838

Mayring, P. (2000). Qualitative content analysis. *Forum Qualitative Social Research*, *1*(2). https://doi.org/10.17169/fqs-1.2.1089

Mayring, P. (2014). *Qualitative content analysis: Theoretical foundation, basic procedures and software solution*. Beltz.

McDonnell, M., & Baxter, D. (2019). Chatbots and gender stereotyping. *Interacting with Computers*, *31*(2), 116–121. https://doi.org/10.1093/iwc/iwz007

Mende, M., Scott, M. L., van Doorn, J., Grewal, D., & Shanks, I. (2019). Service robots rising: How humanoid robots influence service experiences and elicit compensatory

consumer responses. *Journal of Marketing Research*, *56*(4), 535–556. https://doi.org/10.1177/0022243718822827

Meng, J., & Dai, Y. (2021). Emotional support from AI chatbots: Should a supportive partner self-disclose or not? *Journal of Computer-Mediated Communication*, *26*(4), 207–222. https://doi.org/10.1093/jcmc/zmab005

Miner, A., Chow, A., Adler, S., Zaitsev, I., Tero, P., Darcy, A., & Paepcke, A. (2016). Conversational agents and mental health: Theory-informed assessment of language and affect. In W.-Y. Yau, T. Omori, S. Zhao, H. Osawa, & G. Metta (Eds.), *HAI '16: Proceedings of the 4th International Conference on Human Agent Interaction* (pp. 123–130). Association for Computing Machinery. https://doi.org/10.1145/2974804.2974820

Morhart, F., Malär, L., Guèvremont, A., Girardin, F., & Grohmann, B. (2015). Brand authenticity: An integrative framework and measurement scale. *Journal of Consumer Psychology*, *25*(2), 200–218. https://doi.org/10.1016/j.jcps.2014.11.006

Mori, M. (1970). The uncanny valley. *Energy*, *7*(4), 33–35.

Mori, M., MacDorman, K., & Kageki, N. (2012). The uncanny valley. *IEEE Robotics & Automation Magazine*, *19*(2), 98–100. https://doi.org/10.1109/MRA.2012.2192811

Morrissey, K., & Kirakowski, J. (2013). 'Realness' in chatbots: Establishing quantifiable criteria. In M. Kurosu (Ed.), *Human-Computer Interaction. Interaction Modalities and Techniques* (Vol. 8007, pp. 87–96). Springer. https://doi.org/10.1007/978-3-642-39330-3_10

Mozafari, N., Weiger, W. H., & Hammerschmidt, M. (2022). Trust me, I'm a bot—Repercussions of chatbot disclosure in different service frontline settings. *Journal of Service Management*, *33*(2), 221–245. https://doi.org/10.1108/JOSM-10-2020-0380

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

Noble, S. M., Mende, M., Grewal, D., & Parasuraman, A. (2022). The fifth industrial revolution: How harmonious human–machine collaboration is triggering a retail and service [r]evolution. *Journal of Retailing*, *98*(2), 199–208. https://doi.org/10.1016/j.jretai.2022.04.003

Olson, P. (2018, March 8). *This AI has sparked a budding friendship with 2.5 million people*. Forbes. https://www.forbes.com/sites/parmyolson/2018/03/08/replika-chatbot-google-machine-learning/?sh=4ab5327e4ffa [Retrieved December 12, 2023]

Parviainen, J., & Rantala, J. (2022). Chatbot breakthrough in the 2020s? An ethical reflection on the trend of automated consultations in health care. *Medicine, Health Care and Philosophy*, *25*(1), 61–71. https://doi.org/10.1007/s11019-021-10049-w

Pearson, S. D., & Raeke, L. H. (2000). Patients' trust in physicians: Many theories, few measures, and little data. *Journal of General Internal Medicine*, *15*(7), 509–513. https://doi.org/10.1046/j.1525-1497.2000.11002.x

Pelau, C., Dabija, D.-C., & Ene, I. (2021). What makes an AI device human-like? The role of interaction quality, empathy and perceived psychological anthropomorphic characteristics in the acceptance of artificial intelligence in the service industry. *Computers in Human Behavior*, *122*, Article 106855. https://doi.org/10.1016/j.chb.2021.106855

Pentina, I., Hancock, T., & Xie, T. (2023). Exploring relationship development with social chatbots: A mixed-method study of Replika. *Computers in Human Behavior*, *140*, Article 107600. https://doi.org/10.1016/j.chb.2022.107600

Peres, R., Schreier, M., Schweidel, D., & Sorescu, A. (2023). On ChatGPT and beyond: How generative artificial intelligence may affect research, teaching, and practice.

International Journal of Research in Marketing, *40*(2), 269–275. https://doi.org/10.1016/j.ijresmar.2023.03.001

Perri, L. (2023, August 17). *What's new in artificial intelligence from the 2023 Gartner Hype Cycle*. Gartner. https://www.gartner.com/en/articles/what-s-new-in-artificial-intelligence-from-the-2023-gartner-hype-cycle [Retrieved December 8, 2023]

Pitardi, V., Wirtz, J., Paluch, S., & Kunz, W. H. (2022). Service robots, agency and embarrassing service encounters. *Journal of Service Management*, *33*(2), 389–414. https://doi.org/10.1108/JOSM-12-2020-0435

Powell, J. (2019). Trust me, I'm a chatbot: How artificial intelligence in health care fails the Turing test. *Journal of Medical Internet Research*, *21*(10), Article e16222. https://doi.org/10.2196/16222

Puzakova, M., Kwak, H., & Rocereto, J. F. (2013). When humanizing brands goes wrong: The detrimental effect of brand anthropomorphization amid product wrongdoings. *Journal of Marketing*, *77*(3), 81–100. https://doi.org/10.1509/jm.11.0510

Qiu, L., & Benbasat, I. (2009). Evaluating anthropomorphic product recommendation agents: A social relationship perspective to designing information systems. *Journal of Management Information Systems*, *25*(4), 145–182. https://doi.org/10.2753/MIS0742-1222250405

Rauschnabel, P. A., & Ahuvia, A. C. (2014). You're so lovable: Anthropomorphism and brand love. *Journal of Brand Management*, *21*(5), 372–395. https://doi.org/10.1057/bm.2014.14

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

Rempel, J. K., Holmes, J. G., & Zanna, M. P. (1985). Trust in close relationships. *Journal of Personality and Social Psychology*, *49*(1), 95–112. https://doi.org/10.1037/0022-3514.49.1.95

Rouse, W. B., & Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin*, *100*(3), 349–363. https://doi.org/10.1037/0033-2909.100.3.349

Salem, M., Eyssel, F., Rohlfing, K., Kopp, S., & Joublin, F. (2013). To err is human(-like): Effects of robot gesture on perceived anthropomorphism and likability. *International Journal of Social Robotics*, *5*(3), 313–323. https://doi.org/10.1007/s12369-013-0196-9

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

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. https://doi.org/10.17705/1jais.00685

Seitz, L., & Bekmeier-Feuerhahn, S. (2023a). Anthropomorphe Markengestaltung—Wenn Marken menschlich werden. *transfer—Zeitschrift für Kommunikation und Markenmanagement*, *69*(2), 48–55.

Seitz, L., & Bekmeier-Feuerhahn, S. (2023b). 'Please, just make service faster': When human-likeness in chatbots backfires [Abstract]. *Proceedings of the 52nd European Marketing Academy (EMAC 2023)*, Denmark, Article 113944.

Seitz L., & Bekmeier-Feuerhahn, S. (2022). Does anthropomorphism in chatbots enhance customers' willingness to pay? First evidence from a preliminary study [Abstract]. *Proceedings of the 51th European Marketing Academy (EMAC 2022)*, Hungary, Article 107541.

Seitz, L., & Bekmeier-Feuerhahn, S. (2021). Empathic healthcare chatbots: Comparing the effects of emotional expression and caring behavior. *Proceedings of the 42nd International Conference on Information Systems (ICIS 2021)*, USA, Article 4.

Seitz, L., Bekmeier-Feuerhahn, S., Bontrup, F., Wildt, J., & Gohil, K. (2020). Towards a model for building trust and acceptance of artificial intelligence aided medical assessment systems. *Proceedings of the 49th European Marketing Academy (EMAC 2020)*, Hungary (cancelled), Article 64418.

Seitz, L., Woronkow, J., Bekmeier-Feuerhahn, S., & Gohil, K. (2021). The advance of diagnosis chatbots: Should we first avoid distrust before we focus on trust? *Proceedings of the 29th European Conference on Information Systems (ECIS 2021)*, A Virtual Conference, Article 9.

Sharma, M., & Rahman, Z. (2022). Anthropomorphic brand management: An integrated review and research agenda. *Journal of Business Research*, *149*, 463–475. https://doi.org/10.1016/j.jbusres.2022.05.039

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

Shum, H., He, X., & Li, D. (2018). From Eliza to XiaoIce: Challenges and opportunities with social chatbots. *Frontiers of Information Technology & Electronic Engineering*, *19*(1), 10–26. https://doi.org/10.1631/FITEE.1700826

Silva, G. R. S., & Canedo, E. D. (2022). Towards user-centric guidelines for chatbot conversational design. *International Journal of Human–Computer Interaction*. https://doi.org/10.1080/10447318.2022.2118244

Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzaeg, P. B. (2021). My chatbot companion—A study of human-chatbot relationships. *International Journal of*

Human-Computer Studies, *149*, Article 102601. https://doi.org/10.1016/j.ijhcs.2021.102601

Skjuve, M., Haugstveit, I. M., Følstad, A., & Brandtzaeg, P. B. (2019). Help! Is my chatbot falling into the uncanny valley? An empirical study of user experience in human-chatbot interaction. *Human Technology*, *15*(1), 30–54. https://doi.org/10.17011/ht/urn.201902201607

Spezialetti, M., Placidi, G., & Rossi, S. (2020). Emotion recognition for human-robot interaction: Recent advances and future perspectives. *Frontiers in Robotics and AI*, *7*, Article 532279. https://doi.org/10.3389/frobt.2020.532279

Statista (2023). *Weltweites Marktvolumen von Chatbots im Jahr 2022 und Prognose bis 2032*. https://de.statista.com/statistik/daten/studie/1373729/umfrage/weltweites-marktvolumen-chatbots/ [Retrieved December 14, 2023]

Stein, J.-P., & Ohler, P. (2017). Venturing into the uncanny valley of mind—The influence of mind attribution on the acceptance of human-like characters in a virtual reality setting. *Cognition*, *160*, 43–50. https://doi.org/10.1016/j.cognition.2016.12.010

Stricke, B. (2020). People v. robots: A roadmap for enforcing California's new online bot disclosure act. *Vanderbilt Journal of Entertainment & Technology Law*, *22*(4), 839–894.

Sujan, M., Furniss, D., Grundy, K., Grundy, H., Nelson, D., Elliott, M., White, S., Habli, I., & Reynolds, N. (2019). Human factors challenges for the safe use of artificial intelligence in patient care. *BMJ Health & Care Informatics*, *26*(1). https://doi.org/10.1136/bmjhci-2019-100081

Trivedi, J. (2019). Examining the customer experience of using banking chatbots and its impact on brand love: The moderating role of perceived risk. *Journal of Internet Commerce*, *18*(1), 91–111. https://doi.org/10.1080/15332861.2019.1567188

Turing, A. M. (1950). I.—Computing machinery and intelligence. *Mind*, *LIX*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

Uysal, E., Alavi, S., & Bezençon, V. (2022). Trojan horse or useful helper? A relationship perspective on artificial intelligence assistants with humanlike features. *Journal of the Academy of Marketing Science*, *50*(6), 1153–1175. https://doi.org/10.1007/s11747-022-00856-9

van Doorn, J., Mende, M., Noble, S. M., Hulland, J., Ostrom, A. L., Grewal, D., & Petersen, J. A. (2017). Domo arigato Mr. Roboto: Emergence of automated social presence in organizational frontlines and customers' service experiences. *Journal of Service Research*, *20*(1), 43–58. https://doi.org/10.1177/1094670516679272

Velasco, F., Yang, Z., & Janakiraman, N. (2021). A meta-analytic investigation of consumer response to anthropomorphic appeals: The roles of product type and uncertainty avoidance. *Journal of Business Research*, *131*, 735–746. https://doi.org/10.1016/j.jbusres.2020.11.015

Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the Unified Theory of Acceptance and Use of Technology. *MIS Quarterly*, *36*(1), 157–178. https://doi.org/10.2307/41410412

Waizenegger, L., Seeber, I., Dawson, G., & Desouza, K. (2020). Conversational agents—Exploring generative mechanisms and second-hand effects of actualized technology affordances. *Proceedings of the 53rd Hawaii International Conference on System Sciences*, Hawaii, 5180–5189. https://doi.org/10.24251/HICSS.2020.636

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

Waytz, A., & Norton, M. I. (2014). Botsourcing and outsourcing: Robot, British, Chinese, and German workers are for thinking—not feeling—jobs. *Emotion*, *14*(2), 434–444. https://doi.org/10.1037/a0036054

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

Weizenbaum, J. (1976). *Computer power and human reason: From judgment to calculation*. Freeman.

West, M., Kraut, R., & Chew, H. E. (2019). *I'd blush if I could. Closing gender divides in digital skills through education*. UNESCO & EQUALS Skills Coalition. https://unesdoc.unesco.org/ark:/48223/pf0000367416 [Retrieved December 14, 2023]

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

Wood, A. M., Linley, P. A., Maltby, J., Baliousis, M., & Joseph, S. (2008). The authentic personality: A theoretical and empirical conceptualization and the development of the Authenticity Scale. *Journal of Counseling Psychology*, *55*(3), 385–399. https://doi.org/10.1037/0022-0167.55.3.385

Wünderlich, N., & Paluch, S. (2017). A nice and friendly chat with a bot: User perceptions of AI-based service agents. *Proceedings of the 38th International Conference on Information Systems (ICIS 2017)*, South Korea, Article 11.

Yanxia, C., Shijia, Z., & Yuyang, X. (2023). A meta-analysis of the effect of chatbot anthropomorphism on the customer journey. *Marketing Intelligence & Planning*. https://doi.org/10.1108/MIP-03-2023-0103

Yu, S., Xiong, J., & Shen, H. (2022). The rise of chatbots: The effect of using chatbot agents on consumers' responses to request rejection. *Journal of Consumer Psychology*, Article jcpy.1330. https://doi.org/10.1002/jcpy.1330

Yuan, L., & Dennis, A. R. (2019). Acting like humans? Anthropomorphism and consumer's willingness to pay in electronic commerce. *Journal of Management Information Systems*, *36*(2), 450–477. https://doi.org/10.1080/07421222.2019.1598691

Zenthöfer, J. (2023, December 1). *Erste Uni schafft Bachelorarbeiten ab*. Frankfurter Allgemeine. https://www.faz.net/aktuell/karriere-hochschule/hoersaal/erste-uni-schafft-wegen-ki-und-plagiaten-bachelorarbeiten-ab-19353621.html [Retrieved December 11, 2023]

Zhang, L., Pentina, I., & Fan, Y. (2021). Who do you choose? Comparing perceptions of human vs robo-advisor in the context of financial services. *Journal of Services Marketing*, *35*(5), 628–640. https://doi.org/10.1108/JSM-05-2020-0162

Złotowski, J., Yogeeswaran, K., & Bartneck, C. (2017). Can we control it? Autonomous robots threaten human identity, uniqueness, safety, and resources. *International Journal of Human-Computer Studies*, *100*, 48–54. https://doi.org/10.1016/j.ijhcs.2016.12.008

Image Sources

Figure 1

Chatbot "Sophie" (Check24) (n.d.). Screenshot from: https://www.check24.de/unternehmen/kontakt/?chatFlowName=home.welcome&chatFlowRelease=active&chatFlowEntry=start_from_contact_page&sec=c24 [Retrieved November 19, 2023]

Chatbot "Linda" (Sparkasse Bodensee) (n.d.). Screenshot from: https://www.sparkasse-bodensee.de/de/home/service/chatbot-linda.html [Retrieved November 19, 2023]

Social companion robot "Pepper" (n.d.). Screenshot from: https://www.probo-robotics.at/de/humanoider-roboter-pepper/ [Retrieved November 19, 2023]

Figure 6

Replika (2023). *Replika is the #1 AI-powered companion, tailor-made to understand you and transform every chat into an unforgettable journey* [Advertisement screenshot]. Facebook. https://www.facebook.com/myownreplika [Retrieved December 14, 2023]

PART II: PAPER 1

CAN WE TRUST A CHATBOT LIKE A PHYSICIAN? A QUALITATIVE STUDY ON UNDERSTANDING THE EMERGENCE OF TRUST TOWARD DIAGNOSTIC CHATBOTS

Fact Sheet Paper 1

Title	Can We Trust a Chatbot Like a Physician? A Qualitative Study on Understanding the Emergence of Trust Toward Diagnostic Chatbots
Authors	Lennart Seitz, Sigrid Bekmeier-Feuerhahn & Krutika Gohil
Year	2022
Citation	Seitz, L., Bekmeier-Feuerhahn, S., & Gohil, K. (2022). Can we trust a chatbot like a physician? A qualitative study on understanding the emergence of trust toward diagnostic chatbots. *International Journal of Human-Computer Studies*, *165*, 102848.
DOI	https://doi.org/10.1016/j.ijhcs.2022.102848

Abstract

Technological advancements in the virtual assistants' domain pave the way to implement complex autonomous agents like diagnostic chatbots. Drawing on the assumption that chatbots are perceived as both technological tools and social actors, we aim to create a deep understanding of trust-building processes towards diagnostic chatbots compared to trust in medical professionals. We conducted a laboratory experiment in which participants interacted either with a diagnostic chatbot only or with an additional telemedicine professional before we interviewed them primarily on trust-building factors. We identified numerous software-related, user-related, and environment-related factors and derived a model of the initial trust-building process. The results support our assumption that it is equally essential to consider dimensions of physician and technology trust. One significant finding is that trust in a chatbot arises cognitively, while trusting a human agent is affect-based. We argue that the lack of affect-based trust inhibits the willingness to rely on diagnostic chatbots and facilitates the user's desire to keep control. Considering dimensions from doctor-patient trust, we found evidence that a chatbot's communication competencies are more important than empathic reactions as the latter may evoke incredibility feelings. To verify our findings, we applied the derived code system in a larger online survey.

Keywords: trust; chatbot; conversational agent; mHealth; anthropomorphism; telemedicine

1 Introduction

Conversational agents (CAs) are increasingly used in demanding and sensitive environments like healthcare given their advanced capacity to process even complex information. Healthcare CAs offer a simple and efficient form of information, empowering patients to engage in decision-making processes and self-care (Denecke et al., 2019). For instance, CAs are used successfully in psychotherapy, behaviour change interventions, elderly care, and diagnosis (Bickmore et al., 2013; Montenegro et al., 2019; Provoost et al., 2017).

Although today's healthcare CAs show decent performance, their adoption faces specific challenges. One of these is creating trust toward the system, which is vital considering the novelty of healthcare CAs and the situation's sensitivity (Laranjo et al., 2018; Nundy et al., 2019). Trust is originally an interpersonal concept that has frequently been adapted to study interactions between humans and virtual agents (Benbasat and Wang, 2005; Wang et al., 2016). Researchers justify the applicability of interpersonal trust dimensions commonly with the "Computers are Social Actors" paradigm (Reeves and Nass, 1996). Accordingly, humans automatically apply social heuristics to interactions with computers. However, trust toward physicians is significantly based on emotional attachment, human warmth, and reciprocity which is difficult to achieve in interactions with chatbots (Thom and Campbell, 1997). Although a lot of studies have demonstrated the applicability of interpersonal trust models to interactions with recommendation chatbots (e.g., Wang and Benbasat, 2016), it is yet questionable in how far physician trust models are sufficient to explain trust toward healthcare CAs. Since previous research has mostly focused on rigid criteria such as the accuracy of treatment outcomes (Laranjo et al., 2018; Vaidyam et al., 2019), recent early-stage work points to the relevance of studying the emergence of trust toward healthcare CAs (Laumer et al., 2019; Wang and Siau, 2018).

The objective of our study is therefore to better understand how trust in healthcare CAs arises and what the differences are compared to trust in physicians. We do so since CAs show certain social characteristics (Feine et al., 2019) and since interpersonal relations are vital in providing healthcare services (Thom and Campbell, 1997). The latter might make the development of trust toward health CAs significantly different from that toward customer service chatbots, i.e., when functional aspects are more important than emotional attachment (Blut et al., 2021; Nordheim et al., 2019). To approach our research objective, we conducted two studies in which participants interacted with diagnostic chatbots before we interviewed them on their experiences and trust-building factors. We adopted mainly qualitative research methods to deeply understand trust-building processes. To draw parallels between doctor-patient and chatbot-patient trust, we set a focus on corresponding similarities and differences in conducting our studies.

The contribution of our research is twofold. First, we derive a processual and empirically supported model of initial trust-building toward diagnostic chatbots considering internal (software-related) and external (user- and environment-related) factors. Second, we contribute to a better understanding of the applicability of physician and interpersonal trust dimensions to interactions with diagnostic chatbots. In this way, we identify opportunities and boundaries of anthropomorphizing CAs in sensitive environments.

2 Conceptual Background

2.1 The Hybrid Nature of Conversational Agents

The use of artificial intelligence-based conversational agents has significantly increased over the past years (Araujo, 2018). CAs include voice-based personal assistants or text-based chatbots that take on an interaction partner's role to provide several kinds of self-services (Feine et al., 2019). Due to their role as highly responsive interaction partners, CAs represent an exception among software tools. Compared to prior generations of information systems, they

show certain social characteristics as they imitate human intelligence capable of autonomous decision-making (Glikson and Woolley, 2020). In addition, modern CAs are explicitly anthropomorphized to make the conversation more natural (Feine et al., 2019). For instance, text-based chatbots can send verbal social cues like greeting or keeping small talk. Also, concepts associated with human beings like expressing empathy (Fitzpatrick et al., 2017; Liu and Sundar, 2018) and self-disclosure (Lee and Choi, 2017) are implemented to increase the perceived interpersonal communication competence of chatbots (Skjuve and Brandzaeg, 2019).

Consequently, people can perceive CAs as teammates and may even engage in romantic relationships with them (Bickmore et al., 2005; Muresan and Pohl, 2019). A frequently cited approach explaining this phenomenon is the "Computers are Social Actors" (CASA) paradigm. It assumes that humans unconsciously apply basic heuristics of interpersonal behaviour in interactions with computers (Nass et al., 1994; Reeves and Nass, 1996).

Nevertheless, the general adoption of interpersonal communication approaches to human-chatbot interactions is controversial. For instance, people reveal CAs in interactions quickly due to inappropriate or scripted messages (Skjuve et al., 2019). Thus, basic linguistic principles of human communication like coherence are violated. Also, humans' insecurity in communicating with chatbots becomes apparent in the number, structure, and formulation of messages (Hill et al., 2015). Besides, differences can also be attributed to people's divergent expectations toward artificial communication partners (Muresan and Pohl, 2019). According to the "Uncanny Valley of Mind" (UVM), objects with a high level of human-likeness can create a feeling of eeriness (Mori et al., 2012). Although the theory originally refers to physical robots, it has also been investigated in studies on interactions between humans and virtual entities (Skjuve et al., 2019; Stein and Ohler, 2017). From an anthropocentric perspective, researchers argue that people fear the endangerment of human supremacy and perceive an identity threat (Stein et al., 2019; Yogeeswaran et al., 2016). Correspondingly, it is assumed "that people

prefer human-like replicas to be limited to a certain set of characteristics and might not appreciate them to behave in an empathic or social manner" (Stein and Ohler, 2017, p. 48). Although the existence of the UVM in interactions with chatbots has not yet been demonstrated, the desire for human-likeness seems to be limited (Muresan and Pohl, 2019). Also, the service environment in which an artificial entity is used may impact the outcomes of anthropomorphizing agents (Blut et al., 2021; Chi et al., 2021). Thus, it is to be investigated to what extent people perceive and treat healthcare CAs as social actors and how far the concept of doctor-patient trust applies to interactions with them.

2.2 Trust-Building Toward Human Beings, Physicians, and Artificial Entities

2.2.1 Trust in Interpersonal Relationships

Many different disciplines have studied the abstract concept of trust, including sociology, psychology, and organization theory (Lewis and Weigert, 1985; Mayer et al., 1995). Trust is seen as a critical element in interpersonal relationships as it reduces complexity in situations of uncertainty, vulnerability, and risk (Gefen et al., 2008; Luhmann, 1979; Rempel et al., 1985). Although it is difficult to find a common definition, the highly cited quotation of Mayer, Davis, and Schoorman (1995) defines trust as "the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, regardless of the ability to monitor or control that other party" (p. 712). This definition includes several elements that have been adopted by other researchers in their conceptualizations of trust. For instance, it describes the involvement of at least two parties in trusting relationships, i.e., the *trustor* who relies upon and the *trustee* who is to be trusted. In this sense, trust is an interpersonal concept assuming that its emergence depends both on individual characteristics of the trustor (e.g., the subject's propensity to trust) and on their beliefs concerning the characteristics of the trustee (e.g., their trustworthiness) (Gefen et al., 2008; McKnight and Chervany, 2001b). While a trustor's propensity to trust

results from a person's general traits and attitudes (Barber, 1983), beliefs concerning a trustee's trustworthiness are mainly formed by evaluating their *ability*, *benevolence,* and *integrity*. Ability describes a trustee's competence to fulfil a specific task, whereas benevolence and integrity represent time stable morality-related characteristics (Wang and Benbasat, 2016). Beliefs concerning benevolence indicate how far a trustor assumes that the trustee is acting in their best interest while a trustee's integrity is evaluated through beliefs about their adherence to a set of acceptable principles (e.g., honesty) (Mayer et al., 1995). Some researchers also include *predictability* among the trusting beliefs, although it is controversial how far a trustor takes risk when the trustee's actions are predictable (Mayer et al., 1995; McKnight et al., 1998).

Further literature distinguishes between *cognition-based trust* (trust from the head) and *affect-based trust* (trust from the heart) (McAllister, 1995). Cognition-based trust (CBT) describes trusting for good reasons and often occurs within formal relationships. In contrast, affect-based trust (ABT) is based on emotional bonds like friendships, thus having a reciprocal character. Researchers occasionally argue that CBT is based on beliefs concerning a trustee's ability, while ABT reflects general feelings toward a person and is thus stronger associated with benevolence (Chua et al., 2008; Lewis and Weigert, 1985; McAllister, 1995). The quotation of Mayer et al. (1995) refers further to the functional outcomes of trust. Once trust occurs, it manifests in trusting intentions and trust-related behaviour, meaning "that a person voluntarily depends on another person with a feeling of relative security, even though negative consequences are possible" (McKnight and Chervany, 2001b). Thus, regardless of the disciplinary perspective, trust is considered a psychological state, entailing the confident expectation that another party will not exploit one's vulnerability in a situation of uncertainty (Cho, 2006; Kramer, 1999; Mayer et al., 1995).

2.2.2 On the Special Role of Trust in Doctor-Patient Relationships

Considering a person's high vulnerability in medical consultations, the crucial role of trust between doctor and patient becomes apparent (Pearson and Raeke, 2000). Not only can diseases pose a high risk for well-being, but patients must also rely entirely on the information given by their doctors (Buchanan, 1988). A trustful relationship leads to several positive outcomes, like improved information exchange, better adherence to the physician's recommendations, and reduced fear (Hillen et al., 2017). Given its importance, many researchers have conceptualized trust in doctor-patient relationships whereby they often refer to fundamental interpersonal trust concepts (Hall et al., 2002; Hillen et al., 2017). One particularity is the unique role of a physician's interpersonal competencies, including empathy and active listening (Hillen et al., 2017; Thom and Campbell, 1997). In medical consultations, patients expect to be understood and emotionally supported, emphasizing the importance of ABT (Jeffrey, 2016). Also, the communication competencies of a physician have a significant impact on trust-building. Considering the knowledge asymmetry between physician and patient, providing information and shared decision-making can enhance trust as many patients expect collegial rather than hierarchical relationships (Anderson and Dedrick, 1990; Ivbijaro et al., 2014; Thom and Campbell, 1997). Lastly, confidentiality is an integral part of the trusting relationship between doctor and patient since highly sensitive personal information is shared (Anderson and Dedrick, 1990; Hall et al., 2001).

2.2.3 Can We Trust Artificial Entities?

Although trust is an interpersonal concept, it has frequently been adopted in research on information systems. Various studies indicate trust as a significant predictor for the acceptance and adoption of technological artifacts like websites, virtual assistants, and other automated systems (Benbasat and Wang, 2005; Gefen et al., 2003, 2008; Lee and See, 2004). Many of the available publications adopt interpersonal trust models and dimensions to investigate trust-

building toward technology and virtual entities (Al-Natour et al., 2010; Benbasat and Wang, 2005; Gefen et al., 2003; Komiak and Benbasat, 2006; McKnight et al., 2002). While some researchers argue that general trust-related factors like risk or complexity also appear in online environments (Corritore et al., 2003; McKnight et al., 2002), others justify their applicability with the CASA paradigm. Especially research on highly responsive virtual assistants frequently considers interpersonal aspects when studying trust-building (Al-Natour et al., 2006; Komiak and Benbasat, 2006; Wang et al., 2016). Given their social characteristics, AI-driven agents may evoke an emotional attachment, thus triggering ABT (Glikson and Woolley, 2020). Correspondingly, a whole stream of research investigates the effects of humanness on the perceived trustworthiness of virtual assistants (e.g., de Visser et al., 2016; Qiu and Benbasat, 2009).

Nevertheless, the applicability of interpersonal trust models to interactions with virtual entities is not undisputed (Gefen et al., 2008). First, inanimate systems are not morally capable subjects since they do not have a consciousness or intentions (Corritore et al., 2003; Friedman et al., 2000). Therefore, it is questionable to ascribe traits such as benevolence or integrity to them. Researchers suggest applying alternative or equivalent dimensions to capture virtual entities' trustworthiness. For instance, Thatcher et al. (2011) propose considering a technology's predictability instead of integrity since it is more appropriate to refer to a system's consistency. Also, researchers conceptualize a system's trustworthiness by the dimensions of *performance, purpose,* and *process,* frequently distinguishing trust toward the system from trust toward the provider (Lee and See, 2004; Siau and Wang, 2018; Söllner et al., 2012). Second, several technology-related aspects like software failures, usability, and data privacy concerns influence trust-building processes (Flavián et al., 2006; Lee and Moray, 1992; Malhotra et al., 2004). In this sense, the user's missing possibility to emphasize with AI-driven systems' complex inner workings is particularly noteworthy, further increasing uncertainty (Glikson and Woolley,

2020; Lee and Choi, 2017). One way to reduce this information asymmetry is to enhance the systems' transparency as it "allows the user to understand the way it works and explains system choices and behaviour" (Cramer et al., 2008, p. 457). Third, neuroscientific experimental evidence suggests that trusting virtual avatars leads to other brain activations than trusting humans (Riedl et al., 2014). That further supports the assumption that only adopting antecedents of interpersonal trust may not be sufficient to explain trust in artificial entities. And lastly, the capabilities, the features, and the societal role of software systems can rapidly change in a very short period. While twenty years ago the world wide web was still considered an unregulated space and only few individuals used cell phones, people now carelessly transmit their credit card number to online retailers via smartphone. Trust toward technology is thus also a result of the environment's structure it is embedded in and the users' familiarity with it (McKnight et al., 2002). Although research on interpersonal trust also emphasizes the shift from initial trust to history-based trust in recurring interactions (Kramer, 1999), fundamental attributes of human beings and social systems are more stable than technological environments.

Summarizing these considerations, the understanding of trust-building toward diagnostic CAs is crucial for the following reasons: (1) the involved risk, vulnerability, and sensitivity in medical assessments, (2) the uncertainty evoking novelty of diagnostic CAs, and (3) the users' strong dependence on the CA since it takes the role of the better-informed agent.

3 Method

3.1 Study Material and Procedure

3.1.1 Chatbot Prototype

To approach our research questions, we conducted a study using a diagnostic CA prototype from an mHealth company with which we collaborated in a research project. The CA is developed on a custom framework in the "Julia" programming language and uses NLP to function as a self-triage tool. To provide a preliminary medical assessment, the CA obtains

patient's input regarding self-reported symptoms, red flags, and risk factors systematically. The API is from "Infermedica" and consists of medical knowledge database which covers 698 conditions, 1308 symptoms, and 175 risk factors. This data base is constantly audited by medical experts through peer reviews of symptoms, acceptance tests, expert reviews, regression testing, and manual testing. Moreover, the CA has class Ia certification and CE mark following medical device regulations. Achievement of this certification proves that general safety and performance requirements are met with rigorous clinical and usability evidence. It also emphasizes that the risk management was in compliance with the requirements and can be demonstrated through the application of harmonized standards and common specifications. In addition, the device has gone through rigorous internal testing by doctors and health communication scientists before the study took place.

Although the provider was able to further test the device within this study, the research question and the applied methods have been conceptualized independently and were not about evaluating a specific software. Instead, we considered the software as an exemplary representative of diagnostic CAs and thus framed the study material accordingly.

3.1.2 Sample, Preparation, and Pre-Interaction Interviews

We conducted a laboratory experiment with twenty-seven students who interacted with the diagnostic CA prototype and asked them about their experience and trust-building factors in face-to-face interviews. Nineteen females (70.4%) and eight males (29.6%) participated in the study. The mean age is 23.5 years (*SD*=2.46), and none of the subjects suffered from a chronic disease. For a first unbiased impression, subjects were asked about their attitudes, wishes, and expectations toward diagnostic CAs before the actual interaction took place. We conducted semi-structured interviews that contained theoretically derived open-ended questions that allowed us to respond flexibly, thus enabling reciprocity between interviewer and interviewee (Galletta, 2013; Rubin and Rubin, 2011). Following the pre-interaction

interview, the participants were asked to put themselves in a sick person's position. Therefore, one of two scenarios was randomly handed out, in which (1) a cold (fourteen subjects) or (2) a bladder infection (thirteen subjects) was described in detail. We used varying severity of diseases to manipulate the perceived uncertainty and risk to gain deeper insights into trust-building processes. To ensure that both scenarios are perceived equally imaginable but with different degrees of severity, we conducted a pre-test with n=23 participants who had to evaluate the scenarios' imaginability and the diseases' severity on a seven-point Likert scale. Results indicate that both scenarios are equally imaginable ($M_{bladder}$=6.15, $SD_{bladder}$=1.06; M_{cold}=5.81, SD_{cold} =1.50, t(21)=.633, p=.533) while the bladder infection is evaluated significantly more severe than the cold ($M_{bladder}$=4.48, $SD_{bladder}$=1.38; M_{cold}=3.58, SD_{cold}=0.97, t(21)=1.831, p<.05). Furthermore, only six out of twelve subjects with cold symptoms reported that they would visit a doctor, whereas ten out of eleven from the bladder infection condition would do, X^2(1, N=23)=4.537, p<.05. Accordingly, we used both scenarios with minor revisions for our study.

3.1.3 Interaction with Chatbot and Post-Interaction Interviews

After subjects have read the scenario, they interacted with the diagnostic CA at a prepared computer. The CA first asked for personal information (e.g., age and gender) before it captured symptoms in an open answer question. It then asked more specific questions on the entered symptoms using both open ended and pre-defined answer formats (i.e., buttons). After the data collection and analysis were finished, the CA displayed up to four possible diseases in its assessment. Further background information, treatment recommendations, and the share of persons with similar symptoms diagnosed with the relating disease were displayed (see Figure 1). The interaction lasted approximately five minutes (M=5.06, SD=1.69), and participants spent another three minutes (M=3.20, SD=1.19) on examining the CA's assessment. Due to misentries, the CA indicated an emergency in two subjects and recommended to call an

ambulance. However, we showed them the assessment screen of a previous interaction to provide an impression on how the assessment looks like. Technical problems arose another four times (e.g., the conversation did not proceed due to software or connection issues). In all cases, we had been able to fix the problem and restart the interaction successfully. Since software failures are common and can impact trust-building (Toader et al., 2019), we decided to not exclude the corresponding subjects.

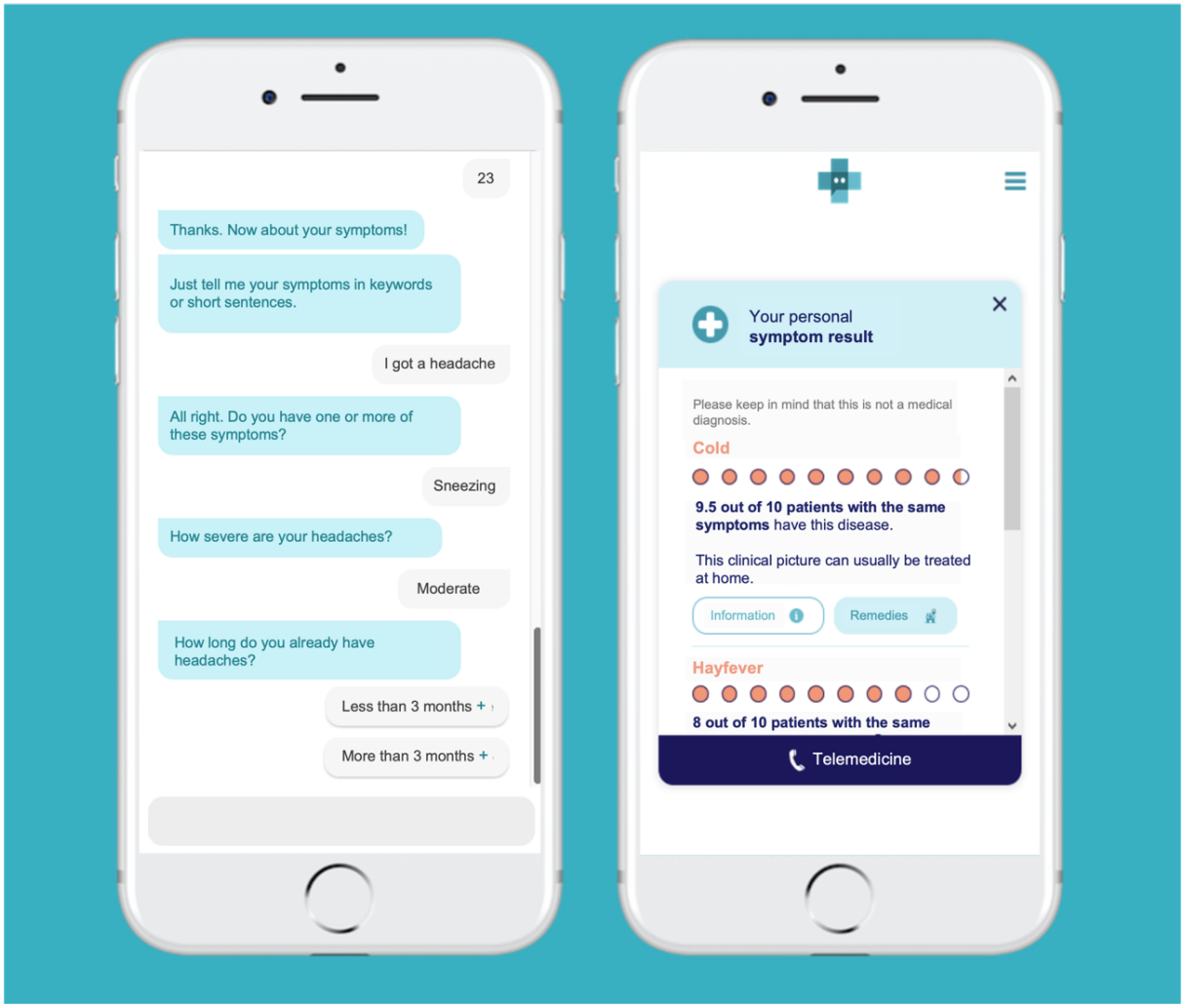


Figure 1. Screenshots of interaction (left) and assessment (right).

Since the software enables its users to be connected to a telemedicine professional and since one of our main goals is to understand better differences in trusting CAs compared to trusting human agents, eleven randomly selected subjects were connected to telemedicine

professionals after the preliminary assessment (see Table 1). These professionals were initiated to the study and had access to the CA's assessment, personal information about the patient, and the symptoms collected. However, they did not know the patient's scenario (i.e., the clinical pictures) to enhance the situations' realism. The telemedicine consultation took place by telephone and enabled patients to talk to an additional human. However, we decided to connect not all participants to medical professionals to get insights on users' concerns when not having the possibility to be assured by a human being. Concurrently, participants who had been able to talk to an additional human may also provide interesting insights on the role of human beings in trust-building processes.

Finally, all participants were interviewed in a semi-structured post-interaction interview, which included pre-defined questions about the general user experience, perceived differences between CA and physician, several aspects of trust-building, the conversation, and the CA's social competencies.

Disease	CA only	CA + Telemedicine
Cold	8	6
Bladder Infection	8	5

Table 1. Allocation of participants to the four conditions.

Note: It was planned to recruit forty participants and to distribute them equally to all conditions. Due to COVID-19 pandemic, the laboratory experiment had to be stopped.

3.2 Process of Data Analysis

The data was interpersonally aggregated due to the relatively large amount of 303 written pages of data material. Therefore, we used systematic categorization and coding as this is fundamental for rigorous qualitative research (Grodal et al., 2020). We used both inductive and deductive methods to explore the data material openly while also relating to previous

research (Locke et al., 2020). The use of both methods enabled us to develop a well-founded, theory-based category system without harming our study's explorative character. More specifically, we decided to apply the "Summarizing Content Analysis" that is an appropriate systematic text analysis method, to reduce interviews to crucial elements and identify latent structures (Mayring, 2000, 2014). In congruence with other related text analysis methods, relevant statements are first identified in the interviews and then classified into concepts, dimensions, and categories under referencing literature (Gioia et al., 2013; Mayring, 2014; Strauss and Corbin, 1998). In this process, a rule- and theory-based approach and a high level of transparency are crucial to reducing subjectivity and ensuring comprehensibility. Throughout, we adhere closely to the steps and rules of "Summarizing Content Analysis", as introduced by Mayring.

In the first step, we reduced the interviews one by one to core statements. To this end, we transcribed the recorded interviews literally before examining the documents for relevant text passages. This examination took place parallel to the data collection to consider interesting aspects in more depth during subsequent interviews. Irrelevant or trivial statements were not further considered to stay focus on our research objectives. The extracted text passages were paraphrased to a comparable form in length and wording without changing their content using "MAXQDA" software (VERBI GmbH, 2020). In this way, we created a total of 1866 paraphrases. Afterward, the paraphrases were further abstracted and generalized to core statements, deleting redundant and invalid paraphrases (see Figure 2). A paraphrase was considered invalid if (1) it did not show any relation to the research questions, (2) it was too generic, thus containing no information, or (3) it resulted from an answer to a suggestive question. The deletions resulted in a total of 1437 generalized statements that were used for the subsequent reduction process.

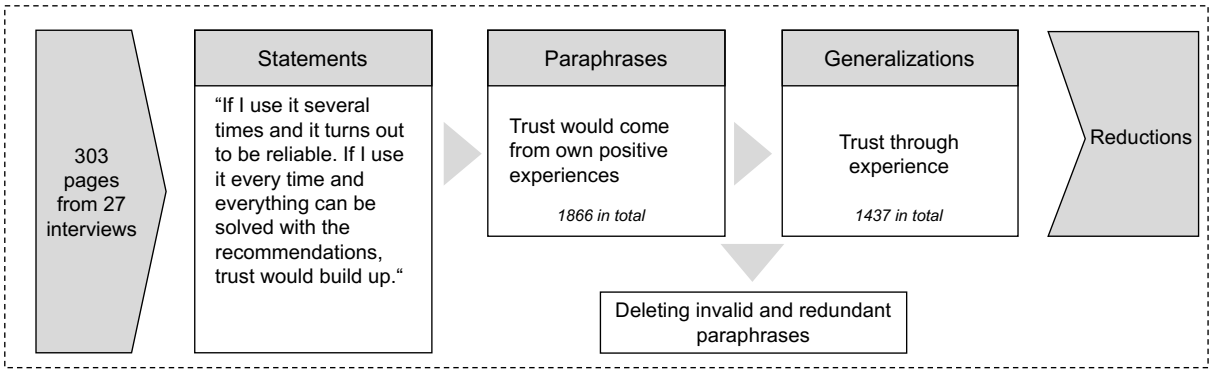


Figure 2. Process of reducing original statements to vital elements.

Note: The example shows a statement of P7.

We created separate intrapersonal reductions for each of the twenty-seven interviewees by bundling and integrating similar and related generalizations (Grodal et al., 2020; Mayring, 2014). These reductions break down the interviews into core elements and topics, enabling us to build interpersonal reductions by integrating similar or identical generalizations interpersonally (see Figure 3). Finally, this resulted in an overview of factors related to our topics of interest (e.g., trust-building factors). Simultaneously, we noticed that various factors were mentioned in different contexts, leading to overlapping. To overcome these ambiguities, we (1) created a mind map of all topics and factors, (2) considered the frequency a factor was mentioned in the context of a specific topic, and (3) considered the relationships between the topics and factors to understand underlying mechanisms and meanings better. Consequently, we derived the first category system for further investigation.

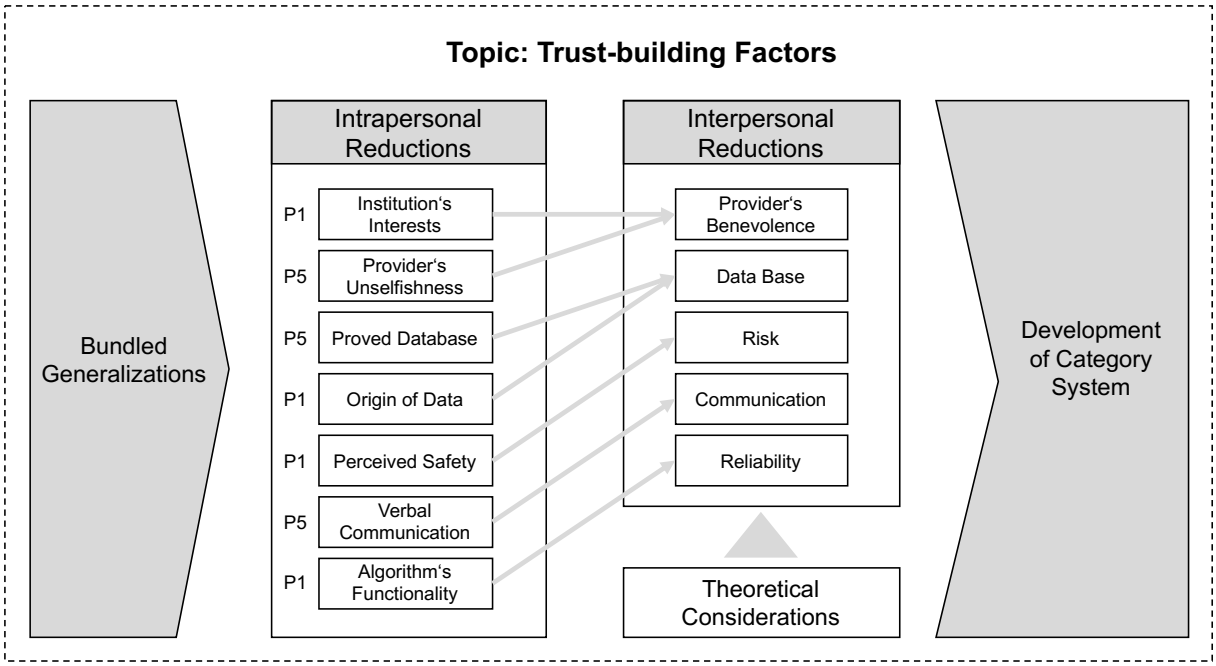


Figure 3. Process of aggregating generalizations from intrapersonal reductions to interpersonal reductions.

Note: The example shows reductions from P1 and P5.

The first category system was developed primarily based on our reduction system, including the inductively identified factors and theoretical considerations. Due to the continuing process of abstracting and simplifying the original material, there is a risk that the derived system does not represent the data adequately. Therefore, we tested our first category system on the data material by coding randomly selected interview passages. We noticed that some categories do not fit precisely, are redundant, or are not selective enough. In this case, we discussed ambiguities in our research group and referred to literature to enhance the accuracy of the coding system. After two revisions, we developed a more decidedly but also reduced system. Due to the systems' complexity and our intensive discussions in advance, we designated one expert in our team to conduct the coding. Following Mayring, he coded approximately 50% of the interviews (i.e., 13 of 27) to test the system's adequacy (Mayring, 2000, 2014). After further minor revisions, he coded the interviews with the final category system. Relevant

statements that have been made proactively by the participants were marked with an additional code ("highly present").

4 Results

In line with prior research on trust-building toward CAs, we identified internal factors (software-related) and external factors (user- and environment-related) as the highest-order factors of trust-building processes (Chi et al., 2021; Nordheim et al., 2019). Software-related factors influencing the development of trust include all aspects related to the specific system, like technical aspects or characteristics of the CA and its provider. In contrast, user-related factors entail the user's characteristics, such as attitudes toward CAs. Lastly, environment-related factors are external factors that influence trust-building despite their independence from the user and system, e.g., a technology's general establishment. Figure 4 shows a conceptualization of the initial trust-building process toward diagnostic CAs that we derived based on our results. In the following, we will provide critical insights and findings from our study, including a quantified overview of the main factors influencing trust at the end of the result section (see Table 2). A complete overview of all categories, including short definitions and examples, can be found in the Appendix.

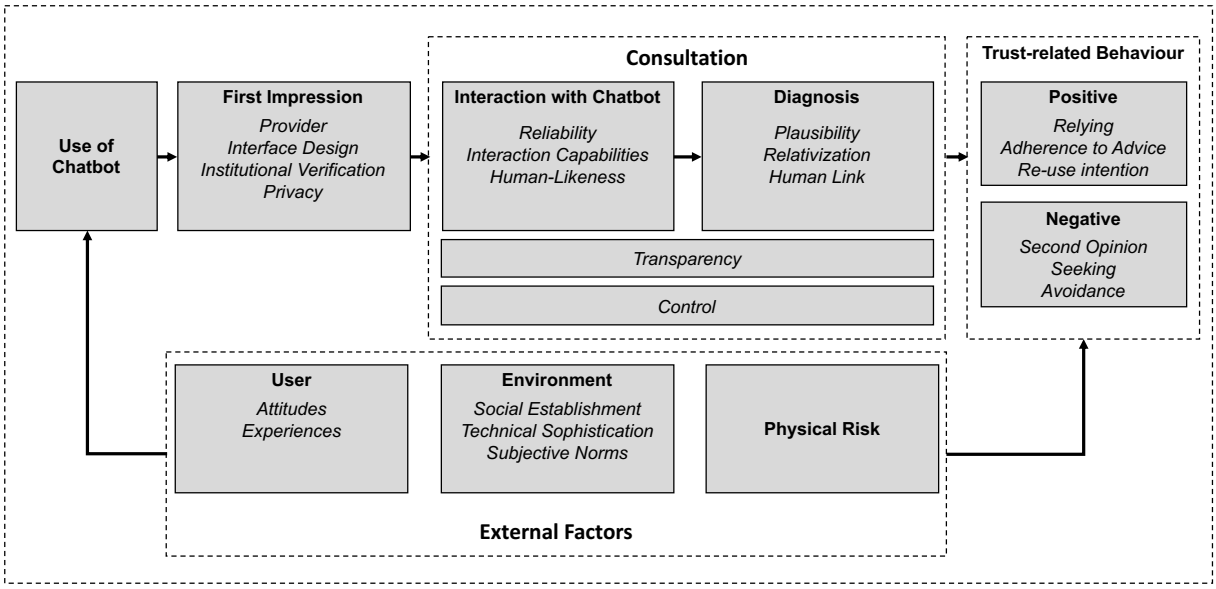


Figure 4. Process of initial trust-building toward diagnostic CAs.

4.1 Trust Influences Chatbot Adoption Even Before Initial Use

The development of trust toward diagnostic chatbots implies their actual usage. Even before the initial contact, trust toward the system plays an important role. We have noticed that attitudes toward diagnostic chatbots, subjective norms, perceived physical risk, and beliefs about the system's structural environment may influence the willingness to use the CA and the subsequent trust-building process. While several subjects were aware of a CA's advantages, some were generally skeptical of computers or CAs, thus thwarting adoption intentions and trust-building. Therefore, subjects indicated that recommendations from close friends or physicians would enhance trust and the likelihood of system adoption. Thus, subjective norms impact attitudes and initial trust-building processes.

> "It would actually give me confidence, even more confidence, if I also know that other people take recourse to it, that it is […] established and accepted by various instances: So, from the medical side […] and that friends of mine also recommend it based on empirical experience." (P21)

Considering the broader environment, the technologies' societal establishment and their technical sophistication also affect trust-building. Trust is more likely to arise when using diagnostic CAs is socially legitimized, and users believe using such a system is embedded in a safe and well-structured environment.

> "It also depends on how many people are using it. If I think it's popular, then I would think that it seems to be quite helpful, and if it's quite new, then maybe I wouldn't use it at first." (P22)

Once participants initially adopted the system, the quality of their own experiences become highly relevant. While negative experiences are likely to harm subsequent trust-building, positive experiences increase familiarity and reliability beliefs, thus changing initial trust to history-based trust, positively influencing the attitudes toward the system.

> "If I use it several times and it turns out to be reliable. If I use it every time and everything can be solved with the recommendations, trust would build up. That would be a time component." (P7)

Lastly, due to the high risk in medical consultations and its crucial role in trust-building processes, many interviewees stated that they would use the CA for preliminary tasks only or when the perceived uncertainty or the disease's severity is low. Correspondingly, subjects would renounce an additional telemedicine professional when they feel a certain level of safety. Our findings indicate that the level of trust toward the CA, respectively, the actual willingness to take risk, is limited by the level of perceived risk.

> "If it's something more severe, I wouldn't give a bot hundred percent trust. If it's something mild, then yes, I would say." (P27)

4.2 A Professional and Reputable First Impression Is the Fundament for Trust

Medical consultations are usually sensitive situations in which patients expect high confidentiality and professionality. In our interviews, we found several indications that users' trust toward the software is significantly impacted by evaluating its reputation and professionality. In particular, we have noted that users evaluate the CA's trustworthiness based on beliefs about the provider's benevolence and integrity. Various interviewees indicated that they expect a provider of integrity to act unselfish and not only for economic benefits. One subject even mentioned that a CA implemented by a profit-oriented company would be unusable.

> "As soon as I would get to know anything that they are working together in any form with different doctors […] in an economic way or with the pharmaceutical industry, then it would be absolutely no longer usable for me." (P10)

Also, the participants emphasized the relevance of the provider's competence and reputation. Since the CA relies on information computerized by its designer, users expect the provider to

have a professional medical background. Alternatively, institutional verifications (e.g., by healthcare organizations) may also indicate the CA's trustworthiness.

> "When there are any labels on it. The Federal Ministry of Health would be something like that. That it is checked, that would increase trust." (P3)

Not only the system's provider, but also the software itself can signal professionality by its interface design and the used language. In a complex scenario of a medical consultation, a trustworthy interface should convey professionality by deducing collected information in a clear and accessible design.

> "That's look and feel, so I think a professional presentation without too big or too shocking images is quite good, so a bit of sobriety in language and presentation is what I think is important. Especially when it is about medical topics." (P15)

Lastly, as confidentiality plays a significant role in doctor-patient relationships and online environments, some interviewees also referred to data privacy concerns since the CA collects and processes very personal data.

> "So, trust depends on whether I really have to log in with email or my name. With mail and name there would already be such a discomfort because the data are stored." (P18)

4.3 The Critical Phase of Trust-Building: The Interaction

The interaction itself determines the primary user experience in the usage of CAs. Considering the CA's ability to communicate with humans via natural language distinguishes it from other technological tools, a successful information exchange is vital. We divided a CA's interaction capabilities into five sub-sections: *understanding*, *specificity*, *coherence*, *expressivity,* and *language*. Beginning with the most fundamental one, the CA's understanding ability indicates the extent to which it can process entered messages correctly. Users see a successful information exchange as a requirement for the reliability of the diagnosis, while they also cast doubt on the general ability of CAs to understand humans. To feel confident in this

regard, users expect sufficient feedback that all entries have been understood correctly.

> "It would have made me feel more secure if it had said 'I understood.' But it just kept going, and I didn't know if it understood." (P7)

The second major factor is the conversation's structure, which also significantly impacts trust-building, i.e., the conversation's specificity and coherence. While specificity describes how detailed and conscientious the CA captured all symptoms, we define coherence as the conversation's inner logic resulting from a meaningful sequence of messages. Both factors convey the feeling of being understood while also making the CA's decision path and working steps more transparent.

> "Just the way it asked me the free questions showed that it understood my free text and was able to understand, okay, she has some urinary tract infection and I'll ask her specific questions. So, it definitely gave me confidence that it somehow has skills." (P8)

Third, as users expect a trustworthy CA to capture all symptoms precisely, the trust level also depends on how much users felt to have shared all relevant information. In contrast to specificity, expressivity describes the extent to which users can freely communicate all their concerns. If users felt they could not share all symptoms precisely, confidence in the diagnosis's accuracy may decrease.

> "...but if I can't answer some question and would like to say more about it and then the next question comes immediately, and I can't explain what pain I have exactly. Then I would trust less the result, so I need a way to add indications." (P6)

Lastly, the language style a CA is using may also impact trust-building. Users expect an appropriate language that is easy to understand, sober, linguistically correct, and polite. In this context, we noted that several subjects only expect the CA to adhere to basic interpersonal communication patterns by showing manners like greeting instead of behaving like a social actor. Ten interviewees (37.0%) did not expect any human-likeness from the CA, while another

fourteen (51.9%) would prefer a moderate level of human-like attributes (e.g., politeness). In contrast, only two (7.4%) mentioned human-likeness as a trust-enhancing factor. Interestingly, subjects also stated that human-likeness might even cause feelings of distrust.

> "However, in a medical context I don't want to joke around so much, but rather to be treated professionally. So, you can surely leave out that component for a chatbot. It should be polite and friendly and communicate clearly, use clear language, but otherwise I don't think social skills are that important." (P15)
>
> "Well, I don't know, this objectivity is very important for me when interacting with the bot and as soon as he communicates with me on another level, I would find it suspicious." (P8)

Thus, polite neutrality is preferred over intensely humanized agents. Instead of simulating fake emotions and empathy, it is promising to increase the conversation's perceived naturalness by providing specific and coherent queries.

> "I think the interaction character. That you have something that you're in exchange with. Mainly through the queries. That's what moves it toward the doctor, no that sounds too much, toward the human being." (P4)

4.4 Offering Transparency and Control During the Consultation Is Vital

Considering the high uncertainty resulting from the novelty of diagnostic CAs and the situation's sensitivity, our interviewees frequently indicated the relevance of a highly transparent handling of information and their desire to keep a sense of control. Starting with transparency, we constructed the four sub-categories *comprehensibility, justifications*, *source transparency*, and *database*. While interacting with CAs, subjects expected to be able to comprehend the decision path the CA is taking to come to the assessment. For instance, interviewees indicated that a consecutive and meaningful conversation flow would make the

decision path appear transparent. Correspondingly, users also require explanations at the end of the consultation on how the CA came to its assessment.

> "The questions were first roughly about where the pain is, for example [...]. And then it just went into detail and then, I can imagine, reached its goal like in a decision tree." (P11)

> "Maybe that it explains the course of the diagnosis at the end. That it explains how it came there. That would significantly increase trust." (P3)

In the diagnosis phase, users further expect the CA to justify its assessment, e.g., by disclosing statistics regarding the number of correct assessments or providing background information on the diagnosis. Since users expect a trustworthy CA to be based on well-founded data, they also wish for information about the used sources and the database's size. Especially in this high-risk situation, users tend to seek information that assures the reliability of the CA.

> "Maybe graphs, statistics or something like that would have been interesting. Or numbers - with these symptoms, one thousand people had the clinical picture cold. That would be another reliability aspect." (P22)

> "If you know who provides this bot, feeds it with data, what kind of people are behind it. If it is only computer scientists who have programmed it, without much medical expertise, it would be difficult." (P27)

The wish to be highly involved in the process is also reflected in statements where participants indicated their desire for control when informing about medical problems. In contrast to web research, a CA's guided assessment provides less control and self-reliance. Thus, subjects prefer the CA to reassure itself and ask for permission before the final assessment to enhance perceived control.

> "So, the bot is a bit more complex and gives me clear results, but on the other hand, I have done more during the internet research and can evaluate by myself. I have read that, and I think it is unlikely so I can kick it out therefore." (P20)
>
> "[…] for example, there was no question at the end for further comments. Instead, the result just appeared, and I don't know if I could go back again. Where it has decided to make a diagnosis, there is perhaps a lack of query whether it can make a diagnosis or whether there are further things." (P23)

4.5 Intention to Trust Depends on User's Attitudes Toward the Diagnosis

Even if the first impression and the perceived interaction quality have been positive, the intention to trust finally depends on the user's attitudes and feelings toward the diagnosis. Most importantly, the diagnosis' plausibility is vital for evaluating the assessment's trustworthiness. Diagnosis' plausibility indicates the extent to which the assessment makes sense to the user and how far expectations regarding the disease have been met. Although they are laymen, some participants indicated that they would only trust when they agree with the diagnosis. This is also supported by statements where participants indicated to renounce an additional telemedicine professional if the CA's diagnosis seems plausible.

> "So, depending on how it feels to me. If I also think it makes sense, I would wait and see." (P16)

Moreover, in this critical phase, participants expected a proper relativization of the CA's assessment. Due to the ambiguity of disease patterns, the missing physical examination, and user entries' dependency, the interviewees prefer several possible diagnoses over a single one. Accordingly, displaying more than one diagnosis enhances trust since the CA appears to be aware of disease patterns' complexity. Thus, a probabilistic assessment makes the CA appear reflected and competent. Further subjects pointed to the importance of openly communicating the CA's limitations. A CA that stresses its assessment's limited validity and refers to an

additional physician is perceived as honest and trustworthy. Besides the provider's evaluation, morality-related beliefs are also formed through the extent to which a CA shows an awareness of its constraints.

> "I found it very good. It gave you the feeling of competence. The easiest thing would be to say it's a cold. But to understand what is possible on the right and left, that gave me confidence." (P7)
>
> "If I notice that they are transparent about it [the limitations], it also increases my trust, because then I think that they are also making an effort and don't think they are ultimate." (P25)

These results further support our assumptions concerning the perceived control and self-reliance as subjects prefer the CA to make suggestions instead of final decisions. Although the participants perceive the CA's dependence on the user's subjective entries as a risk, some stated that they would rather trust their feelings than the assessment of the chatbot.

> "Yes, they should never put their own feeling behind technology. So, if you do feel that's not right, you should always let a feeling take precedence." (P18)

Main category	Sub-category	Int	%	HP	%	Seg
Physical risk		19	70%	13	48%	27
Interaction capabilities	Understanding	14	52%	13	48%	27
	Specificity	16	59%	8	30%	29
	Language	15	56%	5	19%	20
	Coherence	9	33%	4	15%	12
	Expressivity	7	26%	3	11%	12
Diagnosis' plausibility		15	56%	12	44%	19
Reliability		18	67%	11	41%	28
Transparency	Justifications	22	81%	10	37%	38
	Source transparency	12	44%	6	22%	16

	Comprehensibility	17	63%	4	15%	23
	Database	8	30%	4	15%	12
Provider	Purpose	16	59%	8	30%	22
	Competence	9	33%	2	7%	9
Relativization	Alternatives	11	41%	7	26%	12
	Limitations	5	19%	3	11%	6
Technical sophistication		12	44%	6	22%	15
Experiences		12	44%	6	22%	12
Institutional verification		8	30%	6	22%	11
Human link		7	26%	6	22%	8
Subjective norms		5	19%	4	15%	5
Interface design		7	26%	2	7%	7
Privacy		6	22%	2	7%	7
Control		11	41%	2	7%	18
Attitudes (negative)		6	22%	2	7%	10
Attitudes (positive)		8	30%	-	-	12

Table 2. Frequency of factors with codings in at least five interviews.

Notes: Int = interviews: total number of interviews with at least one coding; HP = highly present: number of persons who mentioned the corresponding factor proactively; Seg = segments: total number of coded segments.

4.6 A Comparison of Trust-Building Toward Diagnostic CAs and Physicians

One of our primary goals is to understand differences in trusting diagnostic CAs and trusting physicians to better understand the transferability of interpersonal or physician trust dimensions to the virtual environment. In our interviews, we asked the participants to explain the differences between trusting a CA and trusting a physician. To understand how far the CA is comparable to a physician, we asked the participants to locate the CA's assessment between web research and medical consultation. Ten of our interviewees (37.0%) were quite indifferent

and located the CA in the middle or with just a slight tendency. Eight participants (29.6%) associated the CA with web research, seven (25.9%) located it closer to the physician, and two interviewees (7.4%) made inconsistent statements. This result further emphasizes the hybrid nature of CAs, thus underlying the importance of considering diverse trust models.

While coding the documents and analysing our results, we noticed a considerable difference between the reasons to trust a CA and a physician. From the highest order perspective, trusting a CA is mainly driven by dimensions of CBT (e.g., performance), while trusting a physician (or telemedicine professional) is also based on factors related to ABT (e.g., empathy). Interestingly, some participants stated that they would not automatically trust the telemedicine professional more if their diagnosis differs from the CA. In this case, they would expect the telemedicine professional to justify him- or herself, although overall trust may be higher.

> "I would discuss with the telemedicine professional and ask why he is excluding the opinion from the bot." (P6)

Thus, we assume that subjects may have high levels of CBT toward the CA while trusting humans is also affect-based. To be more concrete, the interviewees saw the advantages of a CA in its vast database, objectivity and unselfishness, higher accuracy and lower error susceptibility, unlimited time capacities, the amount of stored information, and anonymity of the consultation (see Table 3).

> "And in the database are many more data than in a head and the assessment is rather objective then." (P6)

The majority of these factors are cognition-based and mainly associated with reliability aspects and trusting for good reasons. This is also reflected in the participant's desire for high transparency as they expect a CA to justify its outcomes and signal its reliability by numbers and verifications.

Among the mentioned advantages of a physician is the fulfilment of social needs like empathy, the conversation's flexibility, lower user dependency, their qualification, and practical knowledge (see Table 3). Many subjects emphasized the importance of not being left alone in critical situations, getting emotional support, and interacting with an agent who can share feelings and with whom one can identify.

> "But the impression I get when I communicate with a human is different. That I hear a voice or emotions that resonate, or this personal level, for example, "I always do it that way, too." You can identify with that, which is missing with chatbots […]." (P15)

Confirming this, one subject who has been connected to a telemedicine professional indicated that it was only the conversation with the human which reassured her and provided the feeling of a real diagnosis.

> "This made me feel confirmed and calmed down. And I felt that I now have a full diagnosis that I can trust far more than anything from the internet." (P7)

Some subjects also referred to personal and trustful relationships with their doctors, which a CA cannot replace. Besides, we observed one standard answering behaviour that participants could not express why they trust a human more than a CA. Some interviewees even became aware of the irrationality of having higher trust toward a biased and forgetting human being since the CA's assessment is objective and based on countless data sets.

> "But when I think about it more objectively, it [the chatbot] should be more trustworthy because it has much more data than a doctor. He [the doctor] has experience, but data is less error prone than human experience. But that is not my feeling, just the advantage when I think about it objectively." (P7)

This discrepancy was justified with (1) the habit of trusting humans more than machines, (2) subjective gut feelings, or (3) could not be justified at all.

> "I would rather trust the doctor with the complexity, even if that's imbecile. That's maybe a habitual pattern." (P17)

These findings further emphasize that ABT may be necessary for medical consultations. However, this may only hold true for severe and threatening diseases when there is a need for emotional support.

> "And I think that for many people this caring can't be solved by a bot. On the other hand, with cold symptoms nobody needs caring […]." (P19)

Even if only mentioned by three participants, another distinguishing factor is a human being's involvement. Since physicians may fear negative consequences if they misdiagnose a patient, it is assumed that they act more conscientiously than a CA. It is also expected that a human being adheres to certain ethical principles and has a sense of morality that a CA does not have.

> "There is also a medical oath that is made. I don't know exactly what it is about, but I could imagine that he treats patients to the best of his knowledge and belief, so they are committed to a trusting relationship." (P15)

Two difficulties we already addressed in our model's description were mentioned again in comparing a CA with a physician: the flexibility of the conversation and the user dependency. Many subjects stated that specificity and preciseness are significant differences between communicating with a CA and a human. The human-to-human interaction enables better expressivity, and a physician can ask more flexibly.

> "[…] you can speak freely. The chatbot has a pattern. It concentrates on the symptoms, the frequency, pain... You can't enter any additional information." (P11)

Lastly, the user dependency (i.e., the missing physical examination) distinguishes the virtual assessment from a physical consultation. First considered superficial, there is a structural hurdle to trust a CA due to the high dependency on users' subjective feelings and entries.

"With more severe things, I would still go to the doctor, because you don't know what is going on inside you. And to describe that certainly falsifies the results." (P18)

Category	Int	%	Seg
Flexibility	19	70%	32
User independency	18	67%	35
Empathy	16	59%	26
Social presence	12	44%	16
Habit	10	37%	12
Qualification	10	37%	10
Working experience	8	30%	10
Feeling	8	30%	14
Relationship	7	26%	12
Identification	4	15%	5
Morality sense	3	11%	4

Category	Int	%	Seg
Big data	8	30%	16
Accuracy	8	30%	9
Objectivity	6	22%	6
Information quantity	5	19%	7
Anonymity	4	15%	4
Thoroughness	3	11%	5

Table 3. Indicated reasons for trusting a medical professional (left) and trusting a diagnostic CA (right).

5 Complementary Study

5.1 Purpose

It was originally planned to conduct the main study with forty participants and to allocate them equally to the four conditions (see Table 1). However, due to contact restrictions caused by the COVID-19 pandemic, we had to stop data collection after twenty-seven interviews. We thus conducted an additional online study to test and verify the appropriateness of our derived categories and results. The main objective was to identify any shortcomings in our category system and to examine its applicability to slightly different healthcare contexts.

5.2 Method and Sample

To test the category system in a different and more realistic situation, we invited subjects online to interact with a Corona CA developed by the same provider. This CA was intended to assess the risk of a Corona infection based on a person's symptoms, visited locations, and personal contacts. After the participant interacted with the CA and received an individual risk assessment, participants were asked to indicate how much they would trust it compared to (1) a telemedicine professional's assessment and (2) information from websites using a seven-point semantic differential ("much lower (-3)" to "much higher (+3)"). To identify which factors would enhance trust when there is a lack, those who trusted more the telemedicine professional or the information from websites were asked for reasons in an open-ended question. After we had finished coding the main study's interviews, we applied the coding system to these open-ended questions.

The sample consists of 103 females (41.4%) and 143 males (57.4%), with three participants (1.2%) deciding not to report their gender. The mean age is M=29.01 years (SD=8.73).

5.3 Results

Regarding participants' trust level toward the CA's assessment, quantitative results show that users are quite indifferent regarding trusting the CA compared to (1) a telemedicine professional (M=0.00, SD=1.49) and (2) information from websites (M=0.35, SD=1.15). More specifically, n=78 subjects reported higher trust toward the human agent and n=44 felt higher trust toward website information. We then coded the answers from the open-ended questions and results provide support for our coding system's accuracy. We could code 85.9% (reasons for trusting more the telemedicine professional) and 72.7% (reasons for trusting more website information) of answers with existing codes.

Beginning with reasons to trust a human agent more, the importance of a natural and precise conversation was confirmed. With thirty-seven mentions, the higher conversation's flexibility is the most frequently indicated reason why subjects feel higher trust toward a telemedicine professional. Correspondingly, lacking specificity in interactions with CAs was mentioned another six times. Second, in nine responses, we noticed again that participants could not really argue why they trust the CA less, e.g., that it is just a feeling to trust a human more. Also, the lack of interpersonal aspects – mostly missing social presence – has been indicated by further seven subjects. Interestingly, rational arguments (e.g., the qualification of a telemedicine professional) were mentioned less often than subjective feelings and equally frequent as interpersonal aspects. This provides further evidence for the missing ABT in interactions with diagnostic CAs (see Table 4)

Regarding the reasons for trusting a website more than a CA, many subjects commented on transparency aspects. With sixteen mentions, missing source transparency is the main reason for trusting the CA less. Further four participants emphasized that the justification on websites is better, i.e., because they provide more background information. Besides transparency aspects, six participants indicated to trust the CA less than website information since they did not know the provider. Lastly, another six participants justified their higher trust in web research with the possibility to take control and to verify findings considering several sources (see Table 4).

Category	Int	%
Flexibility	37	47%
Feeling	9	12%
Interpersonal aspects	7	9%
Qualification	7	9%
Specificity	6	8%
Source transparency	3	4%

Category	Int	%
Source transparency	16	36%
Provider	6	14%
Control	6	14%
Justifications	4	9%
Other	12	27%

Provider	2	3%
Other	11	14%

Table 4. Indicated reasons for lower trust toward the Corona CA compared to a telemedicine professional (left) and information from websites (right).

6 Discussion

6.1 Trusting a Diagnostic CA Is Driven by Cognition

One of the study's key findings is the considerable difference between trusting a CA and trusting a physician. We found that the participants tend to experience higher trust levels toward a human not only because of their qualification but also for affect-based reasons. Although subjects were aware of their irrationality, the high levels of CBT toward the CA were insufficient to create an equal level of trust. However, some interviewees would not automatically trust a human agent more than a CA when they could not justify a divergent assessment. Therefore, we state that CBT may be necessary but not sufficient to develop trust in sensitive situations. Since subjects tended to trust human agents more than CAs, we assume that ABT is vital in medical consultations. In this sense, these results also provide evidence for the high role-based trust people hold toward physicians, which they do not show toward diagnostic CAs yet (LaRosa and Danks, 2018).

Although it seems promising to identify appropriate antecedents of ABT, some limitations are to be considered. First, our results show that users do not expect the fulfilment of social needs by a chatbot. We state that factors which are necessary for a reliable diagnosis are more important in trusting CAs than emotional ones which may even evoke distrust. Also, considering that diagnostic CAs would be used for mild diseases only, the necessity of ABT is questionable. Second, we strongly encourage researchers to adhere to ethical principles when investigating antecedents of ABT toward artificial entities. The uniqueness of interpersonal relationships characterized by human warmth and emotional attachment should always be

respected. Diagnostic CAs should be seen as a helpful supplement for existing health systems rather than medical professionals' replacements (Powell, 2019). Moreover, cognitive processing and high involvement may also have a protective function in risk situations. Trust toward autonomous systems should not be as great as possible but appropriate (Lee and See, 2004). Therefore, we argue that increasing CBT by enhancing the technology's reliability, transparency, and accuracy may be a more reasonable way to enhance user's trust. Subsuming, we state that diagnostic CAs should not be intended to fulfil social needs. Instead, they provide physicians the opportunity to engage in their interpersonal relationships with patients as autonomous agents can take over routine tasks like data collection (Waizenegger et al., 2020).

6.2 CAs Should Be Humanized Carefully

Although humanizing CAs is considered promising, we could not find evidence for a general superiority of humanized CAs concerning trust development. Instead of imitating human emotions, a CA should show reasonable and credible sociality. Even though researchers point to the high potential of implementing empathy to health CAs (Vaidyam et al., 2019), we argue that the level and type of social cues should be considered to avoid backfiring effects. In our results, there is only little evidence that human-likeness enhances perceived trust toward diagnostic CAs. Expressed emotions may be perceived as gimmicks that are not appropriate in severe situations, highlighting that the effectiveness of anthropomorphic cues depends on contextual circumstances (Blut et al., 2021). In contrast to the UVM, we assume that human-likeness does not necessarily evoke feelings of eeriness but reduces a healthcare agent's credibility. Another hurdle for anthropomorphizing agents can be found when considering the "Theory of Mind", which describes that human beings can form beliefs about another person's cognitive states (Premack and Woodruff, 1978). In the case of CAs, users are aware that chatbots are mindless, thus having no emotions. Hence, they are not able to identify with the

CA or to take its perspective. This missing possibility to empathize with the CA may also explain the users' desire for high transparency concerning underlying functionalities.

However, although participants did not expect empathy from a CA, the importance of communicative aspects shows similarities to research on trust toward physicians. From their doctors, patients expect to be respected, active listening, high involvement in decision-making processes, and the possibility to share all concerns (Meakin, 2002). Our interviewees also expressed the wish to share all relevant information and expected the CA to ask questions conscientiously. Although research already pays attention to the linguistic design of chatbot interactions, the impact of a CA's interaction capabilities on trust-building is still an under-investigated area. We argue that reasonable expressivity and an overall more natural conversation flow could positively influence human-likeness without the usage of anthropomorphic design cues (Go and Sundar, 2019). Expressivity may especially be necessary in situations with a high need for detail and specificity, i.e., complex, and human-like tasks.

6.3 Users' Desire to Keep Control in Interactions with CAs

According to Epley's "Three-Factor Theory of Anthropomorphism", one reason for humans' tendency to anthropomorphize non-human agents is the desire to take control (Epley et al., 2007). Our interviews contain hints on different coping strategies rather than anthropomorphizing the CA, e.g., relying on own feelings or expecting justifications and comprehensibility. Even though users attribute specific competencies to the CA, many perceive it as an expert self-information tool instead of an autonomous trustee. Instead of a final diagnosis, users expect to receive several possible assessments to keep control. Furthermore, the desire to keep control is also reflected in statements indicating that users would like to lead the conversation, i.e., by the wish for high expressivity. We suggest two proposals to explain this phenomenon. First, we argue that users compensate for the missing affective component and reciprocity of trust by relying on their own. Second, humans may overestimate their

abilities compared to a CA due to overconfidence bias (Moore and Healy, 2008) induced by the belief in general human supremacy over technology (Stein and Ohler, 2017). By relying on themselves, subjects can uphold their illusion of keeping control. Therefore, future studies should consider moderating effects of human supremacy beliefs on trust-building processes toward CAs. It is further to be investigated under what circumstances users of CAs adopt different coping strategies to keep a sense of control.

Moreover, keeping control may reduce risk perception, thus reducing the need to trust (Corritore et al., 2003). When keeping control, users can take responsibility for their actions, while it is unclear who takes responsibility for the CA in case of misdiagnosis. Since CAs have no free will or a sense of moral, it is questionable how far CAs can be seen as trustees at all since morality-related trust dimensions require a certain kind of consciousness (Corritore et al., 2003). Following previous research, our interviewees evaluate a CA's morality-related characteristics mainly by beliefs about its designer (Akter et al., 2011; Følstad et al., 2018; Söllner et al., 2012). Although a CA can also convey its integrity by specific design features, it only partially fulfils the agent's role requirements. Further considering the missing ABT, we argue that there is no reciprocal trusting relationship between patient and CA (McAllister, 1995).

6.4 Trusting Diagnostic CA Is Suspect to Change

We also conclude that developing trust toward diagnostic CAs is a matter of time since own experiences and social establishment may change initial trust to history-based or institution-based trust (Kramer, 1999; McKnight and Chervany, 2001b; McKnight et al., 2002). With the increasing dissemination of AI in healthcare, CAs and physicians' role attributions could change (LaRosa and Danks, 2018). Furthermore, many CAs are at the early stages of development, so that users may not ascribe them the necessary competencies for medical consultations yet. This has also been an issue in our study, as there were some initial technical

difficulties. It is foreseeable that future technologies will be more sophisticated and have capabilities that cannot be anticipated today. Thus, trusting healthcare CAs is subject to rapid environmental, technological, and individual changes.

6.5 Practical Implications

The results from our studies revealed that trust-building toward diagnostic CAs is a complex procedure influenced by user-, environment-, and software-related factors. While the first two can barely be influenced immediately by software designers, the latter opens up opportunities of actively enhancing the trustworthiness of healthcare CAs. Therefore, Table 5 provides an overview of major software-related trust-building factors, associated challenges, and suggested solutions that may be considered by software designers.

Trust-building factor	Challenge	Solutions
Purpose	Users may question the intentions of the CA's provider and the purpose of the system.	• Communicating patient-centered intentions • Avoiding advertisements • Respect patients' privacy and emphasize data protection efforts
Reliability	Users may fear the software's performance, reliability, and accuracy.	• Ensuring quality of data base • Ensuring appropriate NLP capabilities • Openly communicating information about provider • Providing external verifications
Interface design	Usability aspects and the software's appearance may harm the trustworthiness of the CA if not appropriate.	• Implementing reduced and clear user interface • Use of a language that is easy to understand • Avoiding too exciting design elements
Interaction capabilities	Conversations with chatbots often feel static and inflexible which is a problem in complex medical consultations.	• Asking detailed queries • Giving users room for expression to enhance perceived control • Enhancing conversation's naturalness • Implementing politeness and moderate human-likeness

Transparency	Complex algorithm-based CAs represent black boxes for users since they cannot emphasize with them, which may be a problem in high-risk situations.	・Making the decision path comprehensible ・Providing justifications for the assessment ・Using numbers and statistics to substantiate argumentation ・Providing information about sources
Relativization	Due to its limited possibilities to examine a patient, a CA is not able to make a final diagnosis.	・Emphasizing the assessment's limited validity ・Displaying a probabilistic assessment showing several diagnoses

Table 5. Practical recommendations for designing trustworthy diagnostic CAs.

7 Future Research Directions and Limitations

Since the aim of our studies was to create a holistic understanding of trust-building processes toward diagnostic CAs, we can only hypothesize specific relations between the dimensions. We thus recommend applying structural equation modelling in quantitative study designs to validate our findings. For instance, we assume that certain factors may be associated with specific dimensions and antecedents of trust, e.g., a professional-looking interface may enhance competence and integrity perceptions, but not necessarily benevolence. It could also be revealing to investigate which factors impact trust directly and which only reduce risk and uncertainty, thus lowering the necessary level of trust. In this course, we encourage researchers to discuss further the role of transparency and predictability in trusting AI. Considering traditional perspectives on trust, we argue that transparency enables the trustor to control the trustee while predictability eliminates risk, both contradicting common conceptualizations of trusting relationships (Mayer et al., 1995). We further assume that certain factors like correct spelling may not increase trust-building as they represent hygiene factors. Thus, users will not notice the presence of correct spelling but only its absence, which may lead to negative feelings and distrust. Theoretical considerations and research on trust and distrust support this assumption, postulating that both concepts are associated with different antecedents, mental states, and consequences (Dimoka, 2010; Lewicki et al., 1998; McKnight and Chervany, 2001a;

McKnight et al., 2004). Lastly, future studies should consider different levels of risk since it significantly impacts the trustee's willingness to depend. Drivers of intention to trust may vary in dependence on perceived risk. Consequently, perceived risk may function as a moderator for the necessity of ABT, i.e., CBT may be sufficient in situations of low need for emotional support.

Our study further has some limitations to be considered when interpreting the findings and developing consecutive research. First, the participants from our main study did not suffer from an actual disease. Future studies should be conducted with real patients to enhance the involvement and validate our findings in a risk situation. However, since we conducted in-depth interviews, we could create an atmosphere of high involvement, allowing the participants to reflect deeply on their statements. Furthermore, our complementary study took place in a situation where risk was at least latently present. Second, the main study's sample is quite homogenous consisting of young and mainly female adults who tend to be familiar with novel technologies. Antecedents of trust could be different for other user groups like elderlies. However, the sample of our additional online study has been more heterogenous and we barely found fundamental differences in trust-building factors. Nevertheless, future studies should be conducted with more heterogenous samples since demographics may impact trust-building towards technological artifacts (Pak et al., 2014; Shao et al., 2019). Third, our findings should always be considered for the specific context of healthcare. Trusting mechanisms and the role of ABT and CBT may vary depending on the domain and task for which a chatbot is used. Thus, it is likely that users' expectations and trust drivers toward a chatbot used for customer service purposes differ from those presented in this paper (Blut et al., 2021; Følstad et al., 2018). Lastly, the findings from our studies may be biased by features of the specific software we applied. Some of the participant's comments could have been impacted by prominent design features of the system, its shortcomings, and highlights. For instance, subjects commented

positive on the claims emphasizing the assessment's limited validity. Thus, the practical recommendations shown in Table 5 should be confirmed by future work. Although some findings might differ slightly in studies using other diagnostic CAs, we framed our study material quite generic and key findings are consistent with results from previous research on trust.

8 Conclusion

Compared to prior research, we considered diagnostic CAs a hybrid between physician and self-information tools since they show social and technological characteristics. By directly comparing trust-building toward diagnostic CAs and human medical professionals, we have created a deep understanding of underlying mechanisms and found evidence that interpersonal trust approaches are only partly appropriate to explain the development of trust toward diagnostic CAs. Concerning physician trust approaches, we noted that communicative aspects are equally important in interactions with chatbots, while subjects do not expect a CA to fulfil social needs. Thus, we conclude that it is promising to enhance the interaction's perceived naturalness instead of focusing on human-like cues. Also, transparent handling of information and justifications is a critical factor since trust mainly arises cognitively underlying the technical character of CAs. Both transparency and a more natural conversation may substantially impact trust development and thus intention to adopt. We also conclude that there is evidence for general human supremacy beliefs, which thwart trust-building toward autonomous agents. As ABT lacks in interactions with CAs, we encourage researchers to investigate its relevance in the specific healthcare context under constant consideration of ethical aspects. Finally, it is important to consider environmental circumstances, technological developments, and user-related factors when investigating trust-building toward diagnostic CAs. Since those factors may change over time, trust-building processes should be depicted continually.

References

Akter, S., D'Ambra, J., Ray, P., 2011. Trustworthiness in mhealth information services: An assessment of a hierarchical model with mediating and moderating effects using partial least squares (PLS). Journal of the American Society for Information Science and Technology. 62 (1), 100–116. https://doi.org/10.1002/asi.21442.

Al-Natour, S., Benbasat, I., Cenfetelli, R. T. 2006. The role of design characteristics in shaping perceptions of similarity: The case of online shopping assistants. Journal of the Association for Information Systems. 7 (12), 821–861. https://doi.org/10.17705/1JAIS.00110.

Al-Natour, S., Benbasat, I., Cenfetelli, R. T. 2010. Trustworthy virtual advisors and enjoyable interactions: Designing for expressiveness and transparency, in: Proceedings of the 18th European Conference on Information Systems (ECIS 2010). AIS, Atlanta, 116.

Anderson, L. A., Dedrick, R. F. 1990. Development of the trust in physician scale: A measure to assess interpersonal trust in patient-physician relationships. Psychological Reports. 67 (3), 1091–1100. https://doi.org/10.2466/pr0.1990.67.3f.1091.

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.

Barber, B. 1983. The logic and limits of trust. Rutgers University Press, New Brunswick.

Benbasat, I., Wang, W. 2005. Trust in and adoption of online recommendation agents. Journal of the Association for Information Systems. 6 (3), 72–101. https://doi.org/10.17705/1jais.00065.

Bickmore, T. W., Caruso, L. B., Clough-Gorr, K. 2005. Acceptance and usability of a relational agent interface by urban older adults, in: CHI '05 Extended Abstracts on

Human Factors in Computing Systems (CHI EA '05). ACM, New York, pp. 1212–1215.

Bickmore, T. W., Schulman, D., Sidner, C. 2013. Automated interventions for multiple health behaviors using conversational agents. Patient Education and Counseling. 92 (2), 142–148. https://doi.org/10.1016/j.pec.2013.05.011.

Blut, M., Wang, C., Wünderlich, N. V., Brock, C. 2021. Understanding anthropomorphism in service provision: A meta-analysis of physical robots, chatbots, and other AI. Journal of the Academy of Marketing Science. https://doi.org/10.1007/s11747-020-00762-y.

Buchanan, A. 1988. Principal/agent theory and decision making in health care. Bioethics. 2 (4), 317–333. https://doi.org/10.1111/j.1467-8519.1988.tb00057.x.

Chi, O. H., Jia, S., Li, Y., Gursoy, D. 2021. Developing a formative scale to measure consumers' trust toward interaction with artificially intelligent (AI) social robots in service delivery. Computers in Human Behavior. 118, 106700. https://doi.org/10.1016/j.chb.2021.106700.

Cho, J. 2006. The mechanism of trust and distrust formation and their relational outcomes. Journal of Retailing. 82 (1), 25–35. https://doi.org/10.1016/j.jretai.2005.11.002.

Chua, R. Y. J., Ingram, P., Morris, M. W. 2008. From the head and the heart: Locating cognition- and affect-based trust in managers' professional networks. Academy of Management Journal, 51 (3), 436–452. https://doi.org/10.5465/amj.2008.32625956.

Corritore, C. L., Kracher, B., Wiedenbeck, S. 2003. On-line trust: Concepts, evolving themes, a model. International Journal of Human-Computer Studies. 58 (6), 737–758. https://doi.org/10.1016/S1071-5819(03)00041-7.

Cramer, H., Evers, V., Ramlal, S., van Someren, M., Rutledge, L., Stash, N., Aroyo, L., Wielinga, B. 2008. The effects of transparency on trust in and acceptance of a content-based art recommender. User Modeling and User-Adapted Interaction, 18 (5), 455–

496. https://doi.org/10.1007/s11257-008-9051-3.

de Visser, E. J., Monfort, S. S., McKendrick, R., Smith, M. A. B., McKnight, P. E., Krueger, F., Parasuraman, R. 2016. Almost human: Anthropomorphism increases trust resilience in cognitive agents. Journal of Experimental Psychology: Applied, 22 (3). 331–349. https://doi.org/10.1037/xap0000092.

Denecke, K., Tschanz, M., Dorner, T. L., May, R. 2019. Intelligent conversational agents in healthcare: Hype or hope?. Studies in Health Technology and Informatics. 259, 77–84. https://doi.org/10.3233/978-1-61499-961-4-77.

Dimoka, A. 2010. What does the brain tell us about trust and distrust? Evidence from a functional neuroimaging study. MIS Quarterly. 34 (2), 373–396. https://doi.org/10.2307/20721433.

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.

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.

Fitzpatrick, K. K., Darcy, A., Vierhile, M. 2017. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (woebot): A randomized controlled trial. JMIR Mental Health. 4 (2), e19. https://doi.org/10.2196/mental.7785.

Flavián, C., Guinalíu, M., Gurrea, R. 2006. The role played by perceived usability, satisfaction and consumer trust on website loyalty. Information and Management. 43 (1), 1–14. https://doi.org/10.1016/j.im.2005.01.002.

Følstad, A., Nordheim, C. B., Bjørkli, C. A. 2018. What makes users trust a chatbot for

customer service? An exploratory interview study, in: Internet Science (INSCI 2018). Springer, Basel, 194–208. https://doi.org/10.1007/978-3-030-01437-7_16.

Friedman, B., Khan, P. H., Howe, D. C. 2000. Trust online. Communications of the ACM, 43 (12), 34–40. https://doi.org/10.1145/355112.355120.

Galletta, A. 2013. Mastering the Semi-Structured Interview and Beyond: From Research Design to Analysis and Publication. NYU Press, New York.

Gefen, D., Benbasat, I., Pavlou, P. 2008. A research agenda for trust in online environments. Journal of Management Information Systems. 24 (4), 275–286. https://doi.org/10.2753/MIS0742-1222240411.

Gefen, D., Karahanna, E., Straub, D. W. 2003. Trust and TAM in online shopping: An integrated model. MIS Quarterly. 27 (1), 51–90. https://doi.org/10.2307/30036519.

Gioia, D. A., Corley, K. G., Hamilton, A. L. 2013. Seeking qualitative rigor in inductive research: Notes on the Gioia Methodology. Organizational Research Methods. 16 (1), 15–31. https://doi.org/10.1177/1094428112452151.

Glikson, E., Woolley, A. W. 2020. Human trust in artificial intelligence: Review of empirical research. Academy of Management Annals. 14 (2), 627–660. https://doi.org/10.5465/annals.2018.0057.

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.

Grodal, S., Anteby, M., Holm, A. L. 2020. Achieving rigor in qualitative analysis: The role of active categorization in theory building. Academy of Management Review. https://doi.org/10.5465/amr.2018.0482.

Hall, M. A., Dugan, E., Zheng, B., Mishra, A. K. 2001. Trust in physicians and medical institutions: What is it, can it be measured, and does it matter? The Milbank Quarterly.

79 (4), 613–639. https://doi.org/10.1111/1468-0009.00223.

Hall, M. A., Zheng, B., Dugan, E., Camacho, F., Kidd, K. E., Mishra, A., Balkrishnan, R. 2002. Measuring patients' trust in their primary care providers. Medical Care Research and Review. 59 (3), 293–318. https://doi.org/10.1177/1077558702059003004.

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

Hillen, M. A., Postma, R.-M., Verdam, M. G. E., Smets, E. M. A. 2017. Development and validation of an abbreviated version of the trust in oncologist scale—the trust in oncologist scale–short form (TiOS-SF). Supportive Care in Cancer. 25 (3), 855–861. https://doi.org/10.1007/s00520-016-3473-y.

Ivbijaro, G. O., Enum, Y., Khan, A. A., Lam, S. S.-K., Gabzdyl, A. 2014. Collaborative care: Models for treatment of patients with complex medical-psychiatric conditions. Current Psychiatry Reports. 16 (11), 506. https://doi.org/10.1007/s11920-014-0506-4.

Jeffrey, D. 2016. Empathy, sympathy and compassion in healthcare: Is there a problem? Is there a difference? Does it matter? Journal of the Royal Society of Medicine, 109 (12), 446–452. https://doi.org/10.1177/0141076816680120.

Komiak, S. X. Y., Benbasat, I. 2006. The effects of personalization and familiarity on trust and adoption of recommendation agents. MIS Quarterly. 30 (4), 941–960. https://doi.org/10.2307/25148760.

Kramer, R. M. 1999. Trust and distrust in organizations: Emerging perspectives, enduring questions. Annual Review of Psychology. 50 (1), 569–598. https://doi.org/10.1146/annurev.psych.50.1.569.

Laranjo, L., Dunn, A. G., Tong, H. L., Kocaballi, A. B., Chen, J., Bashir, R., Surian, D.,

Gallego, B., Magrabi, F., Lau, A. Y. S., Coiera, E. 2018. Conversational agents in healthcare: A systematic review. Journal of the American Medical Informatics Association. 25 (9), 1248–1258. https://doi.org/10.1093/jamia/ocy072.

LaRosa, E., Danks, D. 2018. Impacts on trust of healthcare AI, in: Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society (AIES' 18). ACM, New York, pp. 210–215. https://doi.org/10.1145/3278721.3278771.

Laumer, S., Maier, C., Gubler, F. 2019. Chatbot acceptance in healthcare: Explaining user adoption of conversational agents for disease diagnosis, in: Proceedings of the 27th European Conference on Information Systems (ECIS 2019). AIS, Atlanta, 88.

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, J., Moray, N. 1992. Trust, control strategies and allocation of function in human-machine systems. Ergonomics. 35 (10), 1243–1270. https://doi.org/10.1080/00140139208967392.

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

Lewicki, R. J., McAllister, D. J., Bies, R. J. 1998. Trust and distrust: New relationships and realities. Academy of Management Review. 23 (3), 438–458. https://doi.org/10.5465/amr.1998.926620.

Lewis, J. D., Weigert, A. 1985. Trust as a social reality. Social Forces. 63 (4), 967–985. https://doi.org/10.2307/2578601.

Liu, B., Sundar, S. S. 2018. Should machines express sympathy and empathy? Experiments with a health advice chatbot. Cyberpsychology, Behavior, and Social Networking. 21

(10), 625–636. https://doi.org/10.1089/cyber.2018.0110.

Locke, K., Feldman, M., Golden-Biddle, K. 2020. Coding practices and iterativity: Beyond templates for analyzing qualitative data. Organizational Research Methods. https://doi.org/10.1177/1094428120948600.

Luhmann, N. 1979. Trust and Power. Wiley, Chichester.

Malhotra, N., Kim, S., Agarwal, J. 2004. Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. Information Systems Research. 15 (4), 336–355. https://doi.org/10.1287/isre.1040.0032.

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

Mayring, P. 2000. Qualitative content analysis. Forum Qualitative Social Research. 1 (2), 20. https://doi.org/10.17169/fqs-1.2.1089.

Mayring, P. 2014. Qualitative content analysis: Theoretical foundation, basic procedures and software solution. Beltz, Klagenfurt.

McAllister, D. J. 1995. Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. Academy of Management Journal. 38 (1), 24–59. https://doi.org/10.2307/256727.

McKnight, D. H., Chervany, N. L. 2001a. While trust is cool and collected, distrust is fiery and frenzied: A model of distrust concepts, in: AMCIS 2001 Proceedings. AIS, Atlanta, 171. http://aisel.aisnet.org/amcis2001/171.

McKnight, D. H., Chervany, N. L. 2001b. Trust and distrust definitions: One bite at a time, in: Falcone, R., Singh, M., Tan, Y.-H. (Eds.), Trust in Cyber-societies. Springer, Berlin, Heidelberg, pp. 27–54. https://doi.org/10.1007/3-540-45547-7_3.

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

McKnight, D. H., Cummings, L. L., Chervany, N. L. 1998. Initial trust formation in new organizational relationships. Academy of Management Review. 23 (3), 473–490. https://doi.org/10.5465/amr.1998.926622.

McKnight, D. H., Kacmar, C., Choudhury, V. 2004. Dispositional trust and distrust distinctions in predicting high- and low-risk internet expert advice site perceptions. e-Service Journal. 3 (2), 35–58. https://doi.org/10.2979/esj.2004.3.2.35.

Meakin, R. 2002. The "Medical Interview Satisfaction Scale" (MISS-21) adapted for british general practice. Family Practice. 19 (3), 257–263. https://doi.org/10.1093/fampra/19.3.257.

Montenegro, J. L. Z., da Costa, C. A., da Rosa Righi, R. 2019. Survey of conversational agents in health. Expert Systems with Applications. 129, 56–67. https://doi.org/10.1016/j.eswa.2019.03.054.

Moore, D. A., Healy, P. J. 2008. The trouble with overconfidence. Psychological Review. 115 (2), 502–517. https://doi.org/10.1037/0033-295X.115.2.502.

Mori, M., MacDorman, K., Kageki, N. 2012. The uncanny valley. IEEE Robotics and Automation Magazine. 19 (2), 98–100. https://doi.org/10.1109/MRA.2012.2192811.

Muresan, A., Pohl, H. 2019. Chats with bots: Balancing imitation and engagement, in: CHI '19 Extended Abstracts on Human Factors in Computing Systems (CHI EA '19). ACM, New York, pp. 1–6. https://doi.org/10.1145/3290607.3313084.

Nass, C., Steuer, J., Tauber, E. R. 1994. Computers are social actors, in: CHI '94 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '94), ACM, New York, pp. 72–78. https://doi.org/10.1145/191666.191703.

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

Nundy, S., Montgomery, T., Wachter, R. M. 2019. Promoting trust between patients and physicians in the era of artificial intelligence. Journal of the American Medical Association. 322 (6), 497–498. https://doi.org/10.1001/jama.2018.20563.

Pak, R., McLaughlin, A. C., Bass, B. 2014. A multi-level analysis of the effects of age and gender stereotypes on trust in anthropomorphic technology by younger and older adults. Ergonomics. 57 (9), 1277–1289. https://doi.org/10.1080/00140139.2014.928750

Pearson, S. D., Raeke, L. H. 2000. Patients' trust in physicians: Many theories, few measures, and little data. Journal of General Internal Medicine. 15 (7), 509–513. https://doi.org/10.1046/j.1525-1497.2000.11002.x.

Powell, J. 2019. Trust me, i'm a chatbot: How artificial intelligence in health care fails the turing test. Journal of Medical Internet Research. 21 (10), e16222. https://doi.org/10.2196/16222.

Premack, D., Woodruff, G. 1978. Does the chimpanzee have a theory of mind? Behavioral and Brain Sciences. 1 (4), 515–526. https://doi.org/10.1017/S0140525X00076512.

Provoost, S., Lau, H. M., Ruwaard, J., Riper, H. 2017. Embodied conversational agents in clinical psychology: A scoping review. Journal of Medical Internet Research. 19 (5), e151. https://doi.org/10.2196/jmir.6553.

Qiu, L., Benbasat, I. 2009. Evaluating anthropomorphic product recommendation agents: A social relationship perspective to designing information systems. Journal of Management Information Systems. 25 (4), 145–182. https://doi.org/10.2753/MIS0742-1222250405.

Reeves, B., Nass, C. I. 1996. The media equation: How people treat computers, television, and new media like real people and places. CSLI Publications, Cambridge University Press.

Rempel, J. K., Holmes, J. G., Zanna, M. P. 1985. Trust in close relationships. Journal of Personality and Social Psychology. 49 (1), 95–112. https://doi.org/10.1037/0022-3514.49.1.95.

Riedl, R., Mohr, P. N. C., Kenning, P. H., Davis, F. D., Heekeren, H. R. 2014. Trusting humans and avatars: A brain imaging study based on evolution theory. Journal of Management Information Systems. 30 (4), 83–114. https://doi.org/10.2753/MIS0742-1222300404.

Rubin, H. J., Rubin, I. S. 2011. Qualitative Interviewing: The Art of Hearing Data, third ed. SAGE, Thousand Oaks.

Shao, Z., Zhang, L., Li, Xaotong, Guo, Y. 2019. Antecedents of trust and continuance intention in mobile payment platforms: The moderating effect of gender. Electronic Commerce Research and Applications. 33, 100823. https://doi.org/10.1016/j.elerap.2018.100823

Siau, K., Wang, W. 2018. Building trust in artificial intelligence, machine learning, and robotics. Cutter Business Technology Journal. 31 (2), 47–53.

Skjuve, M., Brandzaeg, P. B. 2019. Measuring user experience in chatbots: An approach to interpersonal communication competence, in: Internet Science (INSCI 2018). Springer, Basel, 113–120. https://doi.org/10.1007/978-3-030-17705-8_10.

Skjuve, M., Haugstveit, I. M., Følstad, A., Brandtzaeg, P. B. 2019. Help! Is my chatbot falling into the uncanny valley? An empirical study of user experience in human-chatbot interaction. Human Technology. 15 (1), 30–54. https://doi.org/10.17011/ht/urn.201902201607.

Söllner, M., Hoffmann, A., Hoffmann, H., Wacker, A., Leimeister, J. M. 2012. Understanding the formation of trust in IT artifacts, in: Proceedings of the 33rd International Conference on Information Systems (ICIS 2012). AIS, Atlanta, 11.

Stein, J.-P., Liebold, B., Ohler, P. 2019. Stay back, clever thing! Linking situational control and human uniqueness concerns to the aversion against autonomous technology. Computers in Human Behavior. 95, 73–82. https://doi.org/10.1016/j.chb.2019.01.021.

Stein, J.-P., Ohler, P. 2017. Venturing into the uncanny valley of mind—The influence of mind attribution on the acceptance of human-like characters in a virtual reality setting. Cognition, 160, 43–50. https://doi.org/10.1016/j.cognition.2016.12.010.

Strauss, A., Corbin, J. 1998. Basics of qualitative research: Techniques and procedures for developing grounded theory, second ed. SAGE, Thousand Oaks.

Thatcher, J. B., McKnight, D. H., Baker, E. W., Arsal, R. E., Roberts, N. H. 2011. The role of trust in postadoption IT exploration: An empirical examination of knowledge management systems. IEEE Transactions on Engineering Management. 58 (1), 56–70. https://doi.org/10.1109/TEM.2009.2028320.

Thom, D. H., Campbell, B. 1997. Patient-physician trust: An exploratory study. The Journal of Family Practice. 44 (2), 169–176.

Toader, D.-C., Boca, G., Toader, R., Măcelaru, M., Toader, C., Ighian, D., Rădulescu, A. T. 2019. The effect of social presence and chatbot errors on trust. Sustainability. 12 (1), 256. https://doi.org/10.3390/su12010256.

Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., Torous, J. B. 2019. Chatbots and conversational agents in mental health: A review of the psychiatric landscape. The Canadian Journal of Psychiatry. 64 (7), 456–464. https://doi.org/10.1177/0706743719828977.

VERBI GmbH 2020. *MAXQDA Plus 2020* (Release 20.2.1) [Computer Software]. VERBI

GmbH.

Waizenegger, L., Seeber, I., Dawson, G., Desouza, K. 2020. Conversational agents -exploring generative mechanisms and second-hand effects of actualized technology affordances, in: Proceedings of the 53rd Hawaii International Conference on System Sciences (HICSS 2020). AIS, Atlanta, pp. 5180–5189. https://doi.org/10.24251/HICSS.2020.636.

Wang, W., Benbasat, I. 2016. Empirical assessment of alternative designs for enhancing different types of trusting beliefs in online recommendation agents. Journal of Management Information Systems. 33 (3), 744–775. https://doi.org/10.1080/07421222.2016.1243949.

Wang, W., Qiu, L., Kim, D., Benbasat, I. 2016. Effects of rational and social appeals of online recommendation agents on cognition- and affect-based trust. Decision Support Systems. 86, 48–60. https://doi.org/10.1016/j.dss.2016.03.007.

Wang, W., Siau, K. 2018. Living with artificial intelligence—Developing a theory on trust in health chatbots, in: Proceedings of the 16th Annual Pre-ICIS Workshop on HCI Research in MIS. AIS, San Francisco.

Yogeeswaran, K., Złotowski, J., Livingstone, M., Bartneck, C., Sumioka, H., Ishiguro, H. 2016. The interactive effects of robot anthropomorphism and robot ability on perceived threat and support for robotics research. Journal of Human-Robot Interaction. 5 (2), 29–47. https://doi.org/10.5898/JHRI.5.2.Yogeeswaran.

Appendix

Trust influencing factors							
Highest-order factor	Main category	Sub-category	Explanation	Example	Segments	Interviews	HP
A1: Internal Factors	A1.1 Reliability	-	Users expect the chatbot's algorithm to have appropriate functions and complexity to ensure reliability and to avoid misdiagnosis.	[…] you don't know whether the algorithm of the bot is correct or whether it is a misinterpretation. Maybe the symptoms are not interpreted correctly, for example. (P16)	28	18	11
	A1.2 Interface Design	-	The graphical user interface (GUI) is the only touchpoint between user and system. To signalize trustworthiness, a diagnosis chatbot's interface should be reduced and clear.	That's look and feel, so I think a professional presentation without too big or too shocking images is quite good, so a bit of sobriety in language and presentation is what I think is important. Especially when it is about medical topics. (P15)	7	7	2
	A1.3 Relativization	A1.3.1 Limitations	An honest chatbot should clearly communicate its restrictions and limitations.	If I notice that they are transparent about it [the limitations], it also increases my trust, because then I think that they are also making an effort and don't think they are ultimate. (P25)	6	5	3
		A1.3.2 Alternatives	Due to the ambiguity of symptoms and missing physical examination, users expect the chatbot to display several possible diseases instead of coming to an ultimate diagnosis.	I found it [display of several diseases] very good. It gave you the feeling of competence. The easiest thing would be to say it's a cold. But to understand what is possible on the right and left, that gave me confidence. (P7)	12	11	7

	A1.4 Interaction Capabilities	A1.4.1 Understanding	The chatbot's ability to understand users' entries are crucial for a reliable assessment. Users expect sufficient feedback that all messages have been understood to increase trust.	It would have made me feel more secure if it had said I understood. But it just kept going and I didn't know if it understood. (P7)	27	14	13
		A1.4.2 Language	The chatbot's language style influences its perception and should fit the task. A trustworthy diagnostic CA is expected to use professional, neutral and correct language.	However, in a medical context I don't want to joke around so much, but rather to be treated professionally. So you can safely leave out that component for a chatbot. It should be polite and friendly and communicate clearly, use clear language, but otherwise I don't think social skills are that important. (P15)	20	15	5
		A1.4.3 Specificity	The specificity of a conversation describes how detailed and thorough the chatbot captures symptoms.	It [level of trust] depends on how determined the questions would have been asked. (P3)	29	16	8
		A1.4.4 Coherence	Textual coherence defines the inner logic and the consecutiveness of a conversation. A conversation is coherent when messages refer to previous messages and conversational content.	The reaction of the chatbot itself, how far do I feel its statements are correct, to what extent has it understood me. If I now enter this or that symptom and then something else comes up. If I say I have a cough and it asks about lower back pain. That the flow of information is stringent. (P17)	12	9	4
		A1.4.5 Expressivity	Users expect to be able to share all symptoms and concerns decidedly and freely since the assessment's reliability depends on the accuracy of users' entries.	[...] but if I can't answer some question and would like to say more about it and then the next question comes directly, and I can't explain what pain I have exactly. Then I would trust less the result, so I need again a way to add indications. (P6)	12	7	3

		A1.4.6 Communication Medium	The interaction type between chatbot and users influences trust perceptions. Verbal interaction is more natural, thus trustworthy.	[…] maybe it would give me more trust if I could talk to it instead of writing. (P22)	2	2	1
	A1.5 Diagnosis Plausibility	-	The diagnosis' plausibility indicates how far the assessment makes sense to the user, meets his or her expectations, and fits the entered symptoms.	Results. So, if something comes out that I think is very unlikely, I wouldn't trust it. (P20)	19	15	12
	A1.6 Transparency	A1.6.1 Comprehensibility	A comprehensible decision path enables the users to understand the chatbot's inner workings and functions better.	Maybe that it explains the course of the diagnosis at the end. That it explains how it came there. That would significantly increase trust. (P3)	23	17	4
		A1.6.2 Justifications	Explanations, background information and justifications make the assessment more transparent and understandable for the users (i.e., by showing statistics).	But I would have liked to see a more precise display. That would give me more confidence. […] Obviously, one also got additional information, which I would consider good. (P4)	38	22	10
		A1.6.3 Source Transparency	Since users expect the chatbot to be based on valid and reliable data, transparency concerning data source is important for trust-building.	I constantly have in the back of my mind, where does this come from? With a trustworthy link, I would trust immediately. (P1)	16	12	6
		A1.6.4. Level	Too much transparency concerning data or functionalities reduces trust since users may be confused.	No, I like it the way it is. Everything else would confuse me, I think. […] That would be too much information. (P8)	4	4	-
		A1.6.5 Database	The reliability and accuracy of a chatbot's assessment depend on the data base's size.	I think to a certain point I trust it because of the large database. (P6)	12	8	4

	A1.7 Provider	A1.7.1 Competence	Since its designer computerizes the chatbot's knowledge, it is expected to have a professional medical background.	If you know who runs this bot, feeds it with data, what kind of people are behind it. If it is only computer scientists who have programmed it, but without much medical expertise, that would be difficult. (P27)	9	9	2
		A1.7.2 Purpose	A medical chatbot should not be intended to earn profits, so users expect an unselfish provider who does not act for their economic benefits. Thus, diagnostic chatbots from pharmaceutical companies would be avoided.	As soon as I would get to know anything that they are working together in any form with different doctors […] in an economic way or with the pharmaceutical industry, then it would be absolutely no longer usable for me. (P10)	22	16	8
	A1.8 Privacy	-	Since users have to provide sensitive and personal data to the chatbot, privacy concerns may inhibit trust-building.	So, trust depends on whether I really have to log in with email or my name. With mail and name, there would be a discomfort, because the data are stored [...]. (P18)	7	6	2
	A1.9 Human Link	-	The possibility to get handed over to a human medical professional enhances trust.	Telemedicine includes the term "medicine", so there is a basic level of trust. This is also transferred to the chatbot. (P15)	8	7	6
	A1.10 Human-Likeness	-	Although too much human-likeness may backfire, latently present social cues hold the potential to enhance trust.	I don't know if that [social skills] would make it seem more human, more trustworthy. That could probably influence it. (P13)	2	2	1
	A1.11 Control	-	Users show the desire to keep control, act self-reliant, and trust own feelings, which illustrates the limits of trusting a chatbot.	Yes, you should never put your own feeling behind technology, so if you do feel that's not right, you should always let a feeling take precedence. (P18)	18	11	2

A2: External Factors	A2.1 Subjective	A2.1.1 Experiences	Trust arises with repeated usage over time, thus replacing initial trust with knowledge-based trust.	If I use it several times and it turns out to be reliable. If I use it every time and everything can be solved with the recommendations, trust would build up. That would be a time component. (P7)	12	12	6
		A2.1.2 Subjective Norms	Recommendations from the immediate social environment enhance trust since subjective norms are crucial for adopting new technologies.	It would actually give me confidence […] if I also know that other people take recourse to it, that it is […] established and accepted by various instances. So, from the medical side, that they rely on it and that they recommend it… that friends of mine also recommend it based on empirical experience. (P21)	5	5	4
		A2.1.3.1 Attitudes (negative)	Users might have generally negative attitudes towards technology or CAs what might inhibit trust-building.	I don't have total confidence in computers. (P10)	10	6	2
		A2.1.3.2 Attitudes (positive)	Users might have generally positive attitudes towards technology or CAs what might facilitate trust-building.	[…] I think the idea is basically good, but I also know that I am a very digitization-open person and just prefer that in many areas and I find it much more useful than manual and conventional methods […]. (P21)	12	8	-
	A2.2 Institution	A2.2.1 Verification	Verifications from authorities, health organizations or other credible institutions enhance trust.	When there are any labels on it. The Federal Ministry of Health would be something like that. That it is checked, that would increase trust. (P3)	11	8	6

		A2.2.2 Technical Sophistication	Trust in the chatbot depends on users' beliefs concerning technology's sophistication and the technological environment's safety.	And I can also imagine that at some day […] a chatbot or AI will sit next to the doctor and the doctor will vote with his knowledge against the bot. But I think it will still take a little while. (P12)	15	12	6
		A2.2.3 Social Establishment	The acceptance of chatbots in the broader social environment (i.e., society) influences trust-building.	It also depends on how many people are using it. If I think it's popular, then I would think that it seems to be quite helpful, and if it's quite new, then maybe I wouldn't use it at first. (P22)	6	4	2
	A2.3 Physical Risk	-	The willingness to trust a chatbot strongly depends on the perceived risk and the disease's severity. Diagnostic chatbots would only be used when the disease is mild or moderate.	If it's something more severe I wouldn't give a bot hundred percent trust. If it's something mild, then yes, I would say. (P27)	27	19	13

Trust comparison						
Highest-order factor	Main category	Sub-category	Explanation	Example	Segments	Interviews
B1: Trusting Chatbot	B1.1 Anonymity	-	The conversation with a chatbot is more anonymous than a conversation with a human. Thus, there is no fear of social judgement or feelings of shame.	[…] it is good that you're not so embarrassed about things that make you feel uncomfortable, what you are with the doctor, even if you shouldn't be. You're always completely honest. (P8)	4	4
	B1.2 Objectivity	-	A chatbot is more objective since it does not underly cognitive biases. Furthermore, it acts not selfish since it has no intentions or a free will.	What it can do better would certainly be impartiality. When I think back to specific doctor visits, different doctors have their area on which they make their diagnosis […]. Or that the bot [does not] benefit from prescribing this drug. (P4)	6	6
	B1.3 Big Data	-	The information quantity in artificial neural networks and databases can significantly exceed the knowledge of a human. From an informational perspective, a chatbot can have much more knowledge than a single physician.	I also believe that you can build up a large amount of data there. That can become more competent than a doctor. (P5)	16	8
	B1.4 Accuracy	-	The diagnosis of a chatbot may be even more accurate than a physician's one since algorithms are less error prone.	Bots are also not as error prone, so the diagnosis would be more accurate. (P1)	9	8
	B1.5 Thoroughness	-	In comparison to often overloaded physicians, chatbots have unlimited time capacities so that they really can take care of a patient.	Because it is objective and doctors often don't ask about the symptoms in such detail and he asked me in great detail and I answered very honestly and that is only possible because he is a machine and not a human being. (P8)	5	3

	B1.6 Information	-	A chatbot can provide more background information, statistical figures and alternative diagnoses.	[…] and I get a more profound result in comparison to a doctor. Not just one diagnosis […], but several cases. (P6)	7	5
B2: Trusting Human	B2.1 Knowledge	B2.1.1 Experience	In practice, a physician makes many experiences resulting in a broad network of implicit knowledge. Thus, a physician can adopt his or her knowledge to grasp even unknown situations.	But the doctor has his assessment, so, his experience, and asks questions. The bot could also do this, but the bot may not have the experience or the subjectivity that the doctor has. (P18)	10	8
		B2.1.2 Qualification	By having studied medicine, a physician can verifiably prove their skills and competencies to diagnose a patient that a chatbot cannot do.	The doctor, he may be more trustworthy and reliable, you can trust that he has the education. (P24)	10	10
	B2.2 Habit		While the use of chatbots is novel, consulting a doctor is very familiar. Therefore, trust in a physician is higher due to learning processes and habits.	I would trust a doctor even more, because one is probably still used to it. […] Chatbots are something completely new in the medical field, and people are not yet familiar with them. (P27)	12	10
	B2.3 Morality		Physicians are expected to have a sense of moral and to adhere to the Hippocratic Oath. Furthermore, a physician has to fear the consequences of his or her actions while it is unclear who takes responsibility for the chatbot.	There is also a medical oath that is taken, I don't know exactly what it is about, but I could imagine that he treats patients to the best of his knowledge and belief, so they are committed to a trusting relationship. (P15)	4	3
	B2.4 Interpersonal Aspects	B2.4.1 Identification	Since a physician is a human being, patients can emphasize with him or her and vice versa. Thus, a physician can comprehend situations and concerns better than a chatbot.	That's what the telemedicine specialist has now taken on and also the identification component. "I also know from myself" or "I always do it like that". You can't believe that without question from a chatbot. (P15)	5	4

	B2.4.2 Social Presence	The interaction with a human conveys the feeling of not being left alone. This can be crucial in medical consultations since patients may feel sick, anxious or uncertain.	First of all, the trust from the feeling. It is simply a different situation when someone really listens to you than when you are alone with yourself. (P22)	17	13
	B2.4.3 Empathy	Beside technical competencies, empathy and emotional support are equally important in medical consultations. Since chatbots cannot feel emotions or compassion, the empathic concern is missing in the interaction with chatbots.	I don't know exactly how it would be for me, but many people care about empathy. […] I think the doctor has a calming function, which the chatbot can't do so well at first glance. (P19)	26	16
	B2.4.4 Relationship	Some patients tend to build up trustful relationships with their doctors, which a chatbot cannot replace.	He can definitely build trust and closeness to you as a person in a different way. Maybe you've known your doctor for ten years. (P14)	12	7
	B2.5 Flexibility	- The interaction with a human being is more flexible and more individual than with a chatbot. Human-to-human communication thus enables precise expressions and dedicated queries.	[…] you can ask questions about what's unclear. You can simply get rid of that in a personal conversation. If he says, "Drink a lot", I can ask: "Which tea?" Then he can answer in detail. (P4)	32	19
	B2.6 Feelings	- Higher trust towards a human is a matter of feeling. Even if there are no concrete reasons or justifications to have higher trust towards a human being, users tend to experience higher trust levels towards a physician.	But when I think about it more objectively, it [the chatbot] should be more trustworthy because it has much more data than a doctor. He [the doctor] has experience, but data is less prone to error than human experience. But that is not my feeling, just the advantage when I think about it objectively. (P7)	14	8

	B2.7 User Independency	-	A chatbot is highly dependent on the user's entries since it is not capable of physical examinations. Thus, a chatbot's assessment is strongly impacted by the quality of the user's subjective descriptions.	With more severe things, I would still go to the doctor, because you don't know what is going on inside you. And to describe that certainly falsifies the results. (P18)	35	18
B3 Classification	B3.1 Web Research	-	Statements indicating that participants associate the chatbot with web research.	Closer to the search engine. Very close. If the search engine is a 1 and the doctor is a 10, then the bot is a 2. (P4)	18	14
	B3.2 Indifferent	-	Statements indicating that participants were indifferent about locating the chatbot in the direction of web research or physician.	I would even say in the middle. (P26)	10	10
	B3.3 Physician	-	Statements indicating that participants associate the chatbot with physicians.	It is definitely better […] than the web research, but it is quite not like the doctor. I wouldn't put it in the middle either, maybe even a little closer to the doctor. (P22)	12	12

Sociality of chatbot					
Main category	Sub-category	Explanation	Example	Segments	Interviews
D1: Relationship	-	Personal recognition and references to previous interactions make the chatbot appear more social. Also individualized components may positively influence relationship-building.	On the other hand, if it could actually learn in the interaction. And links could be created. It would be like having a personal relationship. […] At the next consultation, the patient is asked, "How was your cold?" That one has a more personal reference. (P4)	3	3
D2: Human-Likeness	D2.1 Positive	Statements indicating that participants are open towards human-like cues.	You could give it a personal touch. You could personalize this bot. "Hi I'm Medidoc" That you feel you're not talking to the computer. (P5)	12	7
	D2.2 Moderate	Statements indicating that participants prefer the chatbot to show only a moderate level of human-likeness.	I think that people respond positive to it. It doesn't have to call me by my first name and give me compliments, but that it has a kind of human element. On a formal level. I would find that pleasant. (P7)	17	14
	D2.3 Negative	Statements indicating that participants have negative attitudes towards human-like cues.	It [human-likeness] would annoy me. I know that it is a bot. I just want professional advice. (P1)	32	17
D3: Communication Medium	-	Since verbal communication is the most habitual way of human interaction, a voice-based conversation with the chatbot positively affects perceived sociality. Furthermore, free text-entries enable a more flexible thus natural interaction compared to given answering options.	[…] because it is a chatbot and not a voice. If you compare it to Siri, for example, you still have a voice and therefore a different bond. You imagine something different. Because it is only text and clicking […]. (P13)	6	6
D4: Manners		Users expect the chatbot to adhere to certain manners like greeting, farewell and politeness.	It should be polite and friendly and communicate clearly, use clear language, but otherwise I don't think social ability is that important. (P15)	20	16

D5: Queries		Detailed queries and questions have a positive effect on the perceived sociality as it evokes the feeling of a natural interaction and caring.	I think the interaction character. That you have something with which you are in exchange. Mostly through the queries. That's what moves him in the direction of the doctor, no that sounds too much, to the human being. (P4)	9	7

PART III: PAPER 2

ARTIFICIAL EMPATHY IN HEALTHCARE CHATBOTS: DOES IT FEEL AUTHENTIC?

Fact Sheet Paper 2

Title	Artificial Empathy in Healthcare Chatbots: Does It Feel Authentic?
Author	Lennart Seitz
Year	2024
Citation	Seitz, L. (2024). Artificial empathy in healthcare chatbots: Does it feel authentic? *Computers in Human Behavior: Artificial Humans*, *2*(1), 100067.
DOI	https://doi.org/10.1016/j.chbah.2024.100067

Abstract

Implementing empathy to healthcare chatbots is considered promising to create a sense of human warmth. However, existing research frequently overlooks the multidimensionality of empathy, leading to an insufficient understanding if artificial empathy is perceived similarly to interpersonal empathy. This paper argues that implementing experiential expressions of empathy may have unintended negative consequences as they might feel inauthentic. Instead, providing instrumental support could be more suitable for modeling artificial empathy as it aligns better with computer-like schemas towards chatbots. Two experimental studies using healthcare chatbots examine the effect of *empathetic* (feeling with), *sympathetic* (feeling for), and *behavioral-empathetic* (empathetic helping) vs. *non-empathetic* responses on perceived warmth, perceived authenticity, and their consequences on trust and using intentions. Results reveal that any kind of empathy (vs. no empathy) enhances perceived warmth resulting in higher trust and using intentions. As hypothesized, *empathetic,* and *sympathetic* responses reduce the chatbot's perceived authenticity suppressing this positive effect in both studies. A third study does not replicate this backfiring effect in human-human interactions. This research thus highlights that empathy does not equally apply to human-bot interactions. It further introduces the concept of "perceived authenticity" and demonstrates that distinctively human attributes might backfire by feeling inauthentic in interactions with chatbots.

Keywords: empathy; chatbot; healthcare; authenticity; anthropomorphism

1 Introduction

Driven by the rapid developments in AI and language processing systems, people increasingly interact with virtual assistants like chatbots [1,2]. Chatbots are text-based dialogue systems emulating an interpersonal interaction to serve clients in numerous service domains such as hospitality, retailing, and even healthcare. Since it is foreseeable that interactions with chatbots and other virtual assistants will further increase, researchers argue that the landscape of service provision could be fundamentally changed as bots are expected to supplement or even substitute human agents [3,4].

Despite their increasing capabilities that have been impressively demonstrated by the launch of ChatGPT, interactions with current generations of chatbots often feel mechanical compared to interpersonal interactions [5]. Therefore, chatbots are frequently equipped with social cues, e.g., by giving them names, avatars, or complex communication capabilities [6,7,8]. One major challenge is the missing empathy and warmth that are essential in interpersonal interactions, especially in sensitive environments like healthcare provision where emotional support and trustful relationships are inevitable [9,10]. Imbuing human-bot interactions in such service domains with a sense of empathy is hence considered promising to compensate for the lack of human touch and to facilitate trust-building and using intentions [11,12,13].

However, simply concluding that empathy is equally applicable to interactions with chatbots could be premature for two related reasons. First, existing research on the implementation of empathy to chatbots and other virtual assistants has mostly considered empathy unidimensional (i.e., empathy vs. no empathy) [14]. As empathy is a complex multidimensional concept consisting of cognitive, affective, and behavioral dimensions, this might have led to an incomplete understanding whether humans react in the same way to artificial empathy as they do to interpersonal empathy. Second, due to the insufficient conceptual separation, research has ignored that the different dimensions of empathy may vary

in their suitability for modeling artificial empathy. Cognitive and affective empathy require mindfulness and experiential capabilities as they describe the ability to feel, share, recognize, or understand the mental state of another [15]. These capabilities are, however, considered one of the key distinctions between humans and machines [16,17]. Expressions of empathy in which a chatbot pretends to be able to feel or understand emotion might therefore interfere with computer-like schemas and mechanistic stereotypes towards chatbots hence appearing rather fake than genuine [9,18,19]. Up to this point, there is barely research examining potential drawbacks when social cues feel not authentic, even though there is an increasing number of research articles pointing out potential backfiring effects of humanizing chatbots and other virtual assistants [e.g., 20,21].

To address this research gap, this paper presents two experimental studies using chatbots responding either *empathetic* (feeling *with* the user), *sympathetic* (feeling *for* the user), or *behavioral-empathetic* (empathetic helping). In the selection of an appropriate and realistic service environment, I decided to conduct the studies in a healthcare setting in which empathy is an essential social skill [22]. Drawing on the concept of "anthropomorphism" [23], the related "Social Response Theory" [24], and the "Stereotype Content Model" [25], the present research hypothesizes that all kinds of empathy (vs. no empathy) enhance a chatbot's perceived warmth resulting in a higher willingness to trust and, ultimately, using intentions. In contrast, drawing on "Mind Perception Theory" [17] and the concept of authenticity [26], this research further hypothesizes that *empathetic* and *sympathetic* responses reduce a chatbot's perceived authenticity since chatbots are not believed to have the required cognitive or affective capabilities to feel *with* or *for* a patient [17,27]. This loss in perceived authenticity is hypothesized to suppress the positive effect on the willingness to trust and using intentions since perceived authenticity is vital for evaluating someone's credibility and trustworthiness [28,29]. In contrast, this suppressing effect is not hypothesized for *behavioral-empathetic*

responses as the chatbot does not self-disclose cognitive or affective states but provides instrumental support which might align better with computer-like schemas towards chatbots. Hence, the provision of instrumental support might represent a more authentic way of designing artificial empathy. A third study replicates the research model in an interpersonal communication situation to test if the backfiring effect only occurs in interactions with chatbots and not humans. This aims to substantiate the argument that the potential loss in perceived authenticity by *empathetic* and *sympathetic* responses can be attributed to their interference with computer-like schemas and mechanistic stereotypes towards chatbots.

Subsuming, this paper extends previous knowledge and theory in two ways. First, it shows that not all dimensions of empathy are equally applicable to interactions with chatbots. It therefore provides a critical perspective on anthropomorphism and the "Social Response Theory" by uncovering different reactions to the same social cues in chatbots vs. humans. Second, as a major novelty, it is among the first papers demonstrating that implementing distinctively human attributes to chatbots (i.e., the ability to feel *with* or *for* another) might backfire by feeling inauthentic. It thus takes up the emerging research stream identifying boundary conditions of humanizing bots [18,30,31,32].

The paper is structured as follows. First, it provides a comprehensive literature review on previous research and the theories the research model is based on. Afterwards, the empirical part presents the three studies separately, including a short individual discussion for each study's findings. A general discussion follows in which the theoretical contributions, managerial implications, as well as limitations and future research avenues are presented. The paper closes with a short conclusion summarizing the key findings and their relevance for research and practice.

2 Conceptual Background

2.1 Perceiving Warmth in Chatbots and Anthropomorphism

Technological innovations open new opportunities to use chatbots for complex tasks that require a sense of empathy, e.g., healthcare provision [33]. Since it is not foreseeable that bots will be able to feel emotion soon [2,9], they are sometimes imbued with artificial empathy to make conversations feel more human-like and to facilitate relationship-building [34,35]. Chatbots can, for instance, send emotional supportive messages, use emojis, or express their compassion with a client. Across service domains, research has found several positive effects of artificial empathy on user experience and behavior (see Table 1). For instance, empathetic agents are perceived more likeable, trustworthy, and emotionally supportive [35,36]. Furthermore, users interacting with empathetic agents and AI show higher levels of satisfaction and usage persistence [37,38,39]. Feeling a sense of empathy in bots can also enhance users' mood after social exclusion and even lead to a reduction in depressive symptoms [40,41].

Paper	Year	Study domain	Cue	Modality	Key findings
Klein, Moon, & Picard [66]	2002	Mood induction experiment	Empathy, sympathy, expressivity, emotional support	Text	Participants show higher persistence in playing a frustrating game when they receive emotional support from an agent. No such effect was found for an agent allowing users to vent their feelings.
Bickmore & Picard [39]	2005	Healthcare	Empathy, sympathy, expressivity, facial expressions, gesture	Multimodal	A relational agent expressing empathy is more likeable, creates stronger bonds, and enhances users' willingness to continue usage.
Brave, Nass, & Hutchinson [36]	2005	Entertainment	Empathy, sympathy, facial expressions	Multimodal	Empathetic emotions enhance an agent's likeability, its trustworthiness, and felt support.
Nguyen & Masthoff [65]	2009	Emotional support	Empathy, sympathy, expressivity, facial expressions, gestures	Multimodal	Empathy in an agent enhances perceived enjoyment, perceived caring, and overall attitudes. The effects are stronger for personified vs. non-personified agents.
Liu & Sundar [35]	2018	Healthcare	Empathy (cognitive vs. affective), sympathy	Text	Expressions of affective empathy and sympathy (vs. cognitive empathy) enhance perceived support.

De Gennaro, Krumhuber, & Lucas [40]	2020	Mood induction experiment	Sympathy, emojis	Text	Individuals who have experienced social exclusion report enhanced mood after interacting with an empathetic agent.
Gelbrich, Hagel, & Orsingher [38]	2021	Emotional support	Emotional support	Text	An emotionally supportive digital assistant enhances satisfaction and behavioral persistence.
Lv, Yang, Qin, Cao, & Xu [37]	2022	Hospitality	Empathy, sympathy	Multimodal	Receiving a highly empathetic response from an AI after service failure enhances using intentions.

Table 1. Study overview on artificial empathy in various types of bots, virtual assistants, and AI.

In explaining these positive reactions to artificial empathy (or human-likeness in general), researchers frequently refer to humans' social nature and the resulting tendency to perceive and treat non-human agents like social actors [7,23]. This phenomenon is also known as "anthropomorphism" that is particularly elicited by recognizing social cues in an entity leading to the mindless adoption of social rules [23,42]. Hence, people react in a similar way to social cues and behavior in non-human entities, e.g., users might mirror a virtual agent's smile [43]. This general human tendency to anthropomorphize is also theorized in computer science and information systems research by the "Social Response Theory" [24] and the related "Computers Are Social Actors" paradigm [44].

Given the premise that chatbots are perceived as social actors, receiving empathetic responses might have similar effects like in interpersonal communication. Since empathy is closely associated with concepts like feeling with another, compassion, and pro-social behavior, empathetic individuals are evaluated to be caring and warm [15,45]. According to the "Stereotype Content Model", perceived warmth is – besides perceived competence – one of the core dimensions of social perception and emanates from assuming good intents in another [25]. It is therefore considered vital in interpersonal relations, especially in evaluating someone's trustworthiness. Congruently, many well-established trust models account for the importance of perceived warmth in trust-building processes by introducing the related concept of

"benevolence" which is defined as the extent to which someone is believed to have good intents and that is found to be a major predictor of trusting intentions [29,46,47]. Regarding the relevance of perceived warmth in human-chatbot interactions, previous research lends credence for the predictive power of perceived warmth in facilitating trust-building and – ultimately – using intentions [38,48,49]. A chatbot that provides a sense of warmth, e.g., by behaving empathetic, might therefore appear to be more trustworthy than a chatbot that responds mechanically [12,36]. Furthermore, feeling a sense of humanness might generally enhance the willingness to trust the chatbot due to human's inherent sensitivity and preference for any kind of human-like cues. This might particularly apply in high-risk and intimate service environments like healthcare in which trust is inevitable [33]. Users who do not trust a software system because they consider it unreliable or feel in other ways uncomfortable while using it are unlikely to continue usage [11,50,51].

To conclude, this research hypothesizes that a chatbot responding with any kind of empathy (i.e., *empathetic*, *sympathetic*, or *behavioral-empathetic* responses) enhances perceived warmth resulting in a higher willingness to trust and, ultimately, higher using intentions.

H1: A healthcare chatbot responding with a sense of empathy (empathetic, sympathetic, behavioral-empathetic) is perceived warmer than a healthcare chatbot responding non-empathetic.

H2: Perceived warmth in a healthcare chatbot is positively related to the willingness to trust the chatbot.

H3: The willingness to trust a healthcare chatbot is positively related to the intention to use the chatbot.

2.2 The Multidimensional Concept of Empathy

Besides replicating the well-studied positive consequences of empathy and warmth in chatbots, this research primarily aims at moving towards a more nuanced perspective on artificial empathy. In the past, researchers and practitioners have used various cues to design artificial empathy what might be rooted in the concept's ambiguous definition and conceptualization [52,53]. However, most of the existing studies on artificial empathy have not explicitly accounted for the multidimensionality of empathy, i.e., they either examined only one specific cue of empathy, or they combined different cues and compared it to a non-empathetic agent (see Table 1). In the following, this paper provides a comprehensive overview of widely recognized conceptualizations of empathy in interpersonal communication and psychology. Moreover, it showcases how the different dimensions of empathy can be implemented to chatbots.

From a high-level perspective, literature divides empathy into *cognitive* and *affective* empathy [53]. Cognitive empathy is associated with the ability to take someone's perspective and to accurately recognize emotional states. It is therefore closely related to the "Theory of Mind" referring to humans' ability to ascribe mental states, intentions, emotions, or beliefs to others that might deviate from own ones [15,54]. A requirement for cognitive empathy is thus the ability for self-other distinction that can only be found in higher organisms with complex cognition such as humans [15,53].

Affective empathy, in contrast, does not entail deeper information processing as it is an automatically elicited emotional response to another one's emotion and can therefore be considered a less sophisticated form of empathy [15,55]. Affective empathy is often described in terms like "emotion sharing" and "emotional contagion", meaning that someone mirrors and experiences the same emotion as an observed one [53].

In designing artificial empathy, researchers have made use of both visual cues (e.g., facial expressions) and verbal cues (e.g., emotional statements) to model cognitive and affective empathetic responses. In case of text-based chatbots, cognitive or affective empathy is usually communicated through verbal phrases like "I understand your anxiety" or "I could imagine how annoying that can be" [35]. Obviously, clearly separating cognitive from affective empathetic responses in written communication is challenging. An empathetic response to a message implies that the reader has decoded the message's content accurately before coding an appropriate empathetic response both involving cognitive processes [56]. Affective empathetic reactions, in contrast, usually manifest in emotional responses that instinctively spill out, e.g., starting to cry when seeing a person cry [55]. If someone aims at expressing verbally that s/he feels with another (i.e., affective empathy), s/he might use phrases like "I can really empathize with your fears" for self-disclosing experienced emotions. Due to the difficulty in clearly separating both, this research considers written expressions of cognitive or affective empathy a sender's intent to signalize having empathized with the situation or emotion of another. In the following, this paper uses the generic term *empathetic* responses in referring to such messages.

Besides *empathetic* responses, researchers have made use of *sympathetic* responses for modeling artificial empathy. Sympathy is associated with compassion and describes feeling sorry for someone and might occur in interactions with a person in a demanding situation. Due to its close association with empathy, literature is inconclusive about the relationship between sympathy and empathy as both concepts are even confused [15,22,53]. Some scholars consider sympathy a part of empathy as it is a cognitive or affective response to another person's mental state or situation [15]. Researchers hence also refer to the term "empathic concern" in describing sympathy [52]. However, sympathy can also be an incongruent emotional state as it means feeling *for* another and not feeling *as* another [57,58]. Nevertheless, sympathy is widely accepted as an empathy-related concept as it manifests cognitively when noticing a person

suffering, or affectively when someone's suffering triggers emotional reactions in the observer [15,52,53]. In designing artificial empathy, *sympathetic* responses are typically implemented to chatbots by phrases like "I am sorry to hear that" [39,40].

Although cognitive empathy, affective empathy, and sympathy are somehow distinctive concepts, they all consider empathy a mental process that requires experiential capabilities or complex mindfulness. In addition, there is a behavioral dimension of empathy that is covered less frequently in literature [53]. To worry about someone or to perceive suffering in another might trigger empathetic helping, i.e., helping someone to overcome a distressing event [15]. Theorists consider the emergence of empathetic helping to be either truly altruistic (i.e., helping as an expression of genuine concern and moral beliefs) [59,60], or self-interest driven (i.e., helping with return on benefit expectations or to cope with own negative emotion) [15,61]. Regardless of its origin, empathetic helping usually manifests in efforts of providing support to a person in need of help and can thus be considered a pro-social act. While *empathetic* and *sympathetic* responses mainly provide emotional support, empathetic helping mainly provides instrumental support. Both receiving emotional and instrumental support can be essential in coping with stressful situations [62]. Instrumental support is particularly important in healthcare provision as patients expect an empathetic physician to take care for their issues. For instance, empathetic physicians are expected to listen actively, to be interested in the patients' recovering, and to find solutions for health issues [22,63]. A physician can thus express empathy by indicating being interested in the patient and his or her well-being. Congruently, for modeling artificial empathetic helping, researchers use supporting or caring expressions like "Do you need help?" [64] or "Can you tell me more about how you feel?" [65,66].

2.3 Schemas and Mind Perception Theory

Hitherto, this paper predominantly emphasized the positive consequences of implementing empathy and warmth to interactions with chatbots. However, there is an

increasing number of articles examining differences in the evaluation of bots vs. humans and backfiring effects that might emanate from human-likeness. One of the most well-known theoretical approaches on explaining negative consequences is the "Uncanny Valley Hypothesis" positing that too much human-likeness in inanimate agents can elicit feelings of eeriness or cause a perceived threat to human identity [30,31,67,68]. For instance, recent research has found that humans who feel threatened by machines try to cope with the identity threat by showing compensatory consumption behavior [32] or by emphasizing and valuing human-unique attributes like creativity [69].

However, text-based chatbots have relatively minor social cues thus making them feel more computer-like than humanoid robots with a physical embodiment. Hence, humans usually notice that a chatbot is a software system and might thus apply computer-like schemas to the interaction [8,19,70]. Schemas are cognitive frameworks that organize the knowledge we have about the attributes of certain objects [71,72]. Therefore, they shape our expectations on how objects usually look or operate. In case of chatbots, the activation of computer-like schemas might lead users to apply machine heuristics and mechanistic stereotypes resulting in corresponding expectations [73]. For instance, users might expect a chatbot to have lower problem-solving capabilities compared to a human [18,20,74], but to be able to respond immediately [75]. The perhaps most significant disparity between chatbots and humans lies in their incapacity to experience emotions [2,9,19]. According to "Mind Perception Theory", the ability to think (*agency*) and the ability to feel (*experience*) are the two core dimensions of human mind [17]. While people attribute a somewhat moderate level of agency to bots, the ability to feel is one of the key distinctions between humans and machines [16,76]. Chatbots are thus expected to provide a competent and fast service while lacking interpersonal warmth [19]. Applying this theoretical thought to artificial empathy, humans might not believe a chatbot to be able to accurately understand or even feel emotion. *Empathetic* and *sympathetic*

expressions could thus interfere with computer-like schemas and mechanistic stereotypes and feel ungenuine as feeling *with* or feeling *for* another requires experiential capabilities [19]. Correspondingly, Klein et al. [66] stated more than twenty years ago that the idea of implementing emotional expressions to virtual agents is "perhaps the most problematic one […], since an expression of sympathy really is an expression of feeling, and the computer is incapable of truly feeling anything the user might feel" (p. 126).

2.4 Perceived Authenticity

An issue arising from expressing fake emotions is a reduction in the expressor's perceived authenticity. Authenticity defined as a trait (psychology), or in an existentialism sense (philosophy) means that someone acts in congruence with his or her true self [26,77]. Authenticity is therefore related to dimensions like credibility, sincerity, and honesty and thus closely associated with someone's trustworthiness [26,28,29,78]. Individuals who act inauthentic by pretending to be someone they are not, who display fake emotions, or who can be strongly influenced in their opinion by others might therefore be perceived inconsistent or unreliable. Correspondingly, research in the service domain has demonstrated that customers can expose service employees practicing inauthentic surface acting, and that perceiving inauthenticity in a service provider can lead to unfavorable company outcomes, e.g., lower levels of customer satisfaction [78,79,80]. Also, the concept of authenticity has been applied to non-human entities, e.g., there is a variety of literature on "brand authenticity" that is defined "as the extent to which consumers perceive a brand to be faithful toward itself (continuity), being true to its consumers (credibility), motivated by caring and responsibility (integrity), and able to support consumers in being true to themselves (symbolism)" [28, p. 203]. Just like in interpersonal interactions, perceived authenticity is found to be an important predictor for brand evaluation, e.g., a brand's trustworthiness or brand choice [28,81]. The tendency to favor authentic and to reject inauthentic entities might be explained by humans' sensitivity for

identifying fraudulent individuals that helps to separate cheaters from trustworthy cooperation partners [82].

Interestingly, research has barely addressed the role of perceived (in-)authenticity in interactions with emotional or humanized bots. Considering today's chatbots do not have experiential capabilities and people barely believe them to have, empathy in chatbots might feel inauthentic as it interferes with computer-like schemas and mechanistic stereotypes towards bots [16,19]. This could particularly apply when the chatbot expresses empathy by *empathetic* (feeling *with*) or *sympathetic* (feeling *for*) responses as both provide emotional support and require experiential capabilities. *Empathetic* and *sympathetic* responses might hence reduce the chatbot's perceived authenticity by appearing scripted and fake [19]. This reduction in perceived authenticity might reduce users' willingness to trust the chatbot as it appears to be somehow ungenuine and insincere. In contrast, since empathetic helping rather manifests in providing instrumental than emotional support, *behavioral-empathetic* expressions might interfere less with computer-like schemas and mechanistic stereotypes towards chatbots. *Behavioral-empathetic* expressions could therefore be a more computer-like thus authentic way of implementing empathy to healthcare chatbots.

$H4_a$: A healthcare chatbot responding (1) empathetic, or (2) sympathetic is perceived less authentic than a healthcare chatbot responding non-empathetic.

$H4_b$: There is no significant difference in perceived authenticity between a healthcare chatbot responding behavioral-empathetic and a healthcare chatbot responding non-empathetic.

H5: Perceived authenticity in a healthcare chatbot is positively related to the willingness to trust the chatbot.

2.5 Boundary Conditions and Alternative Explanations

The detrimental effect hypothesized in $H4_a$ is attributed to the incongruence between computer-like schemas towards chatbots and the need for experiential capabilities or complex mindfulness that are required for *empathetic* or *sympathetic* responses. However, many chatbots combine a variety of visual and verbal social cues to elicit a human-like first impression, e.g., when a chatbot is given a human avatar and a name [8,20,43]. In this case, the chatbot has a higher chance to pre-activate human-like schemas leading to the expectation that it feels, thinks, acts, and communicates like a human [20,43]. Congruently, recent research has demonstrated that products humanized by visual design elements such as faces are ascribed with the capacity for experiences like pain or joy [83,84]. The hypothesized backfiring effect could hence be attenuated when an *empathetic* or *sympathetic* responding chatbot is personified (vs. non-personified).

H6: The hypothesized loss in perceived authenticity is attenuated (vs. stays robust) when the chatbot is personified.

Additionally, an alternative explanation for the hypothesized backfiring effect might be that *empathetic* or *sympathetic* responses could generally seem like a phrase, irrespective if expressed by a chatbot or a human. Empathy is considered a socially desirable response leading people to show fake empathy even if they do not really emphasize or sympathize with someone suffering. The apprehension of fake empathy might particularly be present when people anticipate an agent having to respond empathetic due to service environment requirements, e.g., a doctor in assessing patients [85]. If so, the negative effect of *empathetic* and *sympathetic* responses on *perceived authenticity* should also occur when expressed by a human agent. In contrast, if the negative effect truly emanates from the interference with computer-like schemas towards chatbots, the effect should not replicate in human-human interactions.

H7: The hypothesized loss in perceived authenticity does not occur when the healthcare agent is believed to be human.

Figure 1 summarizes all hypotheses in a holistic research model.

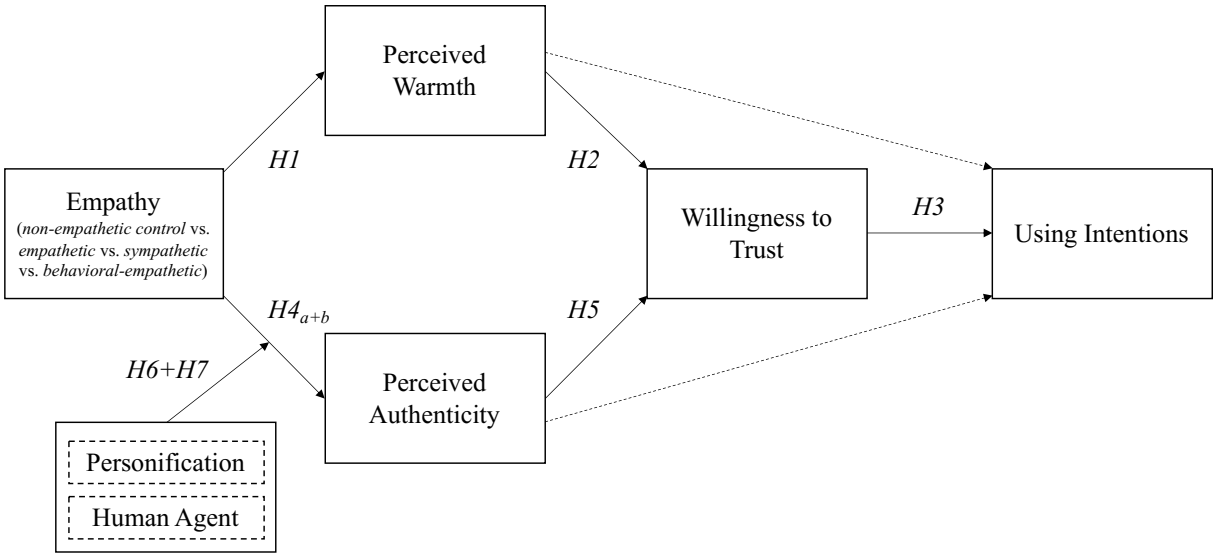


Figure 1. Research model.

3 Study 1

3.1 Method

3.1.1 Scenario and Chatbots

All studies conducted and reported in this paper were independent parts of a larger research project on the design, perception, and evaluation of healthcare chatbots and sought to empirically test the research model presented in Figure 1. Study 1 applied an experimental design in which participants had to take the perspective of a person suffering from a chest pain that radiates to the arms and intensives with breathing and body movement. These symptoms were chosen since they are often associated with worrying diseases like cardiac issues although they are often caused by harmless muscular tensions. The intention was to create a certain level of uncertainty and discomfort so that trust in the chatbot and empathy is relevant at all. Four different healthcare chatbots (*non-empathetic control condition* vs. *empathetic* vs. *sympathetic*

vs. *behavioral-empathetic*) were designed and programmed using the tool "SnatchBot" [86]. All chatbots followed a pre-scripted dialogue asking the participants for some personal information (e.g., age and gender) and their symptoms. Where possible, participants answered questions via buttons to maximize equality of treatment and further preventing text-recognition errors [40]. The conversation flow was identical for all chatbots, except for the empathy manipulations that were conceptualized in accordance with corresponding empathy theories and previous research (see Table 2). After the chatbot has completed questioning, it displayed three clinical pictures associated with the symptoms with varying likelihoods: muscular tension (7 of 10 patients), thoracic spine syndrome (2 of 10), and heart attack (<1 of 10). The chatbot also gave some generic background information and treatment recommendations.

Condition	**Exemplary responses**
Empathetic	I can well understand your concerns.
	I can empathize well with your situation now.
Sympathetic	I am sorry to hear that.
	I feel sorry for you.
Behavioral-empathetic	I will give my best to help you.
	If I can help you in any way, feel free to contact me again any time.
Non-empathetic control condition	No expressions of empathy.

Table 2. Overview of conditions and exemplary responses.

Note: Each chatbot responded three to four times per conversation with corresponding messages.

3.1.2 Pre-Test

A pre-test was conducted to ensure the manipulation's effectiveness. Therefore, screen recordings of all four chatbot conversations were prepared and randomly assigned to *n*=201 participants recruited online. After cleaning the data for invalid respondents (i.e., attention

check failures), the final sample included n=177 participants (53.7% female; M_{age}=35.53, SD_{age}=12.08). Individuals were asked to fill out a standardized questionnaire capturing *perceived empathetic*, *perceived sympathetic*, and *perceived behavioral-empathetic* responses (all self-developed based on empathy theories and measured by multi-item scales). The questionnaire also asked for *perceived overall empathy* to check whether all chatbots equally provide a sense of empathy, except for the *non-empathetic control condition* (single-item measure). Lastly, the questionnaire captured the *scenario's realism* (adapted from Gelbrich et al. [38]) and the *conversation's complexity* (single-item measure) to ensure imaginability and understandability and to rule out confounding effects. All measures used seven-point scales and can be found in Appendix A.

Starting with *perceived overall empathy*, a one-way ANOVA revealed that there is a significant difference across groups (F=4.322, p<.01). The score for the *non-empathetic control condition* (CC; M=3.67) was lower than for the *empathetic* (EM; M=4.57, p<.02), *sympathetic* (SY; M=4.45, p<.04), and *behavioral-empathetic* (BE; M=4.95, p<.001) conditions. Moreover, the empathy conditions did not differ significantly (all ps>.17). Next, three one-way ANOVAs on *perceived empathetic* (α=.93), *sympathetic* (α=.89), and *behavioral-empathetic* (α=.91) responses were run. As intended, the *perceived empathetic* response score was significantly higher for the EM chatbot (M=5.22) than for all other chatbots (M_{CC}=3.07; M_{SY}=4.06; M_{BE}=3.71, all ps<.01), F=13.003, p<.001. Same applied for the *perceived sympathetic* response score that was significantly higher for the SY chatbot (M=5.84) compared to the other conditions (M_{CC}=2.93; M_{EM}=4.23; M_{BE}=4.16, all ps<.001), F=27.539, p<.001. Lastly, the *perceived behavioral-empathetic* response score was significantly higher for the BE chatbot (M=5.48) than for the other chatbots (M_{CC}=3.91; M_{EM}=4.49; M_{SY}=4.86, all ps<.05), F=9.054, p<.001. Moving towards the *scenario's realism* and the *conversation's complexity*, the evaluation of *scenario's realism* was acceptable for a video-based vignette pre-test and did not

differ across groups (M=4.99-5.03; F=.004, p>.99). Same applied for the *conversation's complexity*, that was on a low and comparable level across conditions (M=1.79-2.53; F=2.475, p>.06). In conclusion, the pre-test results verified a successful manipulation and equally realistic and quite easy-to-follow conversations.

3.1.3 Sample and Main Study Procedure

Participants for the main study were recruited on survey platforms and the university's internal recruiting system. The required sample size was calculated a priori using G*Power 3.1. The parameters were set at f=.18 (effect size; small to medium effect according to Cohen [87]), *power level*=.80, and *α error probability*=.05. With four groups, the required minimum sample size was 344. A total of n=366 individuals participated in the study, however, n=11 were excluded due to attention check or technical failures. Hence, the final sample included n=355 individuals (64.5% female; M_{age}=26.05, SD_{age}=7.60).

First, participants read the scenario that described the symptoms in detail. Afterwards, they were redirected to a fictitious healthcare website programmed for the purpose of this study that showed some generic health information and the embedded chatbot (see Appendix B). Participants were randomly assigned to one of the four different chatbots. After having finished the conversation and the assessment, participants returned to the survey and filled out a standardized questionnaire.

3.1.4 Measurements and Control Variables

Most concepts were measured using existing scales. Perceived warmth was measured by three items adapted from Gelbrich et al. [38] and Aaker et al. [88], willingness to trust by six items adapted from Söllner et al. [89] and McKnight et al. [90] and using intentions by three items adapted from Venkatesh et al. [91]. As perceived authenticity has not been examined yet in comparable studies, the scale was based on conceptualizations of authentic personality [26]. The scale included five items capturing "self-alienation" (the extent to which the chatbot is

believed to fake its identity), and "external influences" (the extent to which the chatbot is believed to fake its behavior to please users).

Since the willingness to trust or use a healthcare chatbot does not only depend on chatbot-related, but also on user-related and contextual factors [33], two control variables were included: the participant's general attitudes towards using healthcare chatbots (adapted from Moon & Kim [92]), and the clinical picture's perceived physical risk (self-developed). First, an individual's general attitudes are likely to be related to the chatbot's overall evaluation, i.e., participants holding positive attitudes might be more willing to trust or use a chatbot [33,50]. Second, the perceived physical risk might be a contextual factor determining the willingness to trust the chatbot. If risk perception is high, trusting intentions usually decrease [29]. A full list of items can be found in Appendix C.

3.2 Results

Hypotheses were tested by a serial-mediation-based custom model set up in the PROCESS macro for SPSS [93]. The model included *empathy* as independent variable (multicategorial; 0=CC, 1=BE, 2=EM, 3=SY), *perceived warmth* (M1, α=.77) and *perceived authenticity* (M2, α=.84) as first stage parallel mediators, the *willingness to trust* (M3, α=.90) as second stage mediator, and *using intentions* (α=.93) as dependent variable. The model also controlled for possible direct effects of *perceived warmth* and *perceived authenticity* on *using intentions* (see Figure 1).

The initial calculation estimated parameters on 10,000 bootstrap samples without including the control variables (see Table 3). Confirming *H1*, all chatbots imbued with a sense of empathy enhanced *perceived warmth* compared to the CC (M=4.22) (M_{BE}=4.78, b_{BE}=.56, p<.01; M_{EM}=4.91, b_{EM}=.68, p<.001; M_{SY}=4.94, b_{SY}=.71, p<.001). *Perceived warmth*, subsequently, enhanced the *willingness to trust* the chatbot (b=.36, p<.001) hence lending credence for *H2*. Ultimately, supporting *H3*, the *willingness to trust* strongly predicted *using*

intentions (b=.87, p<.001). To summarize, there was an indirect positive effect for all chatbots imbued with a sense of empathy on *using intentions* serially mediated by *perceived warmth* and *willingness to trust* (b_{BE}=.17 [*CI*=.06;.30]; b_{EM}=.21, [*CI*=.10;.35]; b_{SY}=.22, [*CI*=.11;.36]). Moving towards the second mediator, *perceived authenticity* was lower for both the EM chatbot (M=4.42, b=-1.02, p<.001), and the SY chatbot (M=4.55, b=-.90, p<.001) compared to the CC (M=5.45) thus supporting *H4a*. Contradicting *H4b*, same applied for the BE chatbot, although the effect was weaker (M=4.90, b=-.54, p<.01). However, a one-factor ANOVA applying contrast analysis revealed that *perceived authenticity* was significantly higher for the BE chatbot compared to the EM chatbot (b=.48, p=.014) and tendentially higher compared to the SY chatbot (b=.35, p=.074). Results further confirmed *H5* hypothesizing that *perceived authenticity* is positively related to the *willingness to trust* (b=.28, p<.001). In summary, there was a negative downstream effect on *using intentions* serially mediated by a loss in *perceived authenticity* and *willingness to trust* for the EM chatbot and the SY chatbot (b_{EM}=-.25, [*CI*=-.38;-.15]; b_{SY}=-.22, [*CI*=-.35;-.12]). Note that this – albeit smaller – effect was also observed unexpectedly for the BE chatbot (b=-.13, [*CI*=-.25; -.04]). These opposing indirect effects resulted in an insignificant total effect of *empathy* on *trust* (F=.202, p>.89) and *using intentions* (F=.478, p>.69).

Before adding the control variables, a one-way ANOVA was calculated to examine if *general attitudes* towards using healthcare chatbots and *perceived physical risk* vary across conditions. Results revealed there were no significant differences (all Fs<2.4; all ps>.08). However, simple bivariate correlation analyses revealed significant correlations between both control variables and the *willingness to trust* ($r_{attitudes}$=.59, p<.001; r_{risk}=-.24, p<.001). Both controls were therefore added to the model. The second model calculation showed robustness of the effects (see parameters in parentheses in Table 3).

	Perceived warmth		Perceived authenticity		Willingness to trust		Using intentions	
Predictor	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*
Empathetic	**0.68 (0.61)**	0.16 (0.15)	**-1.02 (-1.09)**	0.18 (0.19)	-	-	-	-
Sympathetic	**0.71 (0.70)**	0.16 (0.15)	**-0.90 (-0.90)**	0.18 (0.19)	-	-	-	-
Behavioral-empathetic	**0.56 (0.59)**	0.17 (0.16)	**-0.54 (-0.50)**	0.18 (0.20)	-	-	-	-
Perceived warmth	-	-	-	-	**0.36 (0.22)**	0.05 (0.05)	**0.16** (0.08)	0.05 (0.05)
Perceived authenticity	-	-	-	-	**0.28 (0.17)**	0.04 (0.04)	**0.09** (0.05)	0.04 (0.04)
Trust	-	-	-	-	-	-	**0.87 (0.65)**	0.05 (0.05)
Controls								
General attitudes	**0.30**	0.04	**0.35**	0.05	**0.39**	0.04	**0.47**	0.04
Perceived physical risk	-0.01	0.04	-0.03	0.04	**-0.07**	0.03	**0.09**	0.03
Without controls	R^2=0.07		R^2=0.09		R^2=0.26		R^2=0.58	
	$F(3, 351)=8.20$, $p<.001$		$F(3, 351)=11.49$, $p<.001$		$F(2, 352)=61.06$, $p<.001$		$F(3, 351)=163.28$, $p<.001$	
With controls	R^2=0.19		R^2=0.22		R^2=0.43		R^2=0.68	
	$F(5, 349)=16.60$, $p<.001$		$F(5, 349)=19.70$, $p<.001$		$F(4, 350)=64.83$, $p<.001$		$F(5, 349)=151.05$, $p<.001$	

Table 3. Results from Study 1 (custom mediation analysis).

Notes: Significant effects (p<.05) are highlighted by ***bold characters****. Parameters inside parentheses show effect sizes when including control variables.*

3.3 Discussion

The first study found robust evidence for most of the hypotheses. Implementing a sense of empathy to healthcare chatbots enhances *perceived warmth* resulting in a higher *willingness to trust* and *using intentions*. In this regard, it confirms previous research positing that perceiving a sense of warmth in interactions with bots facilitates trust-building and using intentions [7,12,38]. However, there was a suppressing negative effect by a loss in *perceived authenticity* – i.e., *empathetic* and *sympathetic* responses appear to be ungenuine. This finding resonates with "Mind Perception Theory" arguing that humans do not attribute experiential abilities or complex mindfulness to inanimate bots [17,76]. Just like in human-human interactions, this somehow insincere behavior reduces the *willingness to trust*. The loss in

perceived authenticity was, however, also observed for the *behavioral-empathetic* chatbot that did not self-disclose affective or complex cognitive states but indicated its intent to help. A potential explanation might be that intentions are associated with *agency* that is the second dimension of mindfulness [17,23]. However, in contrast to *experience*, *agency* is moderately associated with bots which might explain why the *behavioral-empathetic* chatbot was perceived more authentic than the *empathetic* and the *sympathetic* chatbot. Another related explanation for the lower *perceived authenticity* in the *behavioral-empathetic* chatbot might be that any kind of human touch interferes with the prevailing mechanistic stereotypes towards chatbots [19].

4 Study 2

4.1 Purpose

The chatbots used in Study 1 did not show any social cues except for empathy to avoid confounding effects. This lack of human-likeness may have strengthened the perceived incongruence between expected mechanistic responses and the actual level of communicated empathy. Study 2 therefore aimed at testing *H6*, i.e., if the loss in *perceived authenticity* is attenuated (vs. stays robust) when the chatbot is personified.

4.2 Method

4.2.1 Stimuli and Pre-Test

The scenario and the chatbots were similar to those in Study 1, except for the personification that has been implemented by giving the chatbot a name ("Jan") and a profile picture showing a male human physician. A pre-test was conducted with n=32 participants (65.6% female; M_{age}=29.63, SD_{age}=13.47) to ensure the manipulation's effectiveness. The pre-test used a one-factor experimental design with two conditions (personified vs. non-personified chatbot). Participants were randomly assigned to one of the conditions and saw a screenshot of

the website with the embedded chatbot either personified or not (see Appendix D). Afterwards, participants filled out a questionnaire asking if the participants perceived the chatbot like a person by three items adapted from Crolic et al. [20] (α=.92; see Appendix E). Results provided evidence for a successful manipulation as participants perceived the personified chatbot more like a person than the non-personified one ($M_{non\text{-}person}$=2.18; M_{person}=5.33, p<.001).

4.2.2 Sample and Main Study Procedure

Like in Study 1, participants were recruited by means of convenience sampling. A total of n=373 individuals participated, n=28 of which were excluded due to attention check or technical failures. Hence, the final sample included n=345 individuals (66.7% female; M_{age}=26.24, SD_{age}=6.42). For details on materials, procedure, and questionnaire, see Study 1 and Appendix B and C.

4.3 Results

4.3.1 Model Replication

To ensure comparability and further validate the robustness of Study 1's findings, the data analysis was replicated. First, results of the initial model calculation provided mixed evidence for *H6* as the – marginally mitigated – negative effect on *perceived authenticity* (α=.85) remained significant for the EM chatbot (M=4.19, b=-.78, p<.001) and the SY chatbot (M=4.22, b=-.75, p<.001) when comparing to the CC (M=4.97). However, for the BE chatbot, the unexpected negative effect observed in Study 1 disappeared (M=4.82, b=-.15, p=.46) thus partially supporting *H6*. Moreover, the positive effects on *perceived warmth* (α=.81) were replicated for all chatbots providing a sense of empathy (M_{BE}=5.03, b_{BE}=.46, p<.01; M_{EM}=5.06, b_{EM}=.50, p<.01; M_{SY}=4.96, b_{SY}=.39, p<.03) as they were perceived warmer than the CC (M=4.57). Again, both *perceived warmth* (b=.45, p<.001), and *perceived authenticity* (b=.25, p<.001) were positively related to the *willingness to trust* (α=.91) resulting in higher *using*

intentions (α=.93; *b*=.88, *p*<.001). Summarizing, Study 2 replicated the significant positive indirect effect on *using intentions* serially mediated by *perceived warmth* and *willingness to trust* for all empathy-imbued chatbots (b_{BE}=.18 [*CI*=.04;.35]; b_{EM}=.20, [*CI*=.06;.36]; b_{SY}=.16, [*CI*=.03;.30]). However, the negative indirect effect through the loss in *perceived authenticity* and *willingness to trust* only replicated for the EM chatbot (*b*=-.17, [*CI*=-.29;-.07]) and the SY chatbot (*b*=-.16, [*CI*=-.28;-.07]), but not the BE chatbot (*b*=-.03, [*CI*=-.12;.05]). Results stayed robust when adding both control variables (i.e., *general attitudes* and *perceived physical risk*) to the model (see parameters in parentheses in Table 4). Like in Study 1, the total effect of *empathy* on *trust* (*F*=.559, *p*>.64), and *using intentions* (*F*=1.979, *p*>.11) was insignificant.

	Perceived warmth		Perceived authenticity		Willingness to trust		Using intentions	
Predictor	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*
Empathetic	**0.50 (0.41)**	0.17 (0.15)	**-0.78 (-0.88)**	0.20 (0.18)	-	-	-	-
Sympathetic	**0.39 (0.50)**	0.17 (0.15)	**-0.75 (-0.63)**	0.20 (0.18)	-	-	-	-
Behavioral-empathetic	**0.46 (0.47)**	0.17 (0.15)	-0.15 (-0.14)	0.21 (0.19)	-	-	-	-
Perceived warmth	-	-	-	-	**0.45 (0.24)**	0.06 (0.05)	**0.15** (0.08)	0.06 (0.05)
Perceived authenticity	-	-	-	-	**0.25 (0.12)**	0.05 (0.04)	0.06 (0.01)	0.05 (0.04)
Trust	-	-	-	-	-	-	**0.88 (0.67)**	0.05 (0.05)
Controls								
General attitudes	**0.33**	0.04	**0.36**	0.04	**0.41**	0.04	**0.33**	0.04
Perceived physical risk	-0.06	0.03	-0.06	0.04	**-0.09**	0.03	-0.05	0.03
Without controls	R^2=0.03		R^2=0.07		R^2=0.30		R^2=0.62	
	$F(3, 341)=3.76$, $p=.01$		$F(3, 341)=7.95$, $p<.001$		$F(2, 342)=73.03$, $p<.001$		$F(3, 341)=185.75$, $p<.001$	
With controls	R^2=0.25		R^2=0.24		R^2=0.49		R^2=0.68	
	$F(5, 339)=22.89$, $p<.001$		$F(5, 339)=21.39$, $p<.001$		$F(4, 340)=80.38$, $p<.001$		$F(5, 339)=141.98$, $p<.001$	

Table 4. Results from Study 2 (custom mediation analysis).

Notes: Significant effects (p<.05) are highlighted by ***bold characters****. Parameters inside parentheses show effect sizes when including control variables.*

4.3.2 Study Comparison

Next, data from both studies were merged to account for 1) potential main effects of the chatbots' *personification*, and 2) interaction effects between *personification* and *empathy*. First, two two-way ANOVAs (*personification***empathy*) were calculated with 1) *perceived warmth*, and 2) *perceived authenticity* as dependent variables. Regarding *perceived warmth*, there were significant main effects for *personification* ($F(1, 692)=5.364$, $p<.03$), and *empathy* ($F(3, 692)=11.219$, $p<.001$), but no interaction effect, $F(3, 692)=.684$, $p=.56$. Continuing with *perceived authenticity*, there were significant main effects for *personification* ($F(1, 692)=7.851$, $p<.01$), and *empathy* ($F(3, 692)=18.603$, $p<.001$). Again, there was no interaction effect, $F(3, 692)=.673$, $p=.57$.

Diving deeper into the significant main effects of *personification*, the chatbots in Study 2 were perceived warmer ($M=4.57$-5.06) than their non-personified equivalents in Study 1 ($M=4.22$-4.94), with an overall significant difference ($M_{\text{Study 2}}=4.90$; $M_{\text{Study 1}}=4.72$, $p<.04$). Inversely, *perceived authenticity* was lower for the chatbots in Study 2 ($M=4.19$-4.97) compared to their non-personified equivalents in Study 1 ($M=4.42$-5.45), with an overall significant difference ($M_{\text{Study 2}}=4.54$; $M_{\text{Study 1}}=4.82$, $p<.01$).

4.4 Discussion

Study 2 examined if the negative effect of empathy in healthcare chatbots on *perceived authenticity* can be attenuated when the chatbot has an overall more human-like appearance. The idea behind was that personifying the chatbot might elicit anthropomorphic thinking thus reducing the perceived incongruence of the chatbot's empathizing or sympathizing responses. Although this attenuating effect was observed for the *behavioral-empathetic* chatbot, the negative effect stayed robust for the *empathetic* and the *sympathetic* chatbot. The robustness of this negative effect confirms "Mind Perception Theory" positing that experiential capabilities and complex mindfulness are considered one of the key factors distinguishing humans from

machines [17,76]. The mitigation of this negative effect for the *behavioral-empathetic* chatbot supports this line of argumentation as behavioral empathy interferes less with computer-like schemas towards chatbots, particularly when the chatbot has a human-like appearance. This finding supports *H3b* that could not be confirmed in Study 1 in which the chatbots had a computer-like appearance.

Another interesting finding was the negative main effect of *personification* on *perceived authenticity*, i.e., personified chatbots were perceived less authentic than non-personified ones. This finding further supports the hypothesis that human-unique attributes (i.e., having a personality) might reduce a chatbot's *perceived authenticity*. Similar to experiential capabilities, having a human appearance and a personality might interfere with computer-like schemas towards chatbots. Also, this finding could potentially explain the omitted negative effect on *perceived authenticity* for the *behavioral-empathetic* chatbot vs. the *non-empathetic control condition* in two ways. First, the *personification* also reduced the *perceived authenticity* for the *non-empathetic control condition* and thus moved it towards the less authentic empathy expressing chatbots. Second, the used elements for the personification (i.e., a human picture and a name) might have been more salient and human-unique than *behavioral-empathetic* expressions thus overshadowing the effect.

5 Study 3

5.1 Purpose

Even though Study 1 and 2 found evidence that empathy in healthcare chatbots can reduce their *perceived authenticity*, this paper still falls short in proofing that this finding is exclusive to interactions with chatbots and can thus be attributed to the interference with computer-like schemas. Study 3 therefore sought to test *H7*, i.e., if the backfiring does not occur in human-human interactions.

5.2 Method

5.2.1 Stimuli

The main difference between Study 3 and 2 was that participants watched a pre-recorded video of an interaction between a human agent (i.e., a "physician") and a "patient". Video stimuli were used since participants were expected to be able to distinguish an interaction with a chatbot from an interaction with a human. Moreover, using hypothetical scenarios instead of real interactions (e.g., screenshots) is still a common and accepted procedure in the present research area [75]. The four conversations used for Study 3 were almost identical to those in Study 2 and were prepared by two individuals in iMessage (see Appendix F). Only two minor things have changed: first, the "physician" used response time delays since immediate responses are typical for chatbots while being implausible for human agents [75]. Second, the "physician" only presented the main diagnosis since (1) alternative explanations indicated with likelihoods are rather mechanistic, and (2) the clinical pictures' descriptions have been quite extensive, i.e., they might have diverted participant's attention, particularly considering that response time delays would have been unreasonably long.

5.2.2 Sample and Study Procedure

A total of n=454 individuals participated in the study. Besides passing attention checks, participants had to correctly answer if the video showed an interaction with a physician (correct, n=393), or a chatbot (false, n=61) to be included in the analysis. The final sample consisted of n=361 participants (57.6% female; M_{age}=29.49, SD_{age}=10.38).

After a short introduction, participants were randomly assigned to one of the four conditions and watched the video of the interaction that lasted approx. seven minutes. The following questionnaire was similar to the ones used in Study 1 and 2 with minor contextual adoptions (see Appendix C).

5.3 Results

Study 1's and 2's data analysis procedure was replicated. Supporting *H7*, there was no significant difference in *perceived authenticity* (α=.78) for none of the empathy conditions compared to the CC (*M*=5.08) (M_{BE}=5.02, b_{BE}=-.06, *p*=.74; M_{EM}=5.20, b_{EM}=.12, *p*=.53; M_{SY}=4.96, b_{SY}=-.11, *p*=.54). However, the positive effect of empathy on *perceived warmth* (α=.88) remained significant (M_{BE}=5.54, b_{BE}=.37, *p*<.03; M_{EM}=5.59, b_{EM}=.42, *p*<.02; M_{SY}=5.71, b_{SY}=.54, *p*<.001) as all human agents who expressed any kind of empathy were perceived warmer than the agent who did not (*M*=5.17). Both *perceived warmth* (*b*=.60, *p*<.001) and *perceived authenticity* (*b*=.34, *p*<.001) were positively related to the *willingness to trust* the human agent (α=.95) ultimately facilitating *using intentions* (α=.94; *b*=.90, *p*<.001). Hence, there was no empathy-induced negative downstream effect on *using intentions* through a loss in *perceived authenticity* while the positive indirect effect through *perceived warmth* and the *willingness to trust* remained significant (b_{BE}=.20 [*CI*=.02;.39]; b_{EM}=.23, [*CI*=.04;.42]; b_{SY}=.29, [*CI*=.12;.48]). Results stayed robust when adding *general attitudes* as control variable (*perceived physical risk* was neither associated with the *willingness to trust* nor *using intentions* in Study 3; see parameters in parentheses in Table 5). However, despite the presence of the positive indirect effect through *perceived warmth* and the absence of the negative indirect effect through *perceived authenticity*, the total effect on *trust* (*F*=.199, *p*>.89), and *using intentions* (*F*=.957, *p*>.41) was insignificant.

	Perceived warmth		Perceived authenticity		Willingness to trust		Using intentions	
Predictor	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*
Empathetic	**0.42 (0.46)**	0.17 (0.14)	0.12 (0.16)	0.20 (0.17)	-	-	-	-
Sympathetic	**0.54 (0.56)**	0.16 (0.13)	-0.11 (-0.10)	0.18 (0.16)	-	-	-	-
Behavioral-empathetic	**0.37 (0.38)**	0.16 (0.14)	-0.06 (-0.05)	0.19 (0.17)	-	-	-	-
Perceived warmth	-	-	-	-	**0.60 (0.34)**	0.05 (0.05)	0.11 (0.05)	0.05 (0.05)
Perceived authenticity	-	-	-	-	**0.34 (0.17)**	0.05 (0.04)	**0.10** (0.06)	0.04 (0.04)

Trust	-	-	-	-	-	-	**0.90 (0.69)**	0.05 (0.05)
Controls								
General attitudes	**0.39**	0.03	**0.40**	0.04	**0.48**	0.04	**0.34**	0.05
Without controls	R^2=0.03		R^2<0.01		R^2=0.44		R^2=0.70	
	$F(3, 357)=4.16$, $p<0.01$		$F(3, 357)=0.57$, $p=0.63$		$F(2, 358)=139.81$, $p<.001$		$F(3, 357)=282.19$, $p<.001$	
With controls	R^2=0.31		R^2=0.22		R^2=0.61		R^2=0.74	
	$F(4, 356)=39.47$, $p<.001$		$F(4, 356)=24.49$, $p<.001$		$F(3, 357)=186.23$, $p<.001$		$F(4, 356)=258.81$, $p<.001$	

Table 5. Results from Study 3 (custom mediation analysis).

Notes: Significant effects (p<.05) are highlighted by ***bold characters****. Parameters inside parentheses show effect sizes when including control variables.*

5.4 Discussion

The intent of Study 3 was to examine if the negative effect of empathy on *perceived authenticity* disappears when the agent is believed to be human and can thus truly be attributed to computer-like schemas and mechanistic stereotypes towards chatbots [19]. Results substantiated this hypothesis as none of the human agents expressing empathy was perceived less authentic compared to the agent expressing no empathy. Since humans attribute experiential capabilities and complex mindfulness to other humans [17,54], *empathetic* and *sympathetic* expressions seem more genuine from a human agent vs. a chatbot. Study 3 hence confirmed that artificial empathy is perceived different from interpersonal empathy thus showing that the concept of empathy is not equally applicable to chatbots.

6 General Discussion

The present paper provides evidence that expressions of empathy in healthcare chatbots do not only enhance perceived warmth but can also reduce perceived authenticity resulting in detrimental effects on the willingness to trust and using intentions. This backfiring effect is particularly robust for chatbots responding in an *empathetic* (feeling *with*) or *sympathetic*

(feeling *for*) manner as both require experiential capabilities that chatbots do not have. This research hence contributes to the current debate on chances and risks of human-likeness in bots and enables several theoretical and practical implications as well as future research avenues.

6.1 Theoretical Contributions

This paper makes two major theoretical contributions: first, it demonstrates that the interpersonal concept of empathy is not generally applicable to interactions with chatbots. In this regard, it shows that the multidimensionality of empathy should be considered in conceptualizing and studying artificial empathy. And second, it introduces the concept of *perceived authenticity* to the literature on human-bot interaction. In the following, these contributions are elucidated in more detail.

While interpersonal empathy has been extensively conceptualized and well-researched, a nuanced perspective on artificial empathy is still missing. A major issue is the insufficient consideration of the concept's multidimensionality, leading to an incomplete understanding of whether humans react in the same way to artificial empathy as they do to interpersonal empathy. In this regard, it remained obscure if all kinds of empathy are equally appropriate to design artificial empathy. Starting with similarities between interpersonal and artificial empathy, the present findings resonate with "Social Response Theory" [24] and the "Stereotype Content Model" [25] as empathy in a healthcare chatbot creates a sense of warmth resulting in favorable consequences. Precisely, this research aligns with previous studies showing that feeling a sense of warmth and empathy in artificial agents can enhance trust [36], using intentions [37,39], and behavioral persistence [38,66]. These positive effects occurred independently of (1) the kind of empathy, and (2) the presence vs. non-presence of other social cues. Providing emotional (e.g., *empathetic*, or *sympathetic*) or instrumental (e.g., *behavioral-empathetic*) support equally created a sense of warmth in the agent.

However, the major novelty presented in this paper is that a chatbot's expression of empathy might seem scripted and inauthentic therefore contradicting the positive findings observed in previous research. This backfiring effect was particularly robust for *empathetic* (feeling *with*) and *sympathetic* (feeling *for*) responding chatbots. This finding resonates with "Mind Perception Theory" arguing that experiential capabilities and complex mindfulness are considered uniquely human while being poorly associated with bots [17,76]. *Empathetic* or *sympathetic* responses hence interfere with computer-like schemas and mechanistic stereotypes towards chatbots resulting in lower perceived authenticity, even when the chatbot is personified (see Study 2). For *behavioral-empathetic* responses, results were less clear. Study 1 found an unexpected small detrimental effect of behavioral-empathetic responses on perceived authenticity while there was no such effect in Study 2 using personified chatbots. As hypothesized and discussed earlier, *behavioral-empathetic* responses might interfere less with computer-like schemas towards chatbots, particularly when being in congruence with other social cues. It might hence be more appropriate to model artificial empathy by means of providing instrumental rather than emotional support. For instance, chatbots and other virtual assistants could emphasize their purpose to support and help the user instead of expressing empathetic or sympathetic feelings to create a more authentic sense of artificial empathy.

Furthermore, this research is among the first to study the role of *perceived authenticity* in interactions with bots. Although (perceived) authenticity has been studied in interactions with service employees [78] or brands [28], there is barely research on the significance, determinants, and outcomes of perceived authenticity in (chat-)bots. This research demonstrates that social cues that are distinctively human and poorly associated with bots might feel ungenuine and inauthentic. Like in interactions with human service employees, perceiving inauthenticity can reduce trust towards the agent since authenticity is closely associated with dimensions like credibility, sincerity, and honesty [28]. With introducing perceived authenticity

to the literature on human-bot interaction, the present research broadens the understanding of potential negative consequences emanating from humanizing bots and pioneers quantitative research on inauthenticity perception. Previous research on backfiring effects has frequently focused on "Uncanny Valley Theory" [30,31,67,68] or unrealistic high expectations humanized chatbots might elicit in consumers [20]. The present research demonstrates that negative effects or null findings can also be attributed to the fake character that might be inherent to certain social cues, e.g., experiential capabilities or having a personality. The resulting reduction in perceived authenticity was found to act as an opposing mediator for potential positive effects of human-like cues (i.e., empathy) on relevant outcomes dimensions like trust or using intentions. Although humans unconsciously tend to respond positive to human-likeness, this research provides further evidence that humans perceive and evaluate specific social cues differently in interactions with chatbots compared to interactions with other humans [75,94].

6.2 Managerial Implications

Practitioners and software designers frequently equip chatbots with social cues to make interactions more natural and to enhance relationship-building with users [7]. However, this research demonstrates that not all social cues might be equally appropriate to implement. With the intention to make chatbots more human-like, software designers and service providers should be careful in their selection of social cues to not diminish the chatbot's perceived authenticity. Social cues that are poorly associated with bots (e.g., emotional responses or elements of personification) might appear fake and ungenuine that could lead to unintended and unfavorable consequences. Practitioners are hence encouraged to consider the different expectations and stereotypes humans have towards chatbots to not design too human-like and inauthentic agents. This could be particularly important for companies that provide services characterized by a high degree of confidentiality and credibility, e.g., financial services. In such

service domains, it could be advisable to avoid using inauthentic social cues that might mitigate trust.

Regarding the implementation of empathy to chatbots, there is further evidence that empathy might facilitate trust and using intentions in environments that require care-taking and interpersonal relationships, e.g., healthcare [10]. However, since expressions of empathy that require experiential capabilities (i.e., feeling *with* or feeling *for* another) can reduce the chatbot's perceived authenticity, practitioners could decide to design artificial empathy by expressions of instrumental support. A chatbot that indicates its intent to help and to take care for a client equally provides a sense of empathy and warmth without self-disclosing inauthentic experiential capabilities. Also, practitioners should ensure a consistent social design, i.e., empathy should be combined with further social cues to create a congruent experience.

6.3 Limitations and Future Research

Like with any empirical research, this paper has some limitations and implications for future research to discuss. Starting with the finding's generalizability, additional studies are needed to examine the research model's applicability to other service contexts as this paper only focuses on healthcare chatbots. First, healthcare provision is characterized by a high need for empathy and interpersonal relations resulting in a high predictive power for perceived warmth compared to perceived authenticity on the willingness to trust. In service environments with a lower need for warmth and a high need for integrity (e.g., financial services), the detrimental effect of perceived inauthenticity could be even more harmful. Second, it is to be studied if the loss in perceived authenticity induced by empathetic and sympathetic responses replicates for other kinds of bots and in different service environments. Referring to anthropomorphism theories [23] and the findings from Study 2, an overall more human-like appearance (e.g., when a robot has a physical embodiment) is likely to elicit the application of human-like schemas and interpersonal heuristics making human-like behavior appear more reasonable. Multimodal

expressions of empathy (e.g., verbal *and* visual) could appear more consistent thus authentic [37]. Also, the service environment a bot is used in can determine what schemas people apply to the interaction. Previous research has demonstrated that human-like service environments are more likely to elicit anthropomorphic thinking and human-like schemas [10]. However, as healthcare provision is considered one of the most human-like tasks, the backfiring effect of *empathetic* and *sympathetic* responses might apply to many other service context as well. Hence, it seems reasonable that the backfiring effect could be even stronger in computer-like service environments (e.g., receiving product recommendations). Future research could pick-up this idea and conduct further studies in different service environments to seek evidence for the present findings' generalizability.

In seeking for cues to model authentic artificial empathy, it could also be promising to broaden the scope beyond explicit verbal or visual expressions of empathy. For instance, researchers could examine the potentials of equipping a chatbot with the capability to accurately recognize the users' emotional states or needs (e.g., by means of sentiment analysis) [95]. A chatbot that can adopt its behavior to the users' situation (e.g., by sending calming information to a concerned patient) might provide a subliminal sense of empathy. Furthermore, empathetic, or sympathetic responses could feel more authentic when the chatbot only sends them after having accurately recognized the emotional state of a user (vs. sending them by default). In a broader sense, future studies could go beyond empathy and consider in more detail which social cues are perceived (in-)authentic since not all social cues might be equally appropriate for humanizing chatbots. Research in the domain of human-bot interaction has just begun to identify backfiring effects of human-likeness and differences in the perception and evaluation of bots vs. humans [18,20,30,32]. Further examining the sweet spot between human-likeness and robot-likeness regarding authentic vs. inauthentic social cues might provide valuable

insights for both theorists and practitioners, particularly in times of rapidly advancing chatbot technologies [70].

Lastly, it is important to contextualize the findings of the present research within the timeframe the studies were conducted in. Given the rapid developments in AI and chatbot technology, it cannot be excluded that future generations of chatbots will be able to accurately simulate or even experience something we call "emotion" or "empathy" [3]. Regardless of whether artificial emotions become reality or remain fiction, humans' schemas of chatbots could change over time. First, given the increasing performance of chatbots in mimicking human behavior, the attribution of uniquely human capabilities could expand to bots. This might particularly hold true for future generations who grow up with chatbot interactions which are barely distinguishable from interhuman interactions. In this scenario, schemas of chatbots might move closer to humans facilitating their perception as social actors. Hence, empathetic, or sympathetic expressions might be considered authentic. Second, as humans become more experienced and knowledgeable about chatbots, schemas could become more accurate, i.e., computer-like [72,96]. Anthropomorphism and social responses towards computers are considered cognitive biases that are more likely to occur when people have little knowledge about an agent [23,24]. If people get an even higher awareness for the technical nature of bots in future, schemas could remain (or become) more computer-like making empathetic, or sympathetic responses still feel inauthentic. Future research on humanizing chatbots and anthropomorphism should account for this potential shift in schemas.

7 Conclusion

Although making healthcare chatbots more empathetic and human-like seems promising, this research demonstrates that not all kinds of empathy are equally appropriate for designing artificial empathy. Results reveal that expressions of empathy that require experiential capabilities or complex mindfulness feel inauthentic as humans do not believe a

chatbot to have such capabilities. This loss in perceived authenticity is found to have detrimental effects on trust and using intentions. Instead, modeling artificial empathy by providing instrumental support feels more authentic as it aligns better with computer-like schemas and mechanistic stereotypes towards chatbots. Researchers and practitioners are hence encouraged to take a more nuanced perspective on positive and negative consequences that might emanate from the implementation of distinctively human attributes to chatbots. Generally assuming that concepts important in interpersonal interactions, such as empathy, are equally applicable to interactions with chatbots may be an oversimplification and therefore require more clarification.

References

[1] T. Araujo, Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions, Comput. Hum. Behav. 85 (2018) 183–189. https://doi.org/10.1016/j.chb.2018.03.051.

[2] J. Wirtz, P.G. Patterson, W.H. Kunz, T. Gruber, V.N. Lu, S. Paluch, A. Martins, Brave new world: Service robots in the frontline, J. Serv. Manag. 29(5) (2018) 907–931. https://doi.org/10.1108/JOSM-04-2018-0119.

[3] M.-H. Huang, R.T. Rust, Artificial intelligence in service, J. Serv. Res. 21(2) (2018) 155–172. https://doi.org/10.1177/1094670517752459.

[4] B. Larivière, D. Bowen, T.W. Andreassen, W. Kunz, N.J. Sirianni, C. Voss, N.V. Wünderlich, A. De Keyser, "Service encounter 2.0": An investigation into the roles of technology, employees and customers, J. Bus. Res. 79 (2017) 238–246. https://doi.org/10.1016/j.jbusres.2017.03.008.

[5] M.-H. Huang, R.T. Rust, Engaged to a robot? The role of AI in service, J. Serv. Res. 24(1) (2021) 30–41. https://doi.org/10.1177/1094670520902266.

[6] E. Konya-Baumbach, M. Biller, S. von Janda, 2023. Someone out there? A study on the social presence of anthropomorphized chatbots. Comput. Hum. Behav. 139, 107513. https://doi.org/10.1016/j.chb.2022.107513.

[7] M. Blut, C. Wang, N.V. Wünderlich, C. Brock, Understanding anthropomorphism in service provision: A meta-analysis of physical robots, chatbots, and other AI, J. Acad. Mark. Sci. 49 (2021) 632–658. https://doi.org/10.1007/s11747-020-00762-y.

[8] E. Go, S.S. Sundar, Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions, Comput. Hum. Behav. 97 (2019) 304–316. https://doi.org/10.1016/j.chb.2019.01.020.

[9] K.T. Do, H. Gip, P. Guchait, C.-Y. Wang, E.S. Baaklini, Empathetic creativity for frontline employees in the age of service robots: Conceptualization and scale development, J. Serv. Manag. 34(3) (2023) 433–466. https://doi.org/10.1108/JOSM-09-2021-0352.

[10] A.-M. Seeger, J. Pfeiffer, A. Heinzl, Texting with humanlike conversational agents: Designing for anthropomorphism, J. Assoc. Inf. Syst. 22(4) (2021) 931–967. https://doi.org/10.17705/1jais.00685.

[11] O.H. Chi, S. Jia, Y. Li, D. Gursoy, 2021. Developing a formative scale to measure consumers' trust toward interaction with artificially intelligent (AI) social robots in service delivery. Comput. Hum. Behav. 118, 106700. https://doi.org/10.1016/j.chb.2021.106700.

[12] C. Pelau, D.-C. Dabija, I. Ene, 2021. What makes an AI device human-like? The role of interaction quality, empathy and perceived psychological anthropomorphic characteristics in the acceptance of artificial intelligence in the service industry. Comput. Hum. Behav. 122, 106855. https://doi.org/10.1016/j.chb.2021.106855.

[13] J.A. Pepito, H. Ito, F. Betriana, T. Tanioka, R.C. Locsin, 2020. Intelligent humanoid robots expressing artificial humanlike empathy in nursing situations. Nurs. Philos. 21(4), e12318. https://doi.org/10.1111/nup.12318.

[14] Ö.N. Yalçın, S. DiPaola, Modeling empathy: Building a link between affective and cognitive processes, Artif. Intell. Rev. 53(4) (2020) 2983–3006. https://doi.org/10.1007/s10462-019-09753-0.

[15] F.B.M. de Waal, Putting the altruism back into altruism: The evolution of empathy, Annu. Rev. Psychol. 59 (2008) 279–300. https://doi.org/10.1146/annurev.psych.59.103006.093625.

[16] K. Gray, D.M. Wegner, Feeling robots and human zombies: Mind perception and the uncanny valley, Cogn. 125(1) (2012) 125–130. https://doi.org/10.1016/j.cognition.2012.06.007.

[17] H.M. Gray, K. Gray, D.M. Wegner, Dimensions of mind perception, Sci. 315(5812) (2007) 619–619. https://doi.org/10.1126/science.1134475.

[18] S. Yu, J. Xiong, H. Shen, 2022. The rise of chatbots: The effect of using chatbot agents on consumers' responses to request rejection. J. Consum. Psychol., jcpy.1330. https://doi.org/10.1002/jcpy.1330.

[19] J. Meng, Y. Dai, Emotional support from AI chatbots: Should a supportive partner self-disclose or not?, J. Comput. Mediat. Commun. 26(4) (2021) 207–222. https://doi.org/10.1093/jcmc/zmab005.

[20] C. Crolic, F. Thomaz, R. Hadi, A.T. Stephen, Blame the bot: Anthropomorphism and anger in customer–chatbot interactions, J. Mark. 86(1) (2022) 132–148. https://doi.org/10.1177/00222429211045687.

[21] S. Kim, R.P. Chen, K. Zhang, Anthropomorphized helpers undermine autonomy and enjoyment in computer games, J. Consum. Res. 43(2) (2016), 282–302. https://doi.org/10.1093/jcr/ucw016.

[22] D. Jeffrey, Empathy, sympathy and compassion in healthcare: Is there a problem? Is there a difference? Does it matter?, J. R. Soc. Med. 109(12) (2016) 446–452. https://doi.org/10.1177/0141076816680120.

[23] N. Epley, A. Waytz, J.T. Cacioppo, On seeing human: A three-factor theory of anthropomorphism, Psychol. Rev. 114(4) (2007) 864–886. https://doi.org/10.1037/0033-295X.114.4.864.

[24] C.I. Nass, Y. Moon, Machines and mindlessness: Social responses to computers, J. Soc. Issues 56(1) (2000) 81–103. https://doi.org/10.1111/0022-4537.00153.

[25] S.T. Fiske, A.J.C. Cuddy, P. Glick, J. Xu, A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition, J. Pers. Soc. Psychol. 82(6) (2002) 878–902. https://doi.org/10.1037/0022-3514.82.6.878.

[26] A.M. Wood, P.A. Linley, J. Maltby, M. Baliousis, S. Joseph, The authentic personality: A theoretical and empirical conceptualization and the development of the authenticity scale, J. Couns. Psychol. 55(3) (2008) 385–399. https://doi.org/10.1037/0022-0167.55.3.385.

[27] N. Epley, A. Waytz, Mind perception, in: S.T. Fiske, D.T. Gilbert, G. Lindzey (Eds.), Handbook of social psychology, John Wiley & Sons, Inc., Hoboken, 2010, pp. 498–541. https://doi.org/10.1002/9780470561119.socpsy001014.

[28] F. Morhart, L. Malär, A. Guèvremont, F. Girardin, B. Grohmann, Brand authenticity: An integrative framework and measurement scale, J. Consum. Psychol. 25(2) (2015) 200–218. https://doi.org/10.1016/j.jcps.2014.11.006.

[29] R.C. Mayer, J.H. Davis, F.D. Schoorman, An integrative model of organizational trust, Acad. Manag. Rev. 20(3) (1995) 709–734. https://doi.org/10.2307/258792.

[30] M. Appel, D. Izydorczyk, S. Weber, M. Mara, T. Lischetzke, The uncanny of mind in a machine: Humanoid robots as tools, agents, and experiencers, Comput. Hum. Behav. 102 (2020) 274–286. https://doi.org/10.1016/j.chb.2019.07.031.

[31] J. Giger, N. Piçarra, P. Alves-Oliveira, R. Oliveira, P. Arriaga, Humanization of robots: Is it really such a good idea?, Hum. Behav. Emerg. Technol. 1(2) (2019) 111–123. https://doi.org/10.1002/hbe2.147.

[32] M. Mende, M.L. Scott, J. van Doorn, D. Grewal, I. Shanks, Service robots rising: How humanoid robots influence service experiences and elicit compensatory consumer responses, J. Mark. Res. 56(4) (2019) 535–556. https://doi.org/10.1177/0022243718822827.

[33] L. Seitz, S. Bekmeier-Feuerhahn, K. Gohil, 2022. Can we trust a chatbot like a physician? A qualitative study on understanding the emergence of trust toward diagnostic chatbots. Int. J. Hum. Comput. Stud. 165, 102848. https://doi.org/10.1016/j.ijhcs.2022.102848.

[34] L. Zhou, J. Gao, D. Li, H.-Y. Shum, The design and implementation of XiaoIce, an empathetic social chatbot, Comput. Linguist. 46(1) (2020) 53–93. https://doi.org/10.1162/coli_a_00368.

[35] B. Liu, S.S. Sundar, Should machines express sympathy and empathy? Experiments with a health advice chatbot, Cyberpsychol. Behav. Soc. Netw. 21(10) (2018) 625–636. https://doi.org/10.1089/cyber.2018.0110.

[36] S. Brave, C. Nass, K. Hutchinson, Computers that care: Investigating the effects of orientation of emotion exhibited by an embodied computer agent, Int. J. Hum. Comput. Stud. 62(2) (2005) 161–178. https://doi.org/10.1016/j.ijhcs.2004.11.002.

[37] X. Lv, Y. Yang, D. Qin, X. Cao, H. Xu, 2022. Artificial intelligence service recovery: The role of empathic response in hospitality customers' continuous usage intention. Comput. Hum. Behav. 126, 106993. https://doi.org/10.1016/j.chb.2021.106993.

[38] K. Gelbrich, J. Hagel, C. Orsingher, Emotional support from a digital assistant in technology-mediated services: Effects on customer satisfaction and behavioral persistence, Int. J. Res. Mark. 38(1) (2021) 176–193. https://doi.org/10.1016/j.ijresmar.2020.06.004.

[39] T.W. Bickmore, R.W. Picard, Establishing and maintaining long-term human-computer relationships, ACM Trans. Comput. Hum. Interact. 12(2) (2005) 293–327. https://doi.org/10.1145/1067860.1067867.

[40] M. de Gennaro, E.G. Krumhuber, G. Lucas, Effectiveness of an empathic chatbot in combating adverse effects of social exclusion on mood, Front. Psychol. 10(3061) (2019) 1–14. https://doi.org/10.3389/fpsyg.2019.03061.

[41] K.K. Fitzpatrick, A. Darcy, M. Vierhile, 2017. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (woebot): A randomized controlled trial. JMIR Ment. Health. 4(2), e19. https://doi.org/10.2196/mental.7785.

[42] T.M. Holtgraves, S.J. Ross, C.R. Weywadt, T.L. Han, Perceiving artificial social agents, Comput. Hum. Behav. 23(5) (2007) 2163–2174. https://doi.org/10.1016/j.chb.2006.02.017.

[43] N. Krämer, S. Kopp, C. Becker-Asano, N. Sommer, Smile and the world will smile with you - The effects of a virtual agent's smile on users' evaluation and behavior, Int. J. Hum. Comput. Stud. 71(3) (2013) 335–349. https://doi.org/10.1016/j.ijhcs.2012.09.006.

[44] B. Reeves, C.I. Nass, The media equation: How people treat computers, television, and new media like real people and places, CSLI Publications, Cambridge University Press, New York, 1996.

[45] G.T. Kraft-Todd, D.A. Reinero, J.M. Kelley, A.S. Heberlein, L. Baer, H. Riess, 2017. Empathic nonverbal behavior increases ratings of both warmth and competence in a medical context. PLOS ONE. 12(5), e0177758. https://doi.org/10.1371/journal.pone.0177758.

[46] R.Y.J. Chua, P. Ingram, M.W. Morris, From the head and the heart: Locating cognition- and affect-based trust in managers' professional networks, Acad. Manag. J. 51(3) (2008) 436–452. https://doi.org/10.5465/amj.2008.32625956.

[47] D.J. McAllister, Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations, Acad. Manag. J. 38(1) (1995) 24–59. https://doi.org/10.2307/256727.

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

[49] L. Christoforakos, A. Gallucci, T. Surmava-Große, D. Ullrich, S. Diefenbach, 2021. Can robots earn our trust the same way humans do? A systematic exploration of competence, warmth, and anthropomorphism as determinants of trust development in HRI. Front. in Robot. AI. 8, 640444. https://doi.org/10.3389/frobt.2021.640444.

[50] E. Glikson, A.W. Woolley, Human trust in artificial intelligence: Review of empirical research, Acad. Manag. Ann. 14(2) (2020) 627–660. https://doi.org/10.5465/annals.2018.0057.

[51] D. Gefen, I. Benbasat, P. Pavlou, A research agenda for trust in online environments, J. Manag. Inf. Syst. 24(4) (2008) 275–286. https://doi.org/10.2753/MIS0742-1222240411.

[52] J.A. Hall, R. Schwartz, Empathy present and future, J. Soc. Psychol. 159(3) (2019) 225–243. https://doi.org/10.1080/00224545.2018.1477442.

[53] B.M.P. Cuff, S.J. Brown, L. Taylor, D.J. Howat, Empathy: A review of the concept, Emot. Rev. 8(2) (2016) 144–153. https://doi.org/10.1177/1754073914558466.

[54] D. Premack, G. Woodruff, Does the chimpanzee have a theory of mind?, Behav. Brain Sci. 1(4) (1978) 515–526. https://doi.org/10.1017/S0140525X00076512.

[55] U. Frith, Autism: Explaining the enigma, second ed., Blackwell, Oxford, 2003.

[56] A.R. Dennis, R.M. Fuller, J.S. Valacich, Media, tasks, and communication processes: A theory of media synchronicity, MIS Q. 32(3) (2008) 575–600. https://doi.org/10.2307/25148857.

[57] G. Hein, T. Singer, I feel how you feel but not always: The empathic brain and its modulation, Curr. Opin. Neurobiol. 18(2) (2008) 153–158. https://doi.org/10.1016/j.conb.2008.07.012.

[58] J.E. Escalas, B.B. Stern, Sympathy and empathy: Emotional responses to advertising dramas, J. Consum. Res. 29(4) (2003) 566–578. https://doi.org/10.1086/346251.

[59] C.D. Batson, The altruism question: Toward a social-psychological answer, Lawrence Erlbaum Associates, New York, 1991. https://doi.org/10.4324/9781315808048.

[60] C.D. Batson, B.D. Duncan, P. Ackerman, T. Buckley, K. Birch, Is empathic emotion a source of altruistic motivation?, J. Personal. Soc. Psychol. 40(2) (1981) 290–302. https://doi.org/10.1037/0022-3514.40.2.290.

[61] R.B. Cialdini, M. Schaller, D. Houlihan, K. Arps, J. Fultz, A.L. Beaman, Empathy-based helping: Is it selflessly or selfishly motivated?, J. Pers. Soc. Psychol. 52 (1987) 749–758. https://doi.org/10.1037/0022-3514.52.4.749.

[62] R.S. Lazarus, S. Folkman, Stress, appraisal, and coping, Springer, New York, 1984.

[63] J. Halpern, From detached concern to empathy: Humanizing medical practice, Oxford University Press, New York, 2001. https://doi.org/10.1093/acprof:osobl/9780195111194.001.0001.

[64] I. Leite, G. Castellano, A. Pereira, C. Martinho, A. Paiva, Empathic robots for long-term interaction evaluating social presence, engagement and perceived support in children, Int. J. Soc. Robot. 6(3) (2014) 329–341. https://doi.org/10.1007/s12369-014-0227-1.

[65] H. Nguyen, J. Masthoff, Designing empathic computers: The effect of multimodal empathic feedback using animated agent, in: S. Chatterjee, P. Dev (Eds.), Persuasive '09: Proceedings of the 4th International Conference on Persuasive Technology, Association for Computing Machinery, New York, 2009, 7. https://doi.org/10.1145/1541948.1541958.

[66] J. Klein, Y. Moon, R.W. Picard, This computer responds to user frustration: Theory, design, and results, Interact. Comput. 14(2) (2002) 119–140. https://doi.org/10.1016/S0953-5438(01)00053-4.

[67] J.-P. Stein, B. Liebold, P. Ohler, Stay back, clever thing! Linking situational control and human uniqueness concerns to the aversion against autonomous technology, Comput. Hum. Behav. 95 (2019) 73–82. https://doi.org/10.1016/j.chb.2019.01.021.

[68] M. Mori, K. MacDorman, N. Kageki, The uncanny valley, IEEE Robot. Autom. Mag. 19(2) (2012) 98–100. https://doi.org/10.1109/MRA.2012.2192811.

[69] Y.-J. Cha, S. Baek, G. Ahn, H. Lee, B. Lee, J. Shin, D. Jang, Compensating for the loss of human distinctiveness: The use of social creativity under human–machine comparisons, Comput. Hum. Behav. 103 (2020) 80–90. https://doi.org/10.1016/j.chb.2019.08.027.

[70] V. Pitardi, J. Wirtz, S. Paluch, W.H. Kunz, Service robots, agency and embarrassing service encounters, J. Serv. Manag. 33(2) (2022) 389–414. https://doi.org/10.1108/JOSM-12-2020-0435.

[71] S.T. Fiske, P.W. Linville, What does the schema concept buy us?, Pers. Soc. Psychol. Bull. 6(4) (1980), 543–557. https://doi.org/10.1177/014616728064006.

[72] W.B. Rouse, N.M. Morris, On looking into the black box: Prospects and limits in the search for mental models, Psychol. Bull. 100(3) (1986), 349–363. https://doi.org/10.1037/0033-2909.100.3.349.

[73] G.M. Grimes, R.M. Schuetzler, J.S. Giboney, 2021. Mental models and expectation violations in conversational AI interactions. Decis. Support Sys. 144, 113515. https://doi.org/10.1016/j.dss.2021.113515.

[74] D. Belanche, L.V. Casaló, C. Flavián, J. Schepers, Robots or frontline employees? Exploring customers' attributions of responsibility and stability after service failure or

success, J. Serv. Manag. 31(2) (2020) 267–289. https://doi.org/10.1108/JOSM-05-2019-0156.

[75] N. Castelo, J. Boegershausen, C. Hildebrand, A.P. Henkel, 2023. Understanding and improving consumer reactions to service bots. J. Consum. Res., ucad023. https://doi.org/10.1093/jcr/ucad023.

[76] A. Waytz, M.I. Norton, Botsourcing and outsourcing: Robot, British, Chinese, and German workers are for thinking - not feeling - jobs, Emot. 14(2) (2014) 434–444. https://doi.org/10.1037/a0036054.

[77] M. Heidegger, Being and time, State University of New York Press, Albany, 1996.

[78] T. Hennig-Thurau, M. Groth, M. Paul, D.D. Gremler, Are all smiles created equal? How emotional contagion and emotional labor affect service relationships, J. Mark. 70(3) (2006) 58–73. https://doi.org/10.1509/jmkg.70.3.058.

[79] A.T. Lechner, F. Mathmann, M. Paul, Frontline employees' display of fake smiles and angry faces: When and why they influence service performance, J. Serv. Res. 25(2) (2022) 211–226. https://doi.org/10.1177/1094670520975148.

[80] A.A. Grandey, G.M. Fisk, A.S. Mattila, K.J. Jansen, L.A. Sideman, Is "service with a smile" enough? Authenticity of positive displays during service encounters, Organ. Behav. Hum. Decis. Process. 96(1) (2005) 38–55. https://doi.org/10.1016/j.obhdp.2004.08.002.

[81] J.G. Moulard, R.D. Raggio, J.A.G. Folse, Brand authenticity: Testing the antecedents and outcomes of brand management's passion for its products, Psychol. Mark. 33(6) (2016) 421–436. https://doi.org/10.1002/mar.20888.

[82] M. Okubo, A. Kobayashi, K. Ishikawa, A fake smile thwarts cheater detection, J. Nonverbal Behav. 36(3) (2012) 217–225. https://doi.org/10.1007/s10919-012-0134-9.

[83] R. Schroll, "Ouch!" When and why food anthropomorphism negatively affects consumption, J. Consum. Psychol. 33(3) (2023), 561–574. https://doi.org/10.1002/jcpy.1316.

[84] L. Wang, S. Kim, X. Zhou, Money in a "safe" place: Money anthropomorphism increases saving behavior, Int. J. Res. Mark. 40(1) (2023), 88–108. https://doi.org/10.1016/j.ijresmar.2022.02.001.

[85] W.F. Laughey, M.E.L. Brown, A.N. Dueñas, R. Archer, M.R. Whitwell, A. Liu, G.M. Finn, How medical school alters empathy: Student love and break up letters to empathy for patients, Med. Educ. 55(3) (2021), 394–403. https://doi.org/10.1111/medu.14403.

[86] SnatchBot, SnatchBot. https://snatchbot.me, 2023 (accessed 22nd March 2024).

[87] J. Cohen, Statistical power analysis for the behavioral sciences, second ed., Lawrence Erlbaum Associates, Hillsdale, 1988.

[88] J.L. Aaker, K.D. Vohs, C. Mogilner, Nonprofits are seen as warm and for-profits as competent: Firm stereotypes matter, J. Consum. Res. 37(2) (2010) 224–237. https://doi.org/10.1086/651566.

[89] M. Söllner, A. Hoffmann, H. Hoffmann, A. Wacker, J.M. Leimeister, Understanding the formation of trust in IT artifacts, in: ICIS 2012 Proceedings, Association for Information Systems, Atlanta, 2012, 11.

[90] D.H. McKnight, V. Choudhury, C. Kacmar, Developing and validating trust measures for e-commerce: An integrative typology, Inf. Syst. Res. 13(3) (2002) 334–359. https://doi.org/10.1287/isre.13.3.334.81.

[91] V. Venkatesh, J.Y.L. Thong, X. Xu, Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology, MIS Q. 36(1) (2012) 157–178. https://doi.org/10.2307/41410412.

[92] J.-W. Moon, Y.-G. Kim, Extending the TAM for a world-wide-web context, Inf. Manag. 38(4) (2001) 217–230. https://doi.org/10.1016/S0378-7206(00)00061-6.

[93] A.F. Hayes, Introduction to mediation, moderation, and conditional process analysis: A regression-based approach, second ed., Guilford Press, New York, 2018.

[94] E. Efendić, P.P.F.M Van De Calseyde, A.M. Evans, Slow response times undermine trust in algorithmic (but not human) predictions, Organ. Behav. Hum. Decis. Process. 157 (2020), 103–114. https://doi.org/10.1016/j.obhdp.2020.01.008.

[95] S. Diederich, M. Janßen-Müller, A.B. Brendel, S. Morana, Emulating empathetic behavior in online service encounters with sentiment-adaptive responses: Insights from an experiment with a conversational agent, in: ICIS 2019 Proceedings, Association for Information Systems, Atlanta, 2019, 2.

[96] A. Gambino, J. Fox, R. Ratan, Building a stronger CASA: Extending the Computers Are Social Actors Paradigm, Hum.-Mach. Commun. 1 (2020), 71–86. https://doi.org/10.30658/hmc.1.5.

Appendices

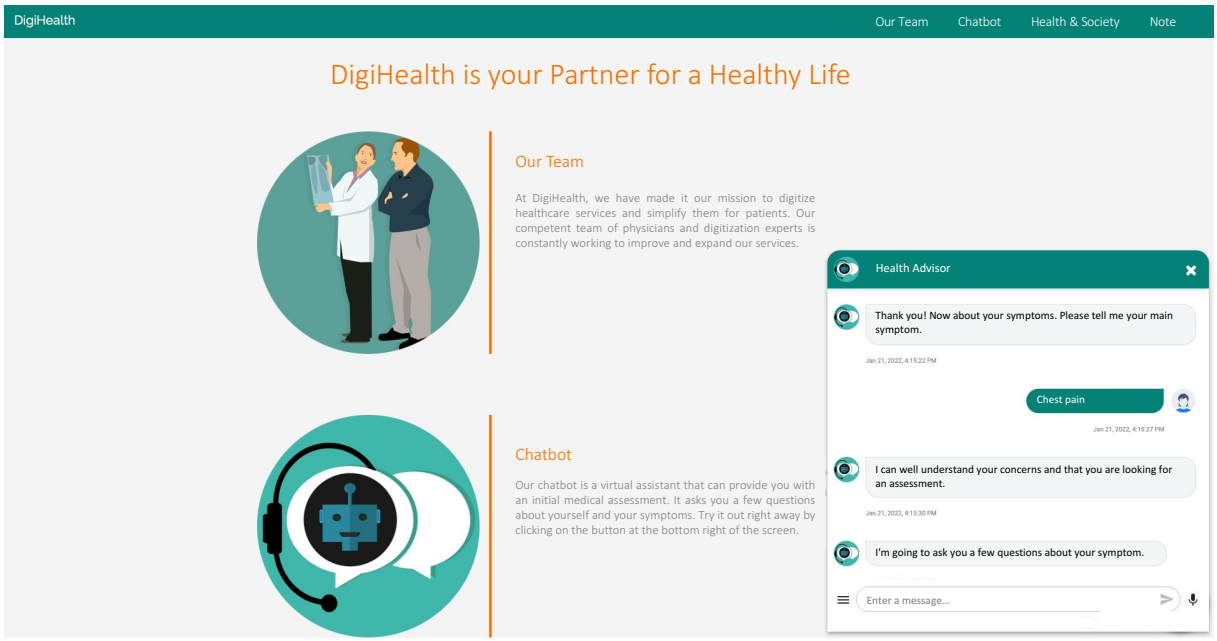
Appendix A. Measurements of Pre-Test (Study 1)

Measurement/Items	Cronbach's Alpha
Perceived empathetic responses (self-developed based on empathy theories)	.93
The chatbot has expressed being able to empathize with the patient's feelings.	
The chatbot has indicated it could put itself well in the patient's shoes.	
The chatbot was able to accurately understand the patient's concerns.	
Perceived sympathetic responses (self-developed based on empathy theories)	.89
The chatbot was compassionate about the patient's situation.	
The chatbot has indicated to feel sorry for the patient.	
That chatbot has expressed its sympathy.	
Perceived behavioral-empathetic responses (self-developed based on empathy theories)	.91
The chatbot has expressed the intention to support the patient.	
The chatbot has encouraged the patient.	
The chatbot was really interested in helping the patient.	
Perceived overall empathy	-
To what extent did you generally feel a sense of empathy in the chatbot? (1) no empathy at all vs. (7) much empathy	
Scenario's realism (adapted from Gelbrich et al. [38])	.93
The chatbot could exist in reality.	
I was able to imagine the situation very well.	
The interaction between the chatbot and patient was realistic.	
Overall, the scenario was credible.	
Conversation's complexity	-
The interaction was complex.	

Note: Seven-point Likert scales with 1 = "strongly disagree" and 7 = "strongly agree" (if not indicated otherwise).

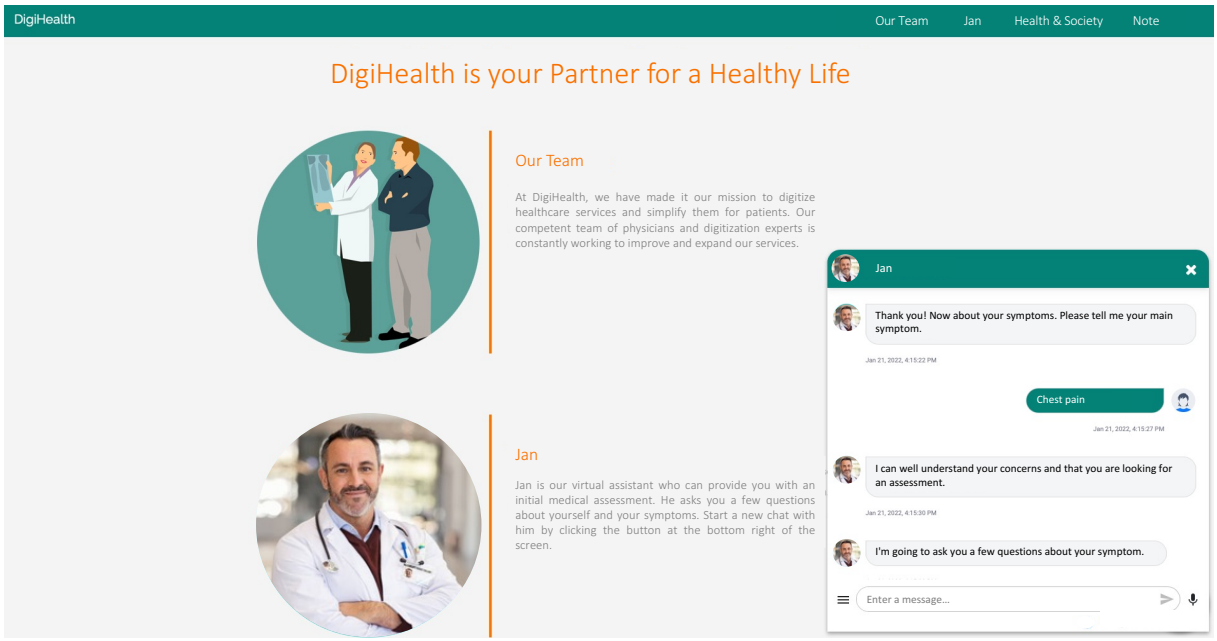
Appendix B. Screenshots of Websites and Chatbot Pop-Ups (Study 1 and Study 2)

Study 1



Note: The example shows a sequence of the interaction with the empathetic chatbot.

Study 2



Note: The example shows a sequence of the interaction with the empathetic chatbot.

Appendix C. Measurements of All Main Studies (Study 1, Study 2, Study 3)

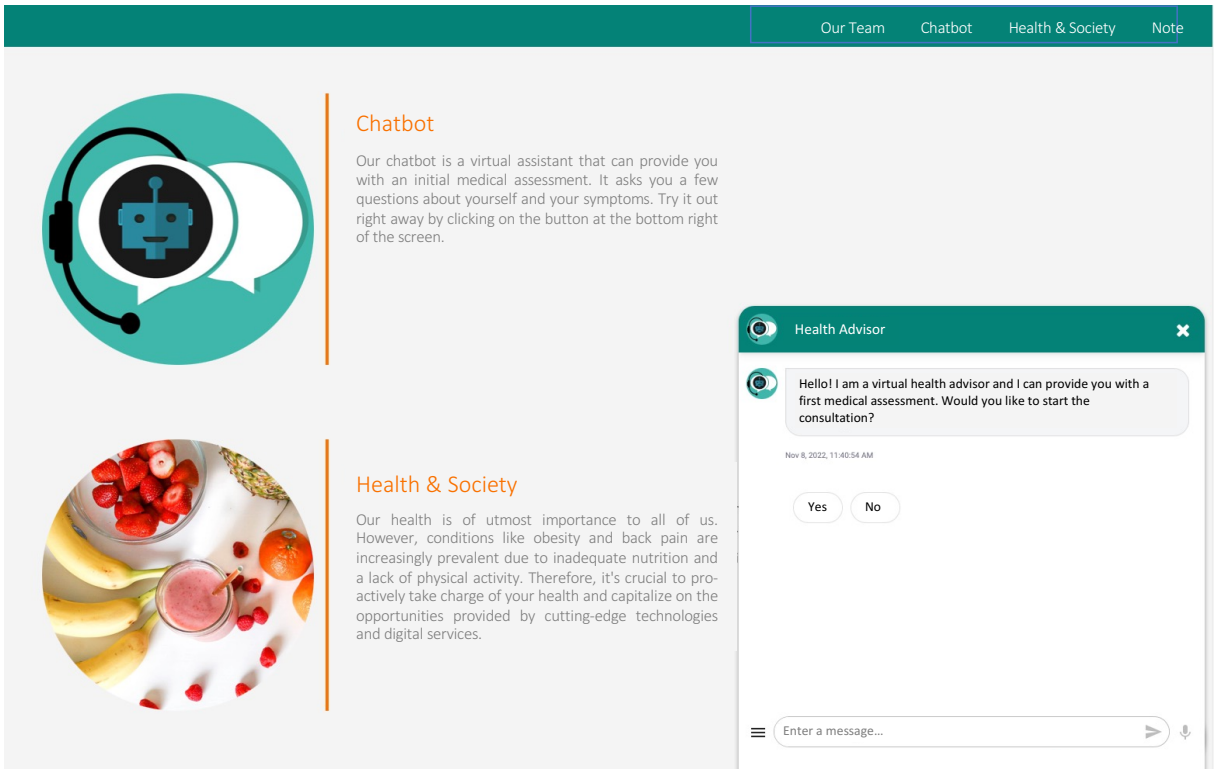
Measurement/Items	Cronbach's Alpha
Perceived warmth (adapted from Gelbrich et al. [38] and Aaker et al. [88])	.77/.81/.88
The chatbot (physician) was…	
…warm.	
…kind.	
…friendly.	
Perceived authenticity (self-developed based on theories on authentic personality)	.84/.85/.78
The chatbot (physician) tried to pretend to be something it (he) is not.	
The chatbot's (physician's) interaction style was credible.	
I sometimes felt the chatbot (physician) was faking out.	
The chatbot's (physician's) messages seemed put-on.	
The chatbot (physician) was play-acting just to please patients.	
Willingness to trust (adapted from Söllner et al. [89] and McKnight et al. [90])	.90/.91/.95
I would feel comfortable relying on the chatbot's (physician's) assessment.	
I would not hesitate to follow the chatbot's (physician's) advice.	
I would confidently follow the chatbot's (physician's) recommendations.	
I would not doubt the chatbot's (physician's) assessment.	
I would count on the chatbot (physician) to help me with health issues.	
Overall, the chatbot (physician) seems trustworthy.	
Using intentions (adapted from Venkatesh et al. [91])	.93/.93/.94
If I had access to the chatbot (health service)…	
…I intend to continue to use it for the next medical assessment.	
…I can well imagine to use it for the next medical assessment.	
…I would always try to use it if I had health issues.	
General attitudes (adapted from Moon & Kim [92])	.91/.94/.94
In general, I consider the idea of using a healthcare chatbot (a chat with a physician)…	
…(1) bad vs. (7) good	
…(1) foolish vs. (7) wise	

…(1) unpleasant vs. (7) pleasant	
…(1) negative vs. (7) positive	
Perceived physical risk (self-developed)	.92/.94/.90
I think the clinical picture in the scenario is a great threat to my health.	
I consider the potential consequences of the clinical picture in the scenario threatening.	
I am concerned that there is a high health risk associated with the clinical picture in the scenario.	

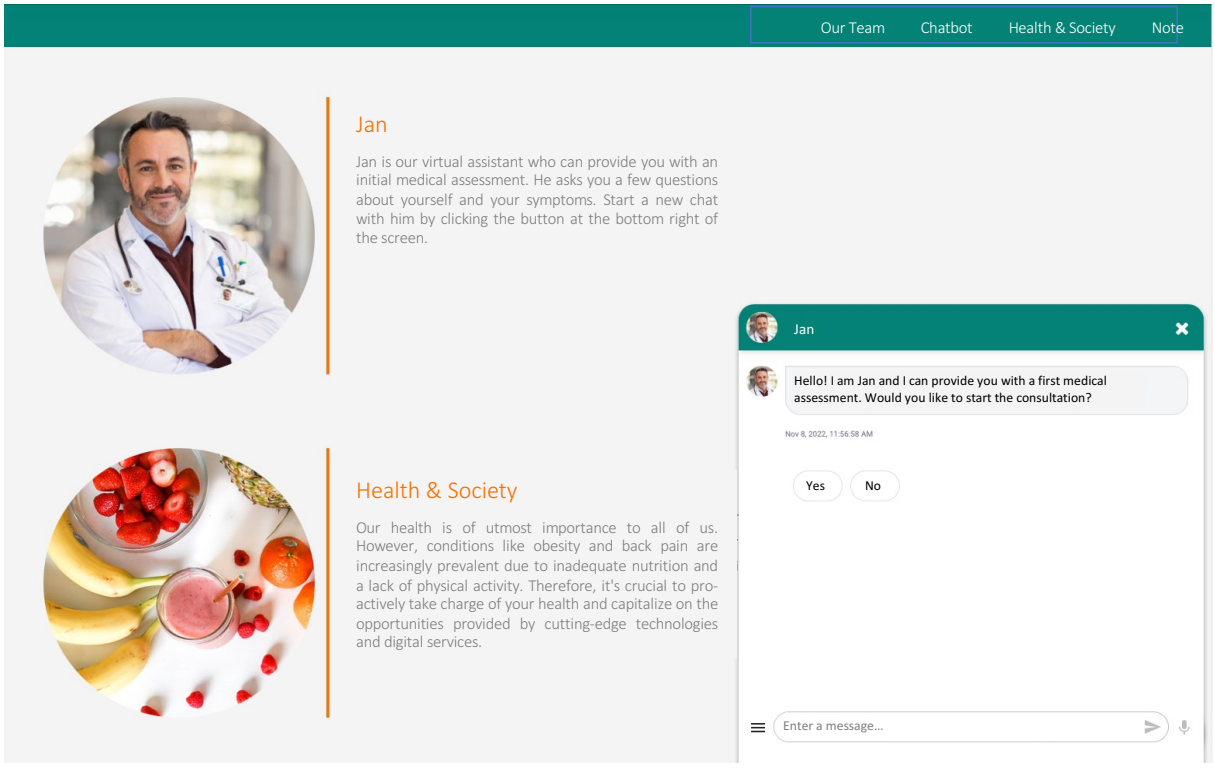
Note: Seven-point Likert scales with 1 = "strongly disagree" and 7 = "strongly agree" (if not indicated otherwise).

Appendix D. Screenshots of Pre-Test (Study 2)

Non-Personified Chatbot



Personified Chatbot



Appendix E. Measurements of Pre-Test (Study 2)

Measurement/Items	Cronbach's Alpha
Personification (adapted from Crolic et al. [20])	.92
The chatbot seemed like a person to me.	
The chatbot seemed human.	
I felt the chatbot has a personality of its own.	

Note: Seven-point Likert scale with 1 = "strongly disagree" and 7 = "strongly agree".

Appendix F. Screenshot of the Human Agent Interaction (Study 3)

PART IV: PAPER 3

BOTS HAVE TO BE FAST: THE DETRIMENTAL EFFECTS OF RESPONSE DELAYS IN SERVICE CHATBOTS AND THE MODERATING ROLE OF ANTHROPOMORPHISM

Fact Sheet Paper 3

Title	Bots Have to Be Fast: The Detrimental Effects of Response Delays in Service Chatbots and the Moderating Role of Anthropomorphism
Authors	Lennart Seitz & Sigrid Bekmeier-Feuerhahn
Year	2023
Target journal	Journal of Service Research
Publication status	Invited to re-submit on February 4, 2024 (status: June 5, 2024)

Abstract

With the increasing prevalence of service chatbots, a broad debate has emerged concerning the advantages and disadvantages of humanizing them. The present research examines whether a social cue reducing a chatbot's efficiency (dynamic response delays) backfires by interfering with the expectation of receiving a fast service that is a key benefit of chatbots over human agents. Across five studies, we find evidence that dynamic response delays have adverse effects on usage intentions and service provider evaluation. We also illuminate the underlying mechanisms by showing that the backfiring effect stems from a violation of usefulness expectations resulting from the activation of computer-like schemas and associations in chatbot interactions. Congruently, the negative effect is attenuated when users apply human-like schemas. This research thus demonstrates that social cues can have adverse effects when reducing a chatbot's efficiency that is crucial for technology acceptance. Furthermore, it shows that the application of computer- vs. human-like schemas guide users' expectation towards a chatbot's behavior, consequently exerting a significant impact on the perception of social cues. Practitioners are therefore recommended to consider potential drawbacks of specific social cues and to align them with users' expectations towards chatbots and their key benefits rather than striving for maximizing human-likeness.

Keywords: chatbot; anthropomorphism; response delays; technology acceptance; expectancy violations

1 Introduction

The launch of "ChatGPT" in November 2022 has initiated a significant surge in the utilization of chatbots in everyday life. Even before, chatbots were frequently employed in service delivery, offering convenience to customers, and benefiting firms by standardizing services, boosting efficiency, and cutting costs through complementing or replacing human employees (Larivière et al. 2017; Sheehan, Jin, and Gottlieb 2020). Chatbots are therefore considered a pivotal technology transforming the service delivery landscape (Huang and Rust 2018; Wirtz et al. 2018; Yu, Xiong, and Shen 2022). The global chatbot market is thus projected to exceed $6 billion in 2023 and to reach $27 billion in 2030 (Grand View Research 2023).

As chatbots and robots increasingly replace service employees, a debate has arisen regarding the desirability of making them more human-like. Although there is meta-analytical evidence for positive effects (Blut et al. 2021), scholars increasingly identify backfiring effects and boundary conditions (e.g., Mende et al. 2019; Crolic et al. 2022; Holthöwer and van Doorn 2022; Han, Deng, and Fan 2023). The objective of the present research is to enhance our understanding of when and why specific social cues in service chatbots may have adverse effects as there is still a lot to learn (Blut et al. 2021; Uysal, Alavi, and Bezençon 2022; Han, Deng, and Fan 2023). Precisely, this paper examines if a social cue backfires when it contradicts one of the main purposes and advantages of chatbots over human agents – enhancing the efficiency of service delivery.

In approaching the research objective, this paper focuses on the non-verbal social cue of response delays that is frequently used to make chatbot conversations feel more natural and human-like (Feine et al. 2019; Schanke, Burtch, and Ray 2021). Chatbots using response delays do not respond immediately but pretend to need some time for typing-in a message. A practical example illustrating their significance is the case of Lufthansa's chatbot "Mildred" who was equipped with response delays after customers complained about its unnaturally fast responses

(Crozier 2017). To bring chatbot interactions even closer to interhuman conversations, literature discusses on "dynamic response delays" that vary depending on the message's length (Schanke, Burtch, and Ray 2021; Gnewuch et al. 2022).

However, this paper assumes that dynamic response delays could surpass the goal of achieving pleasing human-likeness and result in adverse effects. We posit that entering a conversation with a service chatbot might elicit computer-like schemas shaping the expectation that it communicates like a bot, i.e., that it responds immediately and makes service provision more efficient (Grimes, Schuetzler, and Giboney 2021; Meng and Dai 2021). As enhancing efficiency is vital for considering a technology to be useful, we argue that dynamic response delays could lead to expectancy violations leading to a negative evaluation of the chatbot and the service provider (Venkatesh, Thong, and Xu 2012; Blut, Wang, and Schoefer 2016). We further assume that this backfiring effect is attenuated when customers apply human-like schemas in the chatbot interaction (i.e., when they expect it to communicate like a human or when the task is complex).

To empirically test the hypotheses, we conducted five experimental studies in which participants either watched videos of an interaction between a customer and a service chatbot (Study 1 and 2) or interacted with responsive chatbots (Study 3–5). We mainly manipulated the chatbot's response behavior (dynamic response delays vs. no such delays) and examined the impact on usage intentions (Study 1–4) and service provider evaluation (Study 5). To approach the underlying mechanisms, we tested the mediating role of a reduction in perceived usefulness in all studies and the moderating role of the application of computer- vs. human-like schemas. We used both internal indicators for the participants' tendency to apply computer- vs. human-like schemas (i.e., their tendency to anthropomorphize chatbots, Study 1 and 3) or external manipulations by comparing (1) a chatbot with a human agent (Study 2) and (2) a computer- vs. human-like service task (Study 4).

By examining a potential novel backfiring effect of social cues in chatbots and uncovering the underlying mechanisms, this paper contributes to existing literature in several ways: first, it links research on humanizing chatbots, technology acceptance models, and service research. Precisely, it showcases that social cues contradicting central dimensions of technology acceptance (i.e., perceived usefulness) can have adverse effects on the evaluation of the chatbot and the service provider. To the best of our knowledge, there is barely research yet taking this interdisciplinary perspective. Second, this study is among the first to demonstrate that internally or externally triggered schemas significantly influence our expectations of a chatbot's behavior and the service process, consequently exerting a significant impact on the perception and evaluation of social cues. It therefore takes a critical perspective on "Social Response Theory" (Reeves and Nass 1996; Nass and Moon 2000) by showing that humans might not generally apply interpersonal heuristics in their interactions with chatbots. And third, this research demonstrates that there is a further need to study how different factors on the individual and contextual level influence the perception of specific social cues in service chatbots. As AI-driven technology will further improve and is expected to penetrate relational services in future, it is vital to approach how conversations with chatbots can be imbued with a sense of human touch without risking unfavorable consequences (Blut et al. 2021; Huang and Rust 2021). This might not only enhance our theoretical understanding of human-chatbot interactions but can also help companies in designing more satisfying service chatbots.

2 Conceptual Background and Hypotheses

2.1 Social Chatbots

A key distinction between past generations of self-service tools (e.g., online forms) and chatbots is their highly responsive and social nature (Blut et al. 2021; Grimes, Schuetzler, and Giboney 2021). The interaction mode of a service chatbot mirrors that of a human agent, i.e., a customer starts a conversation by messaging a chatbot and subsequently receives personalized

responses aimed to address the service request. To make conversations feel even more natural, chatbots frequently incorporate human-like design elements also known as social cues. For example, chatbots can have a name, a human-like avatar, or use verbal social cues (Feine et al. 2019; Crolic et al. 2022). Integrating social cues can create a sense of social presence, which is crucial for fulfilling social-emotional and relational needs in interactions with bots (van Doorn et al. 2017; Wirtz et al. 2018; Huang and Rust 2021). Because of humans' innate social nature, perceiving a sense of human-likeness in bots can elevate positive emotions and perceived warmth, resulting in higher service satisfaction and usage intentions (Choi, Mattila, and Bolton 2021; Gelbrich, Hagel, and Orsingher 2021).

However, there is also evidence for adverse effects (Blut et al. 2021). First, too much human-likeness can cause a perceived threat to human identity which can elevate coping strategies like compensatory consumption behavior (Mende et al. 2019) or putting higher value on uniquely human attributes such as emotions (Cha et al. 2020). Second, human-like design elements in service chatbots can lead to higher capability attributions that might boost frustration in case of service failure (Crolic et al. 2022). Third, users may have concerns about facing social judgment from human-like bots, potentially leading to adverse outcomes in service environments characterized by a high social risk or situations deemed embarrassing (Holthöwer and van Doorn 2022; Kim et al. 2022). And fourth, the effects of human-like design elements do not only rely on the bot's characteristics or the service environment but also on individual traits. For example, individuals with a competitive mindset may show less favorable responses to human-like AI than those with a collaborative mindset (Han, Deng, and Fan 2023). Subsuming, there is an increasing number of research uncovering backfiring effects of human-like design elements. It is hence crucial to enhance our understanding of when and why social cues have favorable vs. unfavorable consequences as there is ambiguity (Blut et al. 2021; Uysal, Alavi, and Bezençon 2022).

2.2 Response Delays

A frequently used social cue in chatbots are response delays, also known as response latencies (Gnewuch et al. 2022). Instead of responding to user input immediately, the chatbot makes use of typing indicators (e.g., three dots) to pretend to need some time for generating a message. The intent is to make the chatbot appear more human-like as humans usually need some time to read and understand a message before typing-in a response (Moon 1999; Jacquet, Baratgin, and Jamet 2019). Many apps used for interpersonal communication via chat (e.g., "WhatsApp" or "iMessage") use different typing indicators to let the receiver of a message know that the sender is present and already about to respond. For instance, "WhatsApp" shows the information "*typing…*" below a user's name while "iMessage" shows three slightly moving dots before a message appears.

A chatbot that is based on AI or uses pre-scripted conversation flows neither needs to read a message line by line nor does it have to type-in a response letter by letter. The implementation of response delays is hence not necessary in improving a chatbot's performance or accuracy but contributes to enhance perceived social presence and the conversation's naturalness (Schanke, Burtch, and Ray 2021; Gnewuch et al. 2022). Response delays can be either static (e.g., two seconds per message) or dynamic meaning that the length of the response delay depends on the length of the message the chatbot is about to send. Dynamic response delays (e.g., delaying a response by a defined time per character) might move the chatbot's communication behavior even more towards an interpersonal interaction thus maximizing its perceived human-likeness (Holtgraves and Han 2007).

Scientific research on the effects of (dynamic) response delays in chatbots is scarce and ambivalent. On the one hand, response delays can enhance perceived humanness and social presence facilitating satisfaction with the interaction and usage intentions (Gnewuch et al. 2022). Also, users might interpret response delays as a system's effort to generate a high-quality

and accurate outcome (Tsekouras, Li, and Benbasat 2022). On the other hand, long response delays could be interpreted as a system's failure, i.e., that the chatbot is not working properly thus harming perceived service quality. Congruently, previous studies have shown that response delays can decrease the likeability of a virtual assistant (Schanke, Burtch, and Ray 2021) and that long (vs. short) response delays in interactions with computers can reduce the system's persuasiveness (Moon 1999). Also, the positive effect of response delays on perceived social presence and the intention to use a chatbot was found to be moderated by users' experience with chatbots, i.e., it was only positive for novice users but reversed for experienced ones (Gnewuch et al. 2022). Hence, there is evidence that the effect of response delays on user perception could be impacted by individual traits or contextual factors that shape expectations towards the chatbot.

2.3 Expectancy Violations

The previous section closed with a statement that the perception of response delays and the resulting consequences might depend on expectations users have towards the chatbot and its communication behavior. When entering a conversation with an interaction partner – either human or artificial –, people usually have a priori expectancies on how the counterpart will respond. These expectancies are highly influenced by social norms, individual experiences, and contextual factors (Burgoon 1993). This also applies when users enter a conversation with a chatbot, i.e., they have a priori expectancies on how the interaction will move forward (Grimes, Schuetzler, and Giboney 2021). Common expectations towards chatbots include responding in a cold and impersonal manner (Meng and Dai 2021), being less flexible and effective than human agents (Crolic et al. 2022; Yu, Xiong, and Shen 2022), and having overall weaker conversation capabilities (Grimes, Schuetzler, and Giboney 2021). Instead, chatbots are expected to provide convenient, fast, and goal-oriented services as they can respond to customer requests immediately. This enhanced convenience and efficiency is also one of the key reasons

for and advantages of using chatbots compared to consulting human agents (Sheehan, Jin, and Gottlieb 2020; Yu, Xiong, and Shen 2022).

A chatbot with dynamic response delays could potentially violate this expectation by responding unexpectedly slowly, which could result in a negative evaluation of the chatbot (Schanke, Burtch, and Ray 2021). This argument finds theoretical support in "Expectancy Violations Theory" (EVT) that originates from communication studies and aims to explain how people react when their expectations in a communication situation are violated (Burgoon 1993). According to EVT, a violation can either have a positive valence (when prior expectations are exceeded) or a negative valence (when prior expectations are not met). EVT suggests that violations create arousal and stimulate cognitive processing as people try to make sense of the violator's behavior. Previous research in the domains of marketing and information systems has adopted EVT to interactions with chatbots and showed that users can be disappointed when a system does not meet a priori expectations (Grimes, Schuetzler, and Giboney 2021; Crolic et al. 2022). Consequently, the present research argues that dynamic response delays backfire as they negatively violate a priori expectations regarding a chatbot's response time thus contradicting its key benefit, i.e., making service provision faster. Precisely, this research hypothesizes that usage intentions of a chatbot decrease when it is equipped with dynamic response delays.

H1: Dynamic response delays (vs. no such delays) in a service chatbot decrease usage intentions.

2.4 Anticipated Utilitarian Advantages and Perceived Usefulness

To delve deeper into the underlying mechanism, it is essential to theoretically elaborate which specific expectations are violated by dynamic response delays. As discussed, users might expect service chatbots to make service provision faster and to enhance task performance (Yu, Xiong, and Shen 2022). In other words, they anticipate receiving utilitarian advantages that is

a major extrinsic motivator for using service chatbots. Utilitarian advantages refer to the functional benefits a system provides to help users improve their efficiency and productivity in accomplishing specific tasks (Venkatesh, Thong, and Xu 2012). Scholars from information systems research agree that perceiving utilitarian advantages is essential for behavioral intentions and technology acceptance. Hence, well-established theories like the "Unified Theory of Acceptance and Use of Technology" (UTAUT; Venkatesh, Thong, and Xu 2012) and the "Technology Acceptance Model" (TAM; Davis 1989) account for the importance of a technology's utility by the dimensions of "performance expectancy" (UTAUT) or "perceived usefulness" (TAM). To enhance conceptual clarity, this research refers to "perceived usefulness" in defining the extent a technology is perceived to provide utilitarian advantages.

Previous research has repeatedly identified a technology's perceived usefulness to be the primary predictor for usage intentions, including AI and chatbots (Blut, Wang, and Schoefer 2016; Lee and Lyu 2016). Referring to EVT, the extent a technology is evaluated useful highly depends on a priori expectations and posteriori experiences with the system. When a priori expectations are met or exceeded, perceived usefulness is likely to be elevated, but when expectations are not met, perceived usefulness may be diminished. Prior research has shown that meeting or exceeding prior expectations regarding perceived usefulness determine usage intentions and satisfaction with a system (Brown, Venkatesh, and Goyal 2014).

As dynamic response delays may contradict the anticipated utilitarian advantages of using service chatbots, expectations could be violated resulting in a lower perceived usefulness. This research thus hypothesizes that the negative effect of dynamic response delays on usage intentions is mediated by a loss in perceived usefulness.

H2: The negative effect of dynamic response delays on usage intentions is mediated by a loss in perceived usefulness.

2.5 Schemas, Anthropomorphism, and Social Responses to Computers

The previous sections argued that expecting immediate responses stems from the knowledge that chatbots are technological tools designed to enhance service efficiency. The knowledge people have about the attributes of objects is organized in cognitive frameworks that are known as "schemas" in psychology (Fiske and Linville 1980). Schemas specify the defining and relevant attributes of objects and guide us on how things usually look, feel, or behave (Halkias 2015). Hence, schemas are highly relevant in our perception and evaluation of products and brands (Aggarwal and McGill 2007; Aggarwal and McGill 2012), or software systems like chatbots (Grimes, Schuetzler, and Giboney 2021). Regarding software systems in particular, literature often refers to "mental models" when describing schemas people have about a system. Mental models represent users' assumptions about a system's purpose, how it operates, what it is doing, and what it looks like (Rouse and Morris 1986). These mental models help users in understanding and predicting a system's behavior thus being essential for humans' expectations towards a system.

Hitherto, this paper assumed that users generally apply computer-like schemas to chatbots, expecting cold, quick, and impersonal responses (Meng and Dai 2021). However, the human-like interaction style of a chatbot might also elicit human-like schemas leading users to perceive and treat them like social actors. This phenomenon is theorized both in computer science ("Social Response Theory"; Reeves and Nass, 1996; Nass and Moon 2000) and psychology ("anthropomorphism"; Epley, Waytz, and Cacioppo 2007). Put simply, "Social Response Theory" argues that humans mindlessly apply social rules and expectations to computers, particularly when they have social cues. In contrast, anthropomorphism is not limited to computers but all kinds of non-human agents. It also goes beyond mindless social responses as it describes an individual's tendency to ascribe human-like attributes to non-human entities, e.g., emotions, intentions, and behavior. This tendency depends not only on the

interaction object's degree of human-likeness, but also on individual traits and contextual factors. For instance, research has shown that people with a high sociality motivation (e.g., lonely persons) are more susceptible for anthropomorphism (Epley et al. 2008). Also, people are more likely to anthropomorphize non-human agents when they have little knowledge about them and there is a high level of uncertainty. In this case, anthropomorphism is an intuitive method to enhance the feeling of being able to explain and predict an agent's behavior (Epley, Waytz, and Cacioppo 2007).

The application of human-like schemas to objects and interactions can have several perceptual and behavioral consequences. For instance, anthropomorphism has been found to elicit the treatment of objects like cars (Aggarwal and McGill 2007), brands (Puzakova and Kwak 2017), money (Wang, Kim, and Zhou 2023), or chatbots (Blut et al. 2021) as if they were human. Also, "Social Response Theory" and findings from related research suggest that humans might apply the same social rules, heuristics, and expectations to computers and bots (Reeves and Nass 1996; Nass and Moon 2000). Applied to the present research, chatbots might be expected to communicate like humans, i.e., that they need some time to respond. There is supporting evidence as response delays were only found to have a positive impact on perceived social presence for novice but not experienced chatbot users who have richer mental representations of what a chatbot is and how it works (Gnewuch et al. 2022). If the hypothesized backfiring effect of dynamic response delays truly emanates from the application of computer-like schemas, the effect should be attenuated when people apply human-like schemas in the interaction. This research hence hypothesizes that an individual's tendency to anthropomorphize chatbots moderates the backfiring outlined in *H1* and *H2*.

H3: The negative indirect effect of dynamic response delays on usage intentions mediated by a loss in perceived usefulness is moderated by an individual's tendency to anthropomorphize chatbots.

To further enhance validity of our research, we hypothesize that the backfiring effect does not occur when users believe the agent to be a human (vs. a chatbot). As people should apply human-like schemas in human-human interactions, they should not expect a human agent to respond immediately, and there should be no expectancy violations.

H4: There is no expectancy violations-induced backfiring effect of dynamic response delays when users believe the agent to be a human (vs. a chatbot).

Lastly, the effect could also be attenuated when the chatbot performs a human-like (vs. computer-like) service task. First, human-like service tasks (e.g., healthcare provision) might unconsciously activate human-like schemas due to their strong association with the need for a human advisor. Hence, humans might seek for social connectedness and reassurance facilitating the adoption of social heuristics and expectations. In contrast, consumers have no incentive to apply human-like schemas when the task is computer-like (e.g., retrieving data from a data base) as they rather expect the chatbot to maximize efficiency (Seeger, Pfeiffer, and Heinzl 2021). Second, human-like tasks typically entail greater complexity making it more challenging to accomplish them independently. Hence, self-efficacy is lower and expected time effort higher. To test a managerial relevant boundary condition for the assumed backfiring effect, we hypothesize the effect to be moderated by the service task's human-likeness.

H5: The negative indirect effect of dynamic response delays on usage intentions mediated by a loss in perceived usefulness is moderated by the service task's human-likeness.

Figure 1 presents a comprehensive conceptual model summarizing the hypotheses that will be tested empirically (see Figure 1).

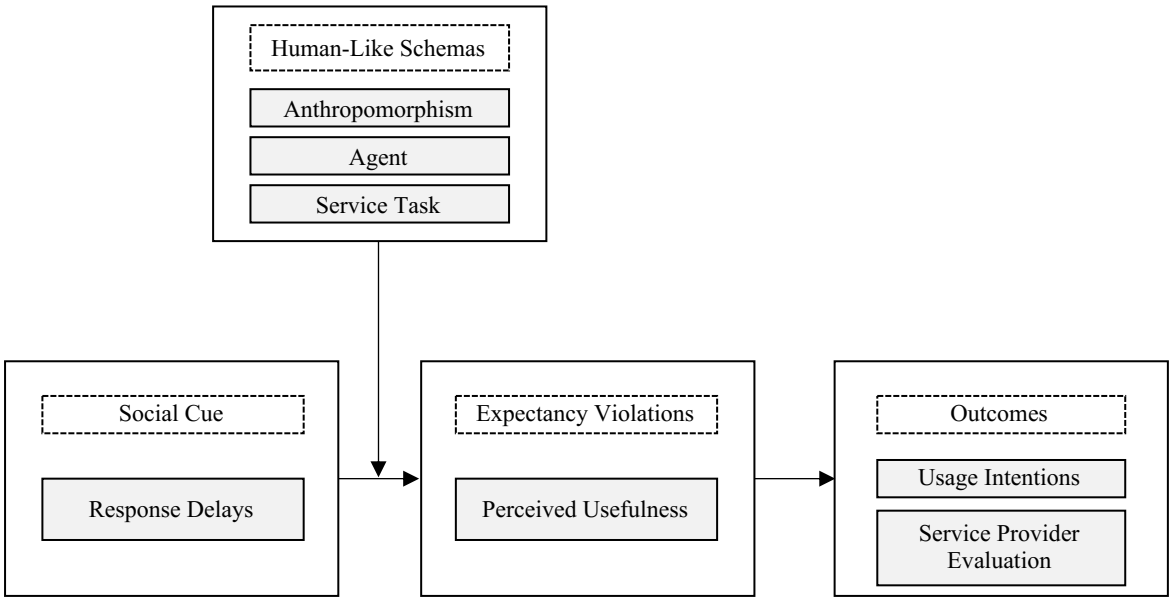


Figure 1. Conceptual model.

3 Study 1: Pilot Study

Study 1 sought to initially test *H1–H3* by showing participants a video of a screen-recorded interaction between a travel booking chatbot and a customer who receives help in finding a hotel for a city trip. We started with using video stimuli instead of real interactions for two reasons: first, it enabled equality of treatment across individuals and conditions, hence increasing internal validity. Second, employing screenshot or video vignettes is a commonly used and economical procedure to simulate human-bot interactions (Castelo et al. 2023).

3.1 Stimuli and Pre-Test

Study 1 manipulated the *chatbot identity* by applying a one-factor design with three levels: *non-social* (NS) vs. *social* (SO) vs. *social delays* (SD). The *non-social* did not have any social cues, e.g., no name and no human-like avatar. We decided to include both a *non-social* baseline condition and a *social* without response delays to account for potential positive effects social cues (except from response delays) might have on perceived usefulness and usage intentions as there is ambivalence in literature (Blut et al. 2021). If the hypothesized adverse effect only occurs for the *social delays* but not the *social* without delays, we can clearly declare

it as backfiring effect emanating from a too human-like response behavior conflicting with users' expectations. This helps us to provide a more nuanced perspective on when and which kind of social cues in chatbots have adverse effects. Besides, the central manipulation was the chatbot's response time. While the *social delays* was equipped with dynamic delays varying between four and eighteen seconds depending on the message's length, the *non-social* and the *social* responded with a latency of only one second. We decided to include a one second delay to prevent confusion and overload for the participants (Crozier 2017).

The screen-recorded interactions were simulated by two involved individuals in Apple's "iMessage" one taking the role of the chatbot and one that of the customer (see Appendix and Web Appendix A1). We conducted a pre-test with n=129 valid respondents (M_{age}=35.27, SD_{age}=12.39, 55.0% female) to ensure the manipulation's effectiveness. Participants were randomly assigned to one of the three conditions and watched the video before filling out a standardized questionnaire capturing *perceived human-likeness* (Kim, Chen, and Zhang 2016; α=.92) and the interaction's *perceived duration* (self-developed, α=.91) by three seven-point items each. We further measured *perceived realism* by four seven-point items (Gelbrich, Hagel, and Orsingher 2021, α=.89) to ensure that all conditions are equally realistic (see Web Appendix A2.1). Three one-way ANOVAs revealed that *perceived human-likeness* was higher for both chatbots with social cues compared to the *non-social* condition (M_{NS}=3.63; M_{SO}=4.70; M_{SD}=4.59, F(2, 126)=5.634, p=.005), and that *perceived duration* was higher for the *social delays* vs. the chatbots without a delay (M_{NS}=3.47; M_{SO}=3.38; M_{SD}=4.68, F(2, 126)=11.685, p<.001). Lastly, *perceived realism* did not differ across conditions, M>5.47, p=.603.

3.2 Sample and Procedure

We calculated the required sample size a priori using G*Power 3.1. Assuming a small to medium effect size (f=.18; Cohen 1988) and striving for a *power level* of .80 and *α error probability* of p=.05, the required minimum sample size is n=301. We recruited n=444

participants using a research panel. After cleaning the data set for invalid respondents (i.e., attention check failures, n=45), the final sample consisted of n=399 individuals (M_{age}=35.57, SD_{age}=12.49, 53.1% female). Participants were randomly assigned to one of the three different conditions and watched the video before filling out a standardized questionnaire capturing *usage intentions* (α=.95) by three items (Venkatesh, Thong, and Xu 2012), *perceived usefulness* (α=.97) by five items (Davis 1989; Venkatesh, Thong, and Xu 2012), and the individual's tendency for *anthropomorphism* (α=.94) by six items (Kozak, Marsh, and Wegner 2006; Epley, Waytz, and Cacioppo 2007) (see Web Appendix A2.2).

3.3 Results

We examined *H1* by applying a one-way ANOVA with planned contrasts using the *chatbot identity* as the independent variable (*non-social* vs. *social* vs. *social delays)* and *usage intentions* as dependent variable. Results revealed a significant difference across groups, F(2, 396)=3.189, p=.042. As hypothesized, *usage intentions* were significantly lower for the *social delays* (M=4.07) compared to the *non-social* (M=4.60, p=.016) and marginally lower compared the *social* (M=4.49, p=.057). The difference between the *non-social* and the *social* was insignificant, p=.595. We proceeded with testing *H2* by conducting a simple mediation analysis using the "PROCESS" macro for SPSS (Model 4; Hayes 2018). We adopted the *chatbot identity* as the independent variable (0=*non-social*, 1=*social*, 2=*social delays*), *perceived usefulness* as mediator, and *usage intentions* as dependent variable. Results confirmed that *perceived usefulness* was significantly lower for the *social delays* compared to the *non-social* (b=-.531, p=.016) while there was no difference between the *non-social* and the *social* (b=-.101, p=.635). Considering the strong positive relation between *perceived usefulness* and *usage intentions* (b=.919, p<.001), the negative effect for the *social delays* on *usage intentions* was found to be fully mediated by a loss in *perceived usefulness* (b=-.488, *95%-CI*[-.879,-.085]; c'=-.040, *95%-CI*[-.207,.128]). For the *social*, there was no significant negative indirect effect (b=-.092, *95%-*

CI[-.455,.281]). To test *H3*, we adopted the individual's tendency for *anthropomorphism* as the moderator to the *a*-path (Model 7; Hayes 2018). We used the common +/-1*SD* criterion to categorize people into being *high* (+1*SD*, *M*=4.10), or *low* (-1*SD*, *M*=1.11) in *anthropomorphism*. Although the index of moderated mediation barely missed statistical significance (MMI_{NSvsSD}=.177, *95%-CI*[-.039,.390]), the negative effect for the *social delays* was about three times larger for people being *low* vs. *high* in *anthropomorphism* (b_{low}=-.778, *95%-CI*[-1.345,-.209]; b_{high}=-.249, *95%-CI*[-.646,.151]. As the effect for the *high anthropomorphism* group even became insignificant, results provided slight evidence for *H3*.

3.4 Discussion

Study 1 initially demonstrated that dynamic response delays reduce a service chatbot's perceived usefulness resulting in lower usage intentions. The effect was found to be stronger for people being low vs. high in anthropomorphism. These findings provided first evidence that the application of computer-like schemas makes customers expect chatbots to respond immediately and to provide an efficient service.

Results further revealed that social cues have no positive effects on usage intentions or perceived usefulness as there were no significant differences between the *non-social* and the *social* without response delays. Although this finding is somehow counterintuitive, it underlines the ambivalence of the effectiveness of social cues in chatbots discussed in literature (Blut et al. 2021). A potential explanation is that travel booking is a quite computer-like task making the utilitarian value more important than relational aspects and need for human contact (Sheehan, Jin, and Gottlieb 2020). Furthermore, the insignificant index of moderated mediation is to be discussed as it contradicted *H3*. An explanation could be that anthropomorphism and treating chatbots like social actors might be unconscious and automatically elicited processes (Nass and Moon, 2000; Kim and Sundar 2012). Asking people explicitly to which extent they believe a chatbot to be able to feel or to think might be inadequate as people are aware that

these believes are not rational (Reeves and Nass 1996; Nass and Moon 2000). This was also evident in the low mean for anthropomorphism in the sample (*M*=2.60). Study 2–4 accounted for this shortcoming by using other indices or manipulations for anthropomorphism or the application of human-like schemas.

4 Study 2: Chatbot Vs. Human Agent

Study 2 dived deeper into the role of computer- vs. human-like schemas in shaping expectations towards service agents as there is some ambivalence in Study 1. Specifically, it aimed to empirically test *H4* positing that the backfiring effect should not occur when users believe the agent to be a human (vs. a chatbot). Study 2 used a setting similar to that of Study 1, however, this time some of the participants were told that the service agent with dynamic response delays is not a chatbot but a human. This aimed at triggering human-like schemas externally enabling us to test whether the application of computer-like schemas is truly the prerequisite for the backfiring effect.

4.1 Stimuli

Study 2 adopted a one factor design with three levels: *social* vs. *social delays* vs. *human agent*. We used the *social* without delays as the baseline condition for two reasons: first, it did not differ from the *non-social* regarding *perceived usefulness* or *usage intentions* in Study 1. And second, we avoided potential confounding effects if all conditions show the same social cues and only the response delay is manipulated.

We used the video stimuli of the *social* and the *social delays* from Study 1 with only one minor adoption: we replaced the cartoon-like human avatar of the chatbot by a picture from a real human to make the situation more plausible for the people in the *human agent* condition (see Appendix). The video for the *human agent* condition was the same one we used for the *social delays*; the only difference is that participants were told to watch an interaction between a customer and a human service agent (vs. a chatbot).

4.2 Sample and Procedure

We recruited *n*=490 participants using a research panel. To be included in the analysis, participants did not only to have pass attention checks (*n*=434), but also had to answer a question asking if the agent in the video was (1) a chatbot vs. (2) a human. We did so since some participants could believe the agent to be a bot although they have been in the *human agent* condition or vice versa. The inclusion of a correct response requirement was intended to ensure the activation of the target schema (computer- vs. human-like); otherwise, results could have been confounded. Further *n*=64 had to be excluded for responding wrong resulting in a final sample of *n*=370 individuals (M_{age}=36.29, SD_{age}=12.23; 54.3% female).

The general procedure was the same as in Study 1, i.e., participants were randomly assigned to one of the three different conditions and watched the video before answering a standardized questionnaire. The measures for *usage intentions* (*α*=.97) and *perceived usefulness* (*α*=.97) were adopted from Study 1. In addition, we measured *expected usefulness* (*α*=.89) before participants watched the video. We did so for two reasons: first, the *expected usefulness* might differ for humans and chatbots. For instance, humans are believed to be more effective than bots, while bots are believed to offer quicker services (Crolic et al. 2022; Yu, Xiong, and Shen 2022). Second, measuring a priori expectations and posteriori experiences regarding the agent's usefulness enabled us to calculate an index for *usefulness expectancy violations*. Lastly, we measured participants' *familiarity* with the use of (1) chatbots or (2) digital consultancies by human agents using a three item seven-point scale (self-developed, *α*=.91) to examine if there are significant differences that could have confounding effects (see Web Appendix B).

4.3 Results

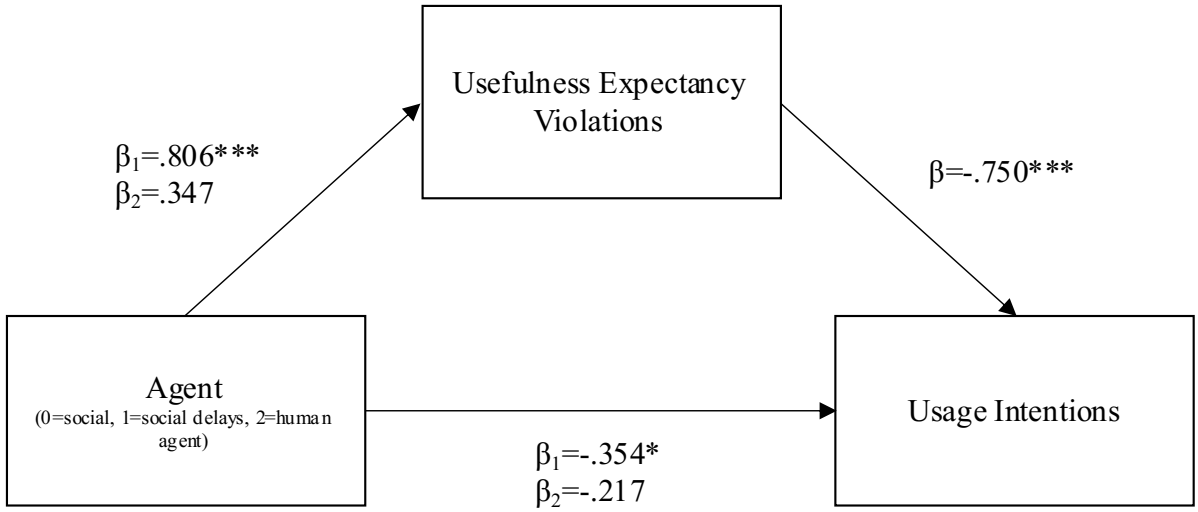
We first checked if participants significantly differ regarding (1) their *familiarity* with using chatbots vs. digital consultancies and (2) *expected usefulness*. Two one-way ANOVAs showed that there were no significant differences across conditions (all *p*s>.18). We continued

with the main effect applying a one-way ANOVA with planned contrasts to examine if there are significant differences across conditions regarding *usage intentions*. Results revealed that all conditions significantly differ from another, $F(2, 367)=8.687$, $p<.001$. Precisely, *usage intentions* were the highest for the *social* ($M=5.05$) followed by the *human agent* ($M=4.58$) and the *social delays* ($M=4.09$). We continued with calculating the index of *expectancy violations* regarding the agent's usefulness by subtracting *perceived usefulness* from *expected usefulness*. We proceeded with testing *H4* by using a simple mediation analysis in "PROCESS" (Model 4; Hayes 2018). We adopted the three conditions as the independent variable (0=*social*, 1=*social delays*, 2=*human agent*), *usefulness expectancy violations* as mediator, and *usage intentions* as dependent variable. Results showed that expectancies were only violated significantly for the *social delays* ($b=.806$, $p<.001$) but not the *human agent* ($b=.347$, $p=.103$) when comparing to the *social*. Results further showed a strong negative relation between *usefulness expectancy violations* and *usage intentions* ($b=-.750$, $p<.001$). We could thus accept *H4* as there was only a negative indirect effect for the *social delays* on *usage intentions* partially mediated by *usefulness expectancy violations* ($b=-.604$, *95%-CI*[-.941,-.278]; $c'=-.354$, *95%-CI*[-.685,-.023]) that was not evident for the *human agent* ($b=-.261$, *95%-CI*[-.589,.050]) (see Figure 2).

4.4 Discussion

Study 2 showed that dynamic response delays only violate usefulness expectations when participants believed the agent to be a chatbot and not a human, although the video stimulus was identical. This substantiates the argument that the application of computer- vs. human-like schemas facilitates the expectation of receiving immediate responses. As dynamic response delays are usual in interpersonal chats, there were no expectancy violations regarding the human agent's efficiency.

However, two things are to be discussed: first, usage intentions were lower for the human agent compared to the social without delays. An explanation could be the computer-like



Indirect effect *social* vs. *social delays*: β=-.604, *95%-CI*[-.941,-.278]
Indirect effect *social* vs. *human agent*: β=-.261, *95%-CI*[-.589,.050]

Figure 2. Results from Study 2.

nature of travel booking making the consultation of a human agent seem inappropriate. Nevertheless, usage intentions for the human agent were still higher than for the social delays and results revealed that the slight negative effect for the human agent could not be explained by usefulness expectancy violations. Second, the expected usefulness was not higher for the chatbot than for the human agent. This might be explained by higher efficacy attributions towards human agents that can also drive the expected usefulness of an agent (Crolic et al. 2022; Yu, Xiong, and Shen 2022).

5 Study 3: Examining Implicit Indices for Anthropomorphism in Real Chatbot Interactions

Study 3 addressed a major shortcoming of Study 1 and 2. Instead of watching videos, participants interacted with travel booking chatbots embedded on a website. Both the chatbots and the website were programmed for this research project. Besides enhancing the situation's realism and increasing external validity, the use of interactive chatbots enabled us to screen the conversation scripts between the participants and chatbots to seek for implicit indices for

anthropomorphism. Anthropomorphism and social responses towards computers are not necessarily mindful but can be mindless phenomena that manifest in automatically elicited social behavior (Nass and Moon 2000; Kim and Sundar 2012; Crolic et al. 2022). As humans are usually aware of the lifelessness of chatbots, using implicit indices might be more accurate to capture an individual's tendency to apply human-like schemas in the interaction.

5.1 Stimuli

Study 3 applied the one factor design from Study 1 (*non-social* vs. *social* vs. *social delays*). The chatbots were programmed using the online tool "SnatchBot" (SnatchBot 2023) and followed a pre-scripted conversation flow that was adaptive to user input. The social cues and the length of the response delays were adopted from Study 1. The fictitious travel booking website ("TravelVista") was employed in "Visual Studio Code" (Microsoft 2023) and showed some generic information on travelling (see Appendix).

5.2 Sample and Procedure

We recruited n=345 individuals on "Prolific" and "SurveyCircle". After excluding participants who failed attention checks or who experienced major technical issues in the interaction (n=30), the final sample included n=315 individuals (M_{age}=31.28, SD_{age}=10.71, 57.8% female). Participants were given the task to seek for a hotel for a city trip using the chatbot (see Web Appendix C1). Some general information on their requirements were given (e.g., the date and the budget) before participants were redirected to the website and chatted with the bot. After having collected all necessary information, the chatbot provided four hotel recommendations matching the participant's requirements. Participants had to select their preferred hotel before entering their individual ID they received at the beginning of the survey (see Web Appendix C2). This ID enabled us to link the conversation script with the survey data. Lastly, participants returned to the survey and filled out a standardized questionnaire including the measures for *usage intentions* (α=.95), *perceived usefulness* (α=.97), and

anthropomorphism (α=.84) adopted from Study 1. We also measured *service outcome satisfaction* by three items (self-developed, α=.89, see Web Appendix C3) as not only the process of service delivery, but also the outcome (i.e., the hotel recommendation) might be a strong determinant for the evaluation of the service (Dabholkar and Overby 2004). For the implicit index of *anthropomorphism*, we screened the conversation scripts for indicators whether an individual anthropomorphized the chatbot or not. Following previous research (Crolic et al. 2022), we coded if a participant adopted norms from interpersonal communication in the interaction, e.g., calling the chatbot by its name, following politeness norms by writing "please" and "thanks", or greeting and farewelling. We coded participants with "0" if there were no indicators or with "1" when at least two messages showed indicators for anthropomorphism.

5.3 Results

We first tested if *service outcome satisfaction* is equal across conditions and positively related to *usage intentions* and *perceived usefulness*. A one-way ANOVA showed no difference across conditions (p>.91) while correlation analysis revealed significant positive relations with *usage intentions* (r=.347, p<.001) and *perceived usefulness* (r=.419, p<.001). We hence included *service outcome satisfaction* as covariate to our further analyses.

A one-way ANCOVA on *usage intentions* showed a significant difference across conditions, $F(2, 311)=21.847$, p<.001.[4] Supporting *H1*, planned contrasts revealed that *usage intentions* were significantly lower for the *social delays* (M=3.43) compared to the *non-social* (M=4.82, p<.001) and the *social* (M=4.42, p<.001). The *non-social* and the *social* did not differ significantly, p=.093. We proceeded with a simple mediation analysis (Model 4; Hayes 2018) adopting the *chatbot identity* as the independent variable (0=*non-social*, 1=*social*, 2=*social delays*), *perceived usefulness* as mediator, *usage intentions* as dependent variable, and *service outcome satisfaction* as covariate. Results revealed that *perceived usefulness* is significantly

[4] Results without covariate (ANOVA): $F(2, 312)=18.102$, p<.001

lower for the *social delays* (b=-1.312, p<.001) but not the *social* (b=-.357, p=.104) when comparing to the *non-social*. As there was a strong relation between *perceived usefulness* and *usage intentions* (b=.764, p<.001), results further showed that the negative effect for the *social delays* was partially mediated by a loss in *perceived usefulness* (b=-1.003, *95%-CI*[-1.361,-.659]; c'=-.360, *95%-CI*[-.636,-.084]) thus supporting *H2*. No such effect was evident for the *social* (b=-.273, *95%-CI*[-.571,.028]). Next, we included the implicit index for *anthropomorphism* (0=*low* vs. 1=*high*) as a moderator to the model (Model 7; Hayes 2018). The index of moderated mediation became significant (MMI_{NSvsSD}=1.453, *95%-CI*[.373,2.638]) as the negative indirect effect was only evident for people who were *low* in *anthropomorphism* (b=-1.341, *95%-CI*[-1.725,-.961]) but not for people who were *high* in *anthropomorphism* (b=.112, *95%-CI*[-.904,1.227]). Results hence lent credence for *H3*.

To rule out alternative explanations and confounding effects, we had to verify if our coding is adequate to capture an individual's tendency to anthropomorphize. Alternatively, those who engaged in full-sentence interactions following interpersonal communication norms might have taken more time, possibly because they enjoyed participating or had no time pressure, leading to less frustration with long response delays. Also, these respondents could have systematically differed regarding demographics or their general response behavior. To rule out these alternative explanations, we examined if the time to fill out the questionnaire, the age, or *service outcome satisfaction* differed between the *low* (n=242) and the *high anthropomorphism* group (n=65)[5]. Results from three independent t-tests showed that there were no significant differences (all ps>.26). However, another t-test on the *anthropomorphism* scale revealed that the *low anthropomorphism* group had a lower score (M=1.75) than the *high anthropomorphism* group (M=2.13), t(77.778)=-2.185, p=.032.[6]

[5] The remaining n=8 participants did not indicate their individual ID at the end of the conversation resulting in missing values for mindless anthropomorphism.
[6] Results without degree of freedom correction: t(305)=-2.838, p=.005

5.4 Discussion

Study 3 replicated the findings from Study 1 in a more realistic service situation. Moreover, Study 3 enabled us to use an implicit index for anthropomorphism (vs. explicit in Study 1). Scientific literature argues that anthropomorphism and social responses towards computers are frequently unconscious, automatically elicited processes (Nass and Moon 2000; Kim and Sundar 2012). Using an implicit instead of an explicit index could hence be a more robust way to capture mindless anthropomorphism. Furthermore, all effect sizes in Study 3 were larger than in Study 1 indicating the participants' higher involvement. Methodologically, this finding emphasizes that using screenshots or video vignettes could potentially underestimate effect sizes. Lastly, we were able to control for participants' service outcome satisfaction that is also an important determinant for the evaluation of a service (Dabholkar and Overby 2004). Results revealed that dynamic response delays might not affect satisfaction with the service outcome but only the service process.

6 Study 4: Computer-Like Vs. Human-Like Task

Study 4 considered a managerial relevant boundary condition by manipulating the service task the chatbot was used for comparing a computer-like task (train ticket booking) and a human-like task (medical assessment). Hence, it tested *H5* hypothesizing the backfiring effect is moderated by the service task's human-likeness. While train ticket booking is an exchange task which can be accomplished independently, receiving a medical assessment is a communal task usually requiring human guidance, social connectedness, and a high time effort (Huang and Rust 2018; Sheehan, Jin, and Gottlieb 2020; Seeger, Pfeiffer, and Heinzl 2021). The backfiring effect might hence be attenuated when the chatbot performs a human-like (vs. computer-like) task.

6.1 Stimuli and Pre-Test

Study 4 applied a 2 (*computer-like task* vs. *human-like task*) x 3 (*non-social* vs. *social* vs. *social delays*) design. Like in Study 3, the chatbots were programmed using "SnatchBot" (SnatchBot 2023) and followed pre-scripted conversation flows being adaptive to user input. The chatbots were embedded either on a fictitious train ticket booking website ("TravelTrain") or a healthcare website ("DigiHealth") programmed in "Visual Studio Code" (Microsoft 2023). The visual appearance of the chatbots and the websites have been almost identical for both services to avoid confounding effects (see Appendix). Also, we ensured the length of the messages the chatbots sent were similar in both services to have comparable dynamic response delays. We delayed responses by 133ms per character, which represents a typing speed of approx. 90 words per minute that very skilled human agents might be capable of (Zhai, Hunter, and Smith 2002; Arif and Stuerzlinger 2009).

We conducted a pre-test for the selection of a computer- vs. human-like service by comparing *travel booking* with *train ticket booking* and *medical assessment*. Although *travel booking* already is a quite computer-like task, *train ticket booking* could be even more computer-like due to its lower complexity and involvement (Sheehan, Jin, and Gottlieb 2020). On the other hand, *medical assessments* are considered one of the most human-like tasks as it is highly complex and characterized by a stronger need for social relatedness (Seeger, Pfeiffer, and Heinzl 2021). To test whether the services significantly differ in their computer- vs. human-likeness, we designed three different scenarios in which situations of (1) *travel booking*, (2) *train ticket booking*, and (3) a *medical assessment* was described (see Web Appendix D1). We randomly assigned participants to one of the three scenarios before they filled out a standardized questionnaire measuring the service task's *complexity* (r=.800) and *human-likeness* (r=.639) by two items each (self-developed) (see Web Appendix D2). A total of n=225 valid respondents participated in the study (M_{age}=35.09, SD_{age}=12.52, 51.6% female). Two one-way ANOVAs

revealed that the service task's *complexity* ($F(2, 222)=55.479$, $p<.001$) and *human-likeness* ($F(2, 222)=43.070$, $p<.001$) significantly differed. Planned contrasts showed that receiving a *medical assessment* (vs. *travel booking* and *train ticket booking*) had a higher *complexity* ($M_{medical}=4.61$; $M_{travel}=3.01$; $M_{train}=2.32$, $ps<.001$) and was more *human-like* ($M_{medical}=5.57$; $M_{travel}=4.04$; $M_{train}=3.49$, $ps<.001$). Moreover, *train ticket booking* was rated less *complex* ($p=.002$) and less *human-like* ($p=.021$) than *travel booking*. We hence decided for using *train ticket booking* vs. *medical assessment*.

6.2 Sample and Procedure

We calculated the required sample size using G*Power 3.1 before recruiting participants by means of convenience sampling. Using the same parameters like in Study 1 and considering the 2x3-design (six groups), the required minimum sample was $n=402$. A total of $n=447$ individuals participated, $n=30$ of which being excluded for attention check failures or experiencing major technical issues in the chatbot interaction. Hence, the final sample included $n=417$ individuals ($M_{age}=27.92$, $SD_{age}=9.40$, 64.5% female).

Participants were randomly assigned to one of the two service scenarios (*train ticket booking* vs. *medical assessment*). In the *train ticket booking* scenario, participants were given the task to purchase train tickets for an intercity trip using the chatbot. In the *medical assessment* scenario, participants had to put themselves into the position of a person suffering from a chest pain before engaging with the chatbot for an initial *medical assessment* (see Web Appendix D3). After having read the scenario, participants were redirected to the fictitious websites and started the interaction with the chatbot. The chatbot collected all necessary data before either (1) recommending three train ticket options or (2) providing an initial medical assessment by showing three different clinical pictures matching the symptoms (see Web Appendix D4). Afterwards, participants returned to the survey and filled out a standardized questionnaire on *usage intentions* ($\alpha=.94$), *perceived usefulness* ($\alpha=.95$), and *service outcome satisfaction*

(α=.92) adopted from the previous studies. Since we have changed the chatbots' appearance, the dialogues, and the service environment, we included the manipulation checks used in the pre-test of Study 1, i.e., we measured the chatbot's *perceived human-likeness* (α=.89), the interaction's *perceived duration* (α=.91), and *perceived realism* (α=.89).

6.3 Results

We started with the manipulation checks by applying two one-way ANOVAs adopting the *chatbot identity* (*non-social* vs. *social* vs. *social delays*) as the independent variable and (1) *perceived human-likeness* and (2) *perceived duration* as dependent variables. Results confirmed a successful manipulation in both service scenarios as *perceived human-likeness* was higher for the *social* (M_{train}=3.32; $M_{medical}$=3.29) and the *social delays* (M_{train}=3.14; $M_{medical}$=3.33) compared to the *non-social* (M_{train}=2.06; $M_{medical}$=2.15), Fs>16.99, ps<.001. Also, *perceived duration* was higher for the *social delays* (M_{train}=5.63; $M_{medical}$=4.50) compared to the *social* (M_{train}=2.51; $M_{medical}$=2.36) and the *non-social* (M_{train}=2.92; $M_{medical}$=2.51) in both service scenarios, Fs>70.53, ps<.001. An independent t-test on *perceived realism* further confirmed that both service scenarios were perceived equally realistic (M_{train}=5.65; $M_{medical}$=5.41), t(415)=1.759, p=.079.

Next, we applied a two-way ANCOVA including the *chatbot identity* and *service task* as independent variables and *usage intentions* as dependent variable to examine the main and interaction effects. Like in Study 3, we controlled for *service outcome satisfaction*. Results indicated significant main effects for both the *chatbot identity*, F(2, 410)=10.530, p<.001[7] and *service task*, F(1, 410)=49.488, p<.001[8]. Additionally, there was a significant interaction effect F(2, 410)=9.463, p<.001.[9] Two one-way ANCOVAs revealed that *usage intentions* for the *social delays* were only lower in the *train ticket booking* scenario (M_{NS}=4.55; M_{SO}=4.60;

[7] Results without covariate (ANOVA): F(2, 411)=12.478, p<.001
[8] Results without covariate (ANOVA): F(1, 411)=5.517, p=.019
[9] Results without covariate (ANOVA): F(2, 411)=5.117, p=.006

M_{SD}=3.13, F(2, 204)=15.553, p<.001[10]) but not in the *medical assessment* scenario (M_{NS}=4.61; M_{SO}=4.58; M_{SD}=4.27, F(2, 205)=.404, p=.668[11]). We proceeded with a simple mediation analysis (Model 4; Hayes 2018) and adopted the *chatbot identity* as the independent variable (0=*non-social*, 1=*social*, 2=*social delays*), *perceived usefulness* as mediator, *usage intentions* as dependent variable, and *service outcome satisfaction* as covariate. Results indicated a significant negative total effect for the *social delays* compared to the *non-social* (c=-.795, p<.001) that was not evident for the *social* (c=.002, p=.993). This effect was found to be fully mediated by a loss in *perceived usefulness* (b=-.742, *95%-CI*[-1.068,-.417]; c'=-.054, *95%-CI*[-.288,.181]). Next, we conducted a moderated mediation analysis (Model 7; Hayes 2018) including *service task* (0=*train ticket booking*, 1=*medical assessment*) as a moderator to the *a*-path. The index of moderated mediation showed a significant interaction effect (MMI_{NSvsSD}=.977, *95%-CI*[.445,1.516]). The negative indirect effect was only evident in the *train ticket booking* scenario (b=-1.173, *95%-CI*[-1.635,-.725]) but not the *medical assessment* scenario (b=-.196, *95%-CI*[-.491,.099]) (see Figure 3). Hence, results supported *H5*.

6.4 Discussion

Study 4 found empirical evidence for *H5*, i.e., the backfiring effect of dynamic response delays on usage intentions mediated by a loss in perceived usefulness was only evident in a computer- but not a human-like service task. The results lent credence for our assumption that a human-like service task characterized by a high complexity and need for human guidance could be more likely to elicit the application of human-like schemas and different expectations. In contrast, chatbots performing a computer-like task were expected to maximize efficiency. Study 4 hence showcased that a service task's computer- vs. human-likeness might guide our perception and evaluation of social cues in service chatbots.

[10] Results without covariate (ANOVA): F(2, 205)=14.231, p<.001
[11] Results without covariate (ANOVA): F(2, 206)=.996, p=.371

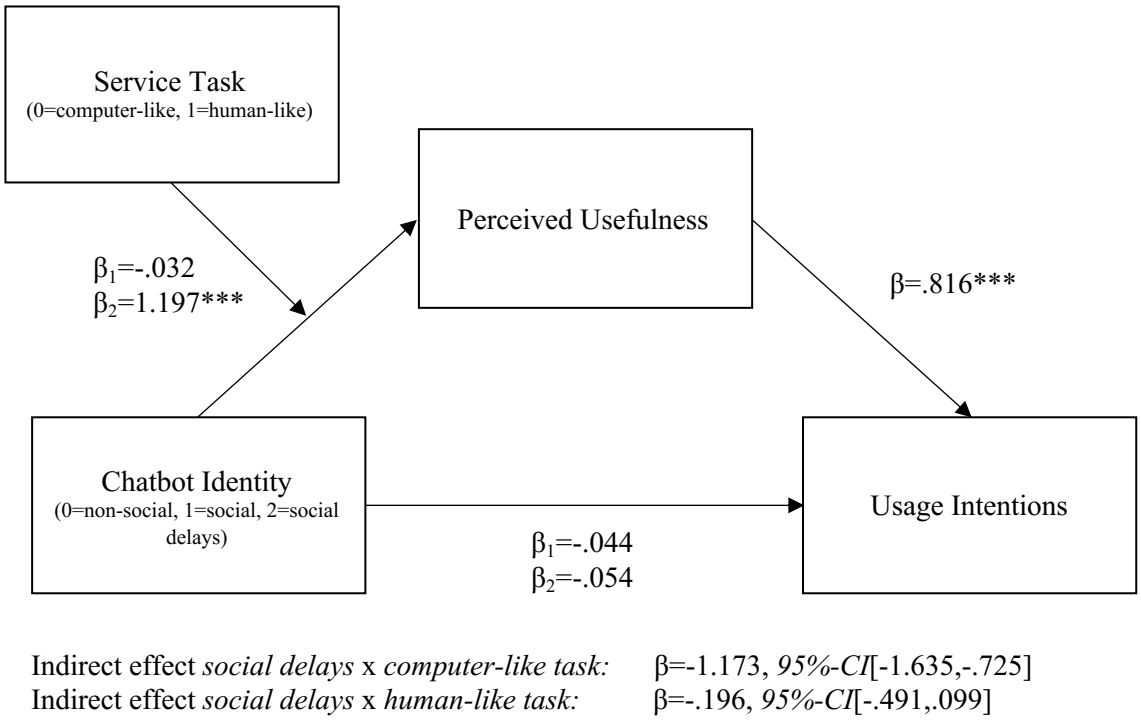


Figure 3. Results from Study 4.

7 Study 5: Usefulness Expectancy Violations and Service Provider Evaluation

The purpose of Study 5 was to examine if expectancy violations regarding the chatbot's perceived usefulness induced by dynamic response delays have adverse effects on the service provider evaluation. As discussed in the conceptual background, humans might enter a conversation with a service chatbot with a priori expectations regarding its usefulness and a negative disconfirmation could result in a negative evaluation of the service provider (Churchill and Surprenant 1982; Crolic et al. 2022). Study 5 extended previous studies by going beyond usage intentions as outcome variable and considering service provider evaluation, testing two additional hypotheses.

H6: Dynamic response delays (vs. no such delays) in a service chatbot diminish service provider evaluation.

H7: The negative effect of dynamic response delays on service provider evaluation is mediated by usefulness expectancy violations.

7.1 Stimuli

Study 5 used the same stimuli as Study 3 (i.e., the travel booking scenario) hence applying a one-factor design with three levels (*non-social* vs. *social* vs. *social delays*).

7.2 Sample and Procedure

We recruited n=354 individuals on "SurveyCircle" and the internal university recruiting system. After excluding participants failing attention checks or experiencing major technical issues (n=51), the final sample included n=303 individuals (M_{age}=26.70, SD_{age}=7.21, 71.2% female). Participants performed the same task like in Study 3 (i.e., travel booking; see Study 3). However, we measured *expected usefulness* (α=.91) before the interaction to have a baseline value for the calculation of *usefulness expectancy violations* (see Study 2). The post-interaction questionnaire captured *service provider evaluation* by four items (α=.95) adopted from Crolic et al. 2022 (see Web Appendix E) as well as *perceived usefulness* (α=.96) and *service outcome satisfaction* (α=.90) using the measures from previous studies.

7.3 Results

We started with examining whether *service outcome satisfaction* is equal across conditions by applying a one-way ANOVA. Results showed that there were no significant differences (p=.157). We proceeded with testing the main effect by applying a one-way ANCOVA adopting the *chatbot identity* (*non-social* vs. *social* vs. *social delays*) as independent variable, *service provider evaluation* as dependent variable, and *service outcome satisfaction* as covariate. Results indicated a significant difference across conditions, $F(2, 299)=34.686$, $p<.001$.[12] Planned contrasts revealed that *service provider evaluation* is lower in the *social delays* condition (M=4.54) compared to the *social* (M=5.70, $p<.001$) and the *non-social* (M=5.79, $p<.001$). We could thus accept *H6*. Next, we calculated the index of *usefulness*

[12] Results without covariate (ANOVA): $F(2, 300)=29.609$, $p<.001$

expectancy violations by subtracting *perceived usefulness* from *expected usefulness*. We tested *H7* by conducting a simple mediation analysis (Model 4; Hayes 2018) adopting the *chatbot identity* as the independent variable (0=*non-social*, 1=*social*, 2=*social delays*), *usefulness expectancy violations* as mediator, *service provider evaluation* as dependent variable, and *service outcome satisfaction* as covariate. Results indicated that expectancies were only violated for the *social delays* (b=1.638, p<.001) but not the *social* (b=.040, p=.865) when comparing to the *non-social*. Results also showed that *usefulness expectancy violations* were negatively related to *service provider evaluation* (b=-.339, p<.001). Consequently, the negative indirect effect for the *social delays* on *service provider evaluation* was partially mediated by *usefulness expectancy violations* (b=-.555, *95%-CI*[-.784,-.361]; c'=-.575, *95%-CI*[-.868,-.283]). The results thus provided support for *H7*.

7.4 Discussion

Study 5 demonstrated that a violation of usefulness expectancies induced by dynamic response delays negatively affects service provider evaluation as customers primarily expect utilitarian advantages when using service chatbots. The study extended previous research by combining dimensions from technology acceptance models with approaches from customer satisfaction and service research. It hence showed that the service provider evaluation does not only depend on the chatbot's outcome, but also on the extent to which a priori expectations towards the process of service delivery have been met.

However, it is important to note that the negative effect on service provider evaluation was only partially mediated by usefulness expectancy violations. The remaining negative effect could potentially be explained by the participants' belief that the chatbot was not working properly as argued in previous research (Schanke, Burtch, and Ray 2021). A company using a service chatbot that seems not sophisticated yet might hence be evaluated worse.

8 General Discussion

Service chatbots are frequently imbued with social cues to make conversations feel more natural and human-like. Although humanization can have positive effects, literature is still inconclusive in understanding under which circumstances humanization has favorable consequences (Mende et al. 2019; Blut et al. 2021; Holthöwer and van Doorn 2022; Uysal, Alavi, and Bezençon 2022; Han, Deng, and Fan 2023). In the present research, we examined if the social cue of dynamic response delays has negative downstream consequences on the evaluation of the chatbot and the service provider as it may interfere with one of the major advantages of chatbots, i.e., making service provision more efficient thus being a useful tool.

In a series of five studies, we found robust evidence that dynamic response delays attenuate usage intentions and service provider evaluation. Findings revealed that the underlying mechanism stems from a violation of usefulness expectations that emanates from the application of computer-like schemas to interactions with service chatbots. Congruently, we found that this backfiring effect is attenuated by the application of human-like (vs. computer-like) schemas in the interaction: the effect disappeared when customers tended to anthropomorphize chatbots (Study 1 and 3), when the service agent was believed to be a human (Study 2), or the service task was human-like (vs. computer-like) (Study 4). This paper therefore shows that social cues can backfire when contradicting the expectation of receiving efficient services that is one of the key benefits of using chatbots. In this regard, this research takes a step toward a better understanding on limitations of humanizing service chatbots and the critical role of schemas in the perception of social cues. Several theoretical contributions, managerial implications, and future research avenues are to be discussed.

8.1 Theoretical Contributions

Starting with high-level theoretical contributions, this research links theories and approaches on social chatbots and anthropomorphism, technology acceptance models, and

service research. Specifically, it demonstrated that social cues reducing a service chatbot's efficiency can result in expectancy violations regarding its perceived usefulness that can be a major extrinsic motivator for adopting new technology (Venkatesh, Thong, and Xu 2012). These expectancy violations culminated in a reduced intention to use the agent and a less favorable service provider evaluation. Therefore, this paper emphasizes the pivotal role of aligning a service chatbot's (social) design elements with core dimensions of technology acceptance as chatbots are still software tools.

The second major theoretical contribution refers to the paper's critical perspective on the "Social Response Theory" (Nass and Moon 2000). "Social Response Theory" argues that humans tend to mindlessly adopt interpersonal heuristics in their interactions with computers and to treat them like social actors. However, this paper shows that social cues are perceived and evaluated differently when shown by a chatbot vs. a human agent. In this regard, the general applicability of the "Social Response Theory" in computer and chatbot interactions should be discussed. "Social Response Theory" originates in the late 1990s when computers were scarcely widespread, and people had limited knowledge about them. This might have facilitated mindless social responses and anthropomorphism that is an intuitive strategy in dealing with unknown and complex agents (Epley, Waytz, and Cacioppo 2007). However, as computers and chatbots have become increasingly prevalent, users have gained more experience and schemas have become richer and more accurate (Rouse and Morris 1986). Consequently, contemporary users are more inclined to apply computer-like schemas in their interactions with chatbots, which can trigger stereotypical associations (Meng and Dai 2021). In this regard, it should also be noted that non-embodied chatbots, when compared to physically existing robots, are generally less human-like, making the application of computer-like schemas even more likely (Blut et al. 2021; Pitardi et al. 2022). These schemas elicit computer-like expectations regarding

the chatbot and its communication behavior (e.g., expecting an immediate response) which in turn may result in expectancy violations and a negative evaluation in case of a mismatch.

Third, this research shows that not all social cues are equally promising to humanize service chatbots. Precisely, the studies have demonstrated that social cues backfire when they disrupt the expected conversation flow. In line with EVT, the resulting expectancy violations might activate cognitive processes and attention shifts toward the disrupting cue (i.e., the response delay). This may trigger sensemaking processes motivating users to elaborate on the discrepancy between the chatbot's expected and the actual behavior, i.e., the slow response time becomes more salient, and users become aware of the technical nature of chatbots (Burgoon 1993; Grimes, Schuetzler, and Giboney 2021). Social cues in a service chatbot could thus be more likely to have positive consequences when aligning with the expected conversation flow and when fulfilling social needs casually. Two points substantiate this argumentation: first, humans' tendency to respond socially and positive to humanized entities is commonly an unconscious process (Nass and Moon 2000; Epley, Waytz, and Cacioppo 2007). The potential positive impact of social cues may manifest on a subconscious level, as people are usually aware of the lifelessness of computers and bots (Nass and Moon 2000). And second, previous research has found that when faced with unsatisfactory services, customers are more inclined to shift their attention towards exchange norms, with decreasing importance of relational aspects (Li, Chan, and Kim 2019). It is further to note that we did not find evidence for any positive effects for social cues on neither the evaluation of the chatbot nor the service provider. Although a recent meta-analysis found an overall positive effect of perceiving a sense of humanness in bots, there is still ambivalence on the drivers and moderators of positive vs. negative outcomes (Blut et al. 2021). The present research suggests that customers might enter service chatbot conversations with computer-like schemas and corresponding expectations as they are still software tools. Prioritizing a chatbot's performance and its provision of utilitarian

value might be more important than humanization. This is supported empirically by the high predictive power of perceived usefulness in all studies and by findings from previous research (Blut, Wang, and Schoefer 2016; Lee and Lyu 2016).

Fourth, this paper significantly contributes to the role of schemas in the perception and evaluation of social cues in chatbots. We found that both individual traits (i.e., an individual's tendency to anthropomorphize chatbots) and contextual factors (i.e., a service task's human-likeness) impact the application of computer- vs. human-like schemas in the interaction. These schemas can shape expectations towards the interaction and guide customers in their perception and evaluation of social cues. Specifically, when customers enter an interaction with human-like schemas, they might be inclined to expect the chatbot to behave like a human agent. In this regard, it is worth mentioning that most of the existing studies consider anthropomorphism a mediator between social cues and relational, functional, or behavioral outcomes (Blut et al. 2021). However, this paper shows that anthropomorphism can also serve as moderator in the perception of social cues in chatbots.

8.2 Managerial Implications

Our findings imply that managers should conduct a thorough assessment of the potential benefits and drawbacks that may emanate from the implementation of a specific social cue to service chatbots. In the present paper, we found that a social cue making a chatbot less efficient (i.e., dynamic response delays) decreases its perceived utilitarian value diminishing the evaluation of the chatbot and the service provider. This might particularly hold true for merely outcome-oriented, computer-like services in which receiving a fast and convenient service is more important than the fulfillment of social needs (Huang and Rust 2018; Sheehan, Jin, and Gottlieb 2020). We thus suggest managers to anticipate their target groups' expectations towards the service and align the chatbot's design accordingly.

Second, our studies did not find any positive effects of social cues on chatbot or service provider evaluation. The benefits of making service chatbots more human-like by simple visual or verbal design elements (e.g., an avatar and a name) might thus be quite limited for enhancing service experience. Focusing resources on enhancing a chatbot's performance could be more beneficial for companies in facilitating the acceptance of chatbots and enhancing customer satisfaction. Results across studies support this as perceived usefulness and service outcome satisfaction were found to strongly predict usage intentions and service provider evaluation. Furthermore, we encourage managers and software designers to abstain from the general attempt of making service chatbots as human-like as possible. Instead, they should ensure to retain the major benefits of chatbots, i.e., to make services faster which is a key advantage of chatbots vs. human agents. Results from Study 2 showed that efficient bots may be even preferred over human agents, particularly for computer-like tasks that can be handled independently and do not require social relations.

Third, we did not find evidence that dynamic response delays have any positive impact on the chatbot's evaluation. Although their purpose is to perfectly mimic an interhuman conversation, our pre-tests and studies did not reveal a significant difference regarding perceived human-likeness between the social without delays and the social with dynamic delays. Hence, dynamic response delays do not add much in enhancing a chatbot's perceived human-likeness. Reflecting this finding against customers' complaints about Lufthansa's chatbot for responding too fast (Crozier 2017), we provide two explanations: first, customers' schemas about chatbots could have become more sophisticated with the increasing pervasiveness of chatbots over the last years, i.e., they might have learned that chatbots can respond immediately. And second, the effects of response delays might follow an inverted U-shape with the length of the response delay on the x-axis and its impact on the chatbot's evaluation on the y-axis. In other words, while too short response times may overwhelm

customers, too long response delays may lead to frustration. Hence, there might be a sweet spot, e.g., delaying a message by only a few seconds (Moon 1999). Lastly, innovative software designers might consider alternative methods for implementing response delays to service chatbots. For example, users of "ChatGPT" can read along while the chatbot generates a response. This might make the conversation feel more natural and human-like without risking the backfiring effect of too long response times.

8.3 Limitations and Future Research

Like any empirical research, this paper has some limitations that provide avenues for future research. Starting with methodological limitations, our studies were survey-based limiting external validity. However, we addressed this issue in two ways: first, we employed websites, interactive chatbots, and provided the participants with realistic service tasks to maximize realism. Second, to enhance robustness in approaching the underlying mechanisms, we used different indices and manipulations for the application of computer- vs. human-like schemas on individual, agent-related, and contextual levels (i.e., explicit and implicit indices for anthropomorphism, comparing a chatbot with a human agent, and manipulating the service task's human-likeness). Future studies could replicate and extend our research by (1) manipulating anthropomorphism by encouraging individuals to imagine the chatbot has come alive (vs. is a computer) (Aggarwal and McGill 2012), (2) considering other relevant outcome dimensions (e.g., customer's willingness to pay for a product or service), and (3) conducting field studies using dropout rates as indicators for customer satisfaction.

We also encourage researchers to further elaborate on chances and risks of implementing social cues to chatbots more granularly. Many of the existing research considers the implementation of social cues binary by comparing a non-anthropomorphic with an anthropomorphic chatbot. However, our studies suggest that the perception and evaluation of social cues can depend on (1) cue-related, (2) agent-related, (3) customer-related, and (4)

context-related factors. Applied to the findings from this research, we have shown that (1) a social cue reducing the efficiency of (2) a service chatbot can backfire for (3) people who have a low tendency to anthropomorphize, or (4) in computer-like services. Future research could examine the interplay of other specific social cues (e.g., emotional support) with agent-related (e.g., non-embodied agent vs. embodied agent), customer-related (e.g., need for social belongingness), or context-related (e.g., information-seeking vs. complaining) factors. Scholars interested in delving deeper into studying inefficient social cues might examine other cues (e.g., messages that add unnecessary filler content), or extent our perspective by considering different types of bots (e.g., physical bots pretending to need some time to think about a response). In addition, although Study 4 manipulated the service task's human-likeness, consulting a chatbot to receive a medical assessment is still an outcome-oriented service. Inefficient social cues might have a different impact in conversations with chatbots in which utilitarian value has a secondary role, e.g., when the chatbot serves as companion like "Replika" (Pentina, Hancock, and Tianling 2023).

Lastly, we encourage researchers to conduct a systematic literature review or a meta-analysis on negative effects of social cues in bots and AI. With increasing empirical evidence for backfiring effects and boundary conditions, a comprehensive overview and systematization might help both theorists and practitioners to better understand which kind of human-likeness has negative consequences for whom, under which circumstances, and for which reasons. As bots and AI will significantly change the landscape of service provision, it is important to not only consider the bright sights of humanization, but also to anticipate and understand negative consequences and limitations.

References

Aggarwal, Pankaj and Ann L. McGill (2007), "Is That Car Smiling at Me? Schema Congruity as a Basis for Evaluating Anthropomorphized Products," *Journal of Consumer Research*, 34 (4), 468-79.

Aggarwal, Pankaj and Ann L. McGill (2012), "When Brands Seem Human, Do Humans Act Like Brands? Automatic Behavioral Priming Effects of Brand Anthropomorphism," *Journal of Consumer Research*, 39 (2), 307-23.

Arif, Ahmed S. and Wolfgang Stuerzlinger (2009), "Analysis of Text Entry Performance Metrics,", in 2009 IEEE Toronto International Conference Science and Technology for Humanity (TIC-STH), (September), 100-05.

Blut, Markus, Cheng Wang and Klaus Schoefer (2016), "Factors Influencing the Acceptance of Self-Service Technologies: A Meta-Analysis," *Journal of Service Research*, 19 (4), 396-416.

Blut, Markus, Cheng Wang, Nancy V. Wünderlich and Christian Brock (2021), "Understanding Anthropomorphism in Service Provision: A Meta-analysis of Physical Robots, Chatbots, and other AI," *Journal of the Academy of Marketing Science*, 49 (4), 632-58.

Brown, Susan A., Viswanath Venkatesh and Sandeep Goyal (2014), "Expectation Confirmation in Information Systems Research: A Test of Six Competing Models," *MIS Quarterly*, 38 (3), 729-56.

Burgoon, Judee K. (1993), "Interpersonal Expectations, Expectancy Violations, and Emotional Communication," *Journal of Language and Social Psychology*, 12 (1-2), 30-48.

Castelo, Noah, Johannes Boegershausen, Christian Hildebrand and Alexander P. Henkel (2023), "Understanding and Improving Consumer Reactions to Service Bots," *Journal of Consumer Research*, ucad023.

Cha, Young-Jae, Sojung Baek, Grace Ahn, Hyoungsuk Lee, Boyun Lee, Ji-eun Shin and Dayk Jang (2020), "Compensating for the Loss of Human Distinctiveness: The Use of Social Creativity under Human–Machine Comparisons," *Computers in Human Behavior*, 103 (February), 80-90.

Choi, Sungwoo, Anna S. Mattila and Lisa E. Bolton (2021), "To Err Is Human(-oid): How Do Consumers React to Robot Service Failure and Recovery?," *Journal of Service Research*, 24 (3), 354-71.

Churchill, Gilbert A. and Carol Surprenant (1982), "An Investigation into the Determinants of Customer Satisfaction," *Journal of Marketing Research*, 19 (4), 491-504.

Cohen, Jacob (1988), *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates.

Crolic, Cammy, Felipe Thomaz, Rhonda Hadi and Andrew T. Stephen (2022), "Blame the Bot: Anthropomorphism and Anger in Customer–Chatbot Interactions," *Journal of Marketing*, 86 (1), 132-48.

Crozier, R. (2017), "Lufthansa Delays Chatbot's Responses to Make It More 'Human'", (accessed September 13, 2023), [available at https://www.itnews.com.au/news/lufthansadelays-chatbots-responses-to-make-it-more-human-462643]

Dabholkar, Pratibha A. and Jeffrey W. Overby (2004), "Linking Process and Outcome to Service Quality and Customer Satisfaction Evaluations: An Investigation of Real Estate Agent Service," *International Journal of Service Industry Management*, 16 (1), 10-27.

Davis, Fred D. (1989), "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *MIS Quarterly*, 13 (3), 319-40.

Epley, Nicholas, Adam Waytz, Scott Akalis and John T. Cacioppo (2008), "When We Need A Human: Motivational Determinants of Anthropomorphism," *Social Cognition*, 26 (2), 143-55.

Epley, Nicholas, Adam Waytz and John T. Cacioppo (2007), "On Seeing Human: A Three-factor Theory of Anthropomorphism," *Psychological Review*, 114 (4), 864-86.

Feine, Jasper, Ulrich Gnewuch, Stefan Morana and Alexander Maedche (2019), "A Taxonomy of Social Cues for Conversational Agents," *International Journal of Human-Computer Studies*, 132 (December), 138-61.

Fiske, Susan T. and Patricia W. Linville (1980), "What Does the Schema Concept Buy Us?," *Personality and Social Psychology Bulletin*, 6 (4), 543-57.

Gelbrich, Katja, Julia Hagel and Chiara Orsingher (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), 176-93.

Gnewuch, U., Stefan Morana, Marc T. P. Adam and Alexander Maedche (2022), "Opposing Effects of Response Time in Human–Chatbot Interaction: The Moderating Role of Prior Experience," *Business & Information Systems Engineering*, 64 (7), 773-91.

Grand View Research (2023), "Chatbot market size, share, trends & growth report, 2030", (accessed December 15, 2023), [available at https://www.grandviewresearch.com/industry-analysis/chatbot-market]

Grimes, G. Mark, Ryan M. Schuetzler and Justin Scott Giboney (2021), "Mental Models and Expectation Violations in Conversational AI Interactions," *Decision Support Systems*, 144 (May), 113515.

Halkias, Georgius (2015), "Mental Representation of Brands: A Schema-based Approach to Consumers' Organization of Market Knowledge," *Journal of Product & Brand Management*, 24 (5), 438-48.

Han, Bing, Xun Deng and Hua Fan (2023), "Partners or Opponents? How Mindset Shapes Consumers' Attitude Toward Anthropomorphic Artificial Intelligence Service Robots," *Journal of Service Research*, 26 (3), 441-58.

Hayes, Andrew F. (2018), *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-based Approach*, 2nd ed. New York: Guilford Press.

Holtgraves, Thomas and Tai-Lin Han (2007), "A Procedure for Studying Online Conversational Processing Using a Chat Bot," *Behavior Research Methods*, 39 (1), 156-63.

Holthöwer, Jana and Jenny van Doorn (2022), "Robots Do Not Judge: Service Robots Can Alleviate Embarrassment in Service Encounters," *Journal of the Academy of Marketing Science*, 51 (4), 767-84.

Huang, Ming-Hui and Roland T. Rust (2018), "Artificial Intelligence in Service," *Journal of Service Research*, 21 (2), 155-72.

Huang, Ming-Hui and Roland T. Rust (2021), "Engaged to a Robot? The Role of AI in Service," *Journal of Service Research*, 24 (1), 30-41.

Jacquet, Baptiste, Jean Baratgin and Frank Jamet (2019), "Cooperation in Online Conversations: The Response Times as a Window into the Cognition of Language Processing," *Frontiers in Psychology*, 10 (April), 727.

Kim, Sara, Rocky P. Chen and Ke Zhang (2016), "Anthropomorphized Helpers Undermine Autonomy and Enjoyment in Computer Games," *Journal of Consumer Research*, 43 (2), 282-302.

Kim, Tae W., Li Jiang, Adam Duhachek, Hyejin Lee and Aaron Garvey (2022), "Do You Mind if I Ask You a Personal Question? How AI Service Agents Alter Consumer Self-Disclosure," *Journal of Service Research*, 25 (4), 649-66.

Kim, Youjeong and Shyam S. Sundar (2012), "Anthropomorphism of Computers: Is It Mindful or Mindless?," *Computers in Human Behavior*, 28 (1), 241-50.

Kozak, Megan N., Abigail A. Marsh and Daniel M. Wegner (2006), "What Do I Think You're Doing? Action Identification and Mind Attribution," *Journal of Personality and Social Psychology*, 90 (4), 543-55.

Larivière, Bart, David Bowen, Tor W. Andreassen, Werner H. Kunz, Nancy J. Sirianni, Chris Voss, Nancy V. Wünderlich and Arne De Keyser (2017). "'Service Encounter 2.0': An Investigation into the Roles of Technology, Employees and Customers," *Journal of Business Research*, 79 (October), 238-46.

Lee, Hyun-Joo and Jewon Lyu (2016), "Personal Values as Determinants of Intentions to Use Self-service Technology in Retailing," *Computers in Human Behavior*, 60 (July), 322-32.

Li, Xueni, Kimmy W. Chan and Sara Kim (2019), "Service with Emoticons: How Customers Interpret Employee Use of Emoticons in Online Service Encounters," *Journal of Consumer Research*, 45 (5), 973-87.

Mende, Martin, Maura L. Scott, Jenny van Doorn, Dhruv Grewal and Ilana Shanks (2019), "Service Robots Rising: How Humanoid Robots Influence Service Experiences and Elicit Compensatory Consumer Responses," *Journal of Marketing Research*, 56 (4), 535-56.

Meng, Jingbo and Yue Dai (2021), "Emotional Support from AI Chatbots: Should a Supportive Partner Self-Disclose or Not?" *Journal of Computer-Mediated Communication*, 26 (4), 207-22.

Microsoft (2023), "Visual Studio Code", (accessed September 13, 2023), [available from https://code.visualstudio.com]

Moon, Youngme (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), 379-92.

Nass, Clifford I. and Youngme Moon (2000), "Machines and Mindlessness: Social Responses to Computers," *Journal of Social Issues*, 56 (1), 81-103.

Pentina, Iryna, Tyler Hancock and Tianling Xie (2023), "Exploring Relationship Development with Social Chatbots: A Mixed-method Study of Replika," *Computers in Human Behavior* (March), 140, 107600.

Pitardi, Valentina, Jochen Wirtz, Stefanie Paluch and Werner H. Kunz (2022), "Service Robots, Agency and Embarrassing Service Encounters," *Journal of Service Management*, 33 (2), 389-14.

Puzakova, Marina and Hyokjin Kwak (2017), "Should Anthropomorphized Brands Engage Customers? The Impact of Social Crowding on Brand Preferences," *Journal of Marketing*, 81 (6), 99-115.

Reeves, Byron and Clifford I. Nass (1996), *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. New York: Cambridge University Press.

Rouse, William B. and Nancy M. Morris (1986), "On Looking into the Black Box: Prospects and Limits in the Search for Mental Models," *Psychological Bulletin*, 100 (3), 349-63.

Schanke, Scott, Gordon Burtch and Gautam Ray (2021), "Estimating the Impact of 'Humanizing' Customer Service Chatbots," *Information Systems Research*, 32 (3), 736-51.

Seeger, Anna-Maria, Jella Pfeiffer and Armin Heinzl (2021), "Texting with Humanlike Conversational Agents: Designing for Anthropomorphism," *Journal of the Association for Information Systems*, 22 (4), 931-67.

Sheehan, Ben, Hyun S. Jin and Udo Gottlieb (2020), "Customer Service Chatbots: Anthropomorphism and Adoption," *Journal of Business Research*, 115 (July), 14-24.

SnatchBot (2023), "SnatchBot" (accessed September 13, 2023), [available from https://www.snatchbot.me]

Tsekouras, Dimitrios, Ting Li and Izak Benbasat (2022), "Scratch My Back and I'll Scratch Yours: The Impact of User Effort and Recommendation Agent Effort on Perceived Recommendation Agent Quality," *Information & Management*, 59 (1), 103571.

Uysal, Ertugrul, Sascha Alavi and Valéry Bezençon (2022), "Trojan Horse or Useful Helper? A Relationship Perspective on Artificial Intelligence Assistants with Humanlike Features," *Journal of the Academy of Marketing Science*, 50 (6), 1153-75.

van Doorn, Jenny, Martin Mende, Stephanie M. Noble, John Hulland, Amy L. Ostrom, Dhruv Grewal and J. Andrew Petersen (2017), "Domo Arigato Mr. Roboto: Emergence of Automated Social Presence in Organizational Frontlines and Customers' Service Experiences," *Journal of Service Research*, 20 (1), 43-58.

Venkatesh, Viswanath, James Y. L. Thong and Xin Xu (2012), "Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology," *MIS Quarterly*, 36 (1), 157-78.

Wang, Lili, Sara Kim and Xinyue Zhou (2023), "Money in a 'Safe' Place: Money Anthropomorphism Increases Saving Behavior," *International Journal of Research in Marketing*, 40 (1), 88-108.

Wirtz, Jochen, Paul G. Patterson, Werner H. Kunz, Thorsten Gruber, Vinh N. Lu, Stephanie Paluch and Antje Martins (2018), "Brave New World: Service Robots in the Frontline," *Journal of Service Management*, 29 (5), 907-31.

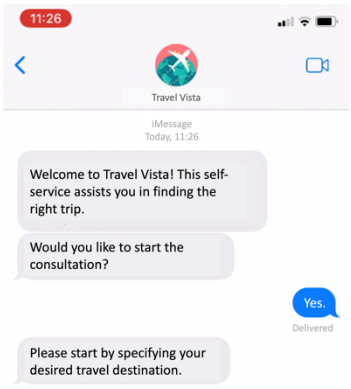
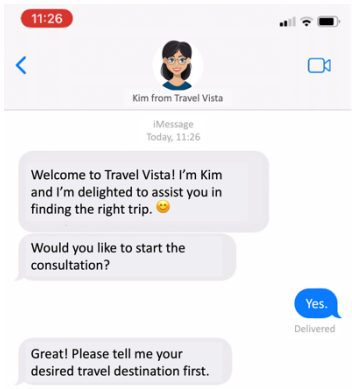
Yu, Shubin, Ji Xiong and Hao Shen (2022), "The Rise of Chatbots: The Effect of Using Chatbot Agents on Consumers' Responses to Request Rejection," *Journal of Consumer Psychology*, jcpy.1330.

Zhai, Shumin, Michael Hunter and Barton A. Smith (2002), "Performance Optimization of Virtual Keyboards," *Human–Computer Interaction*, 17 (2-3), 229-69.

Appendix

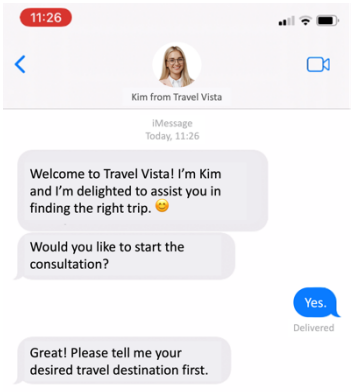
Stimuli

Study 1

Non-social	Social and social delays
	

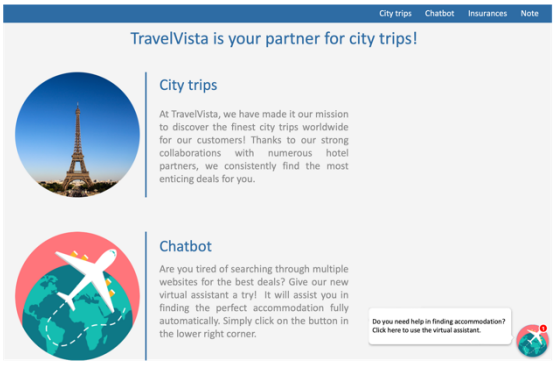
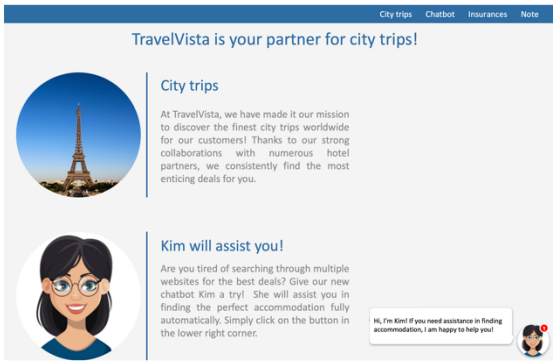
Note: The "social delays" had varying response times between 4 and 18 seconds.

Study 2

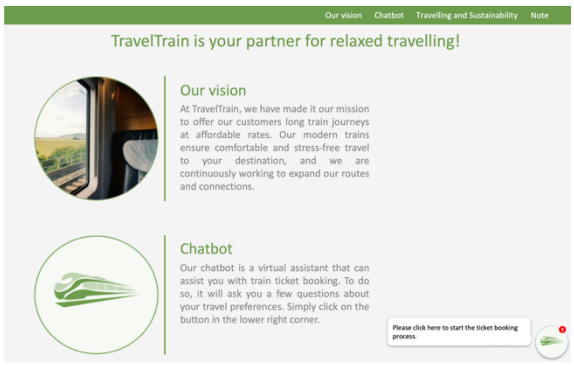
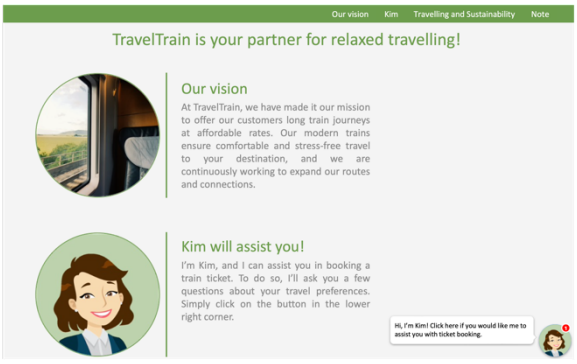
All conditions


Note: The "social delays" and the "human agent" had varying response times between 4 and 18 seconds.

Study 3

Non-social	Social and social delays
	

Study 4

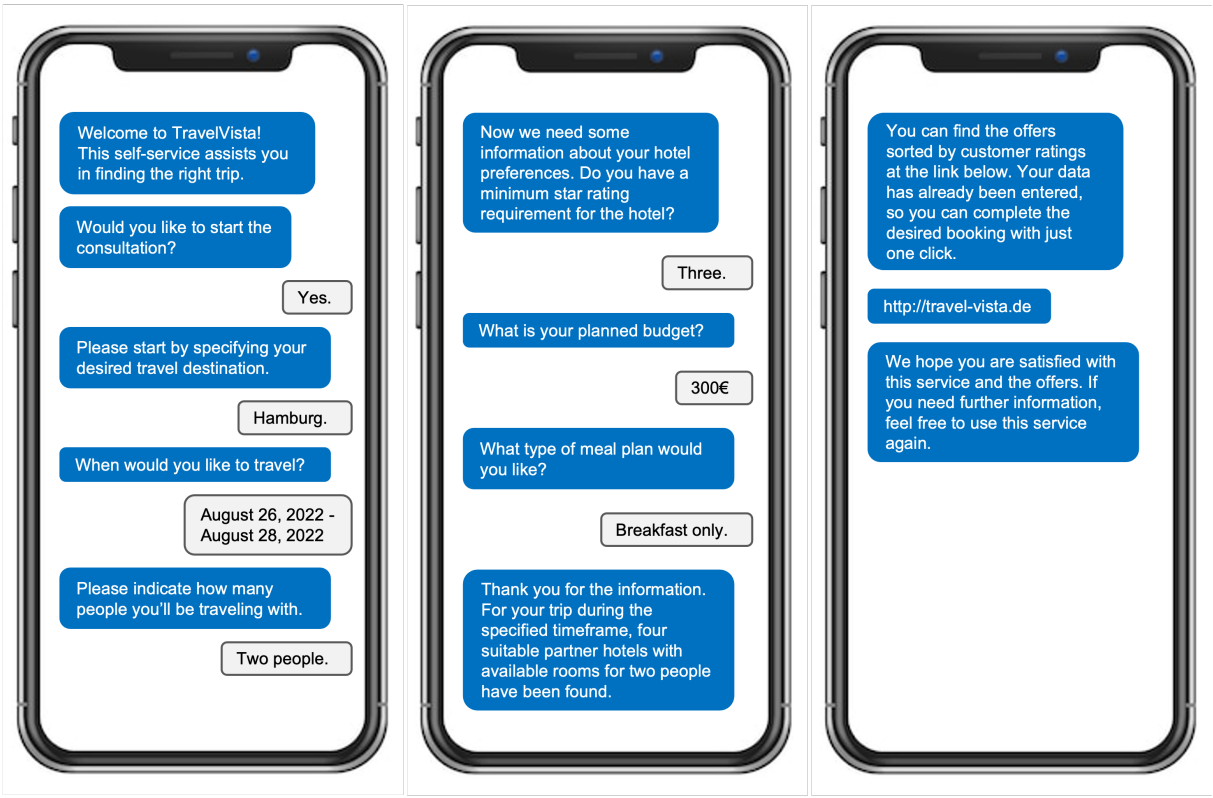
Train ticket booking website	Healthcare website
Non-social	Non-social
	
Social and social delays	Social and social delays
	

Web Appendices

Web Appendix A: Study 1

A1 Conversation Scripts

A1.1 Non-Social



A1.2 Social and Social Delays

A2 Measures

A2.1 Pre-Test

Concept	Items	Origin
Perceived human-likeness	The chatbot seems like a person. The chatbot appears human-like. I feel like the chatbot has its own personality.	Kim, Chen, and Zhang 2016
Perceived duration	The conversation was… (1) as short as possible vs. (7) longer than necessary (1) fast vs. (7) slow (1) very short vs. (7) very long	Self-developed
Perceived realism	The chatbot could exist in reality. I was able to imagine the situation well. The interaction between chatbot and user was realistic. Overall, the scenario was credible.	Gelbrich, Hagel, and Orsingher 2021

A2.2 Main Study

Concept	Items	Origin
Usage intentions	If I had access to the chatbot… …I would intent to continue using it in the future. …I could imagine well to use it for travel booking frequently. …I would always try to use it for travel booking,	Venkatesh, Thong, and Xu 2012
Perceived usefulness	Using the chatbot would help me accomplish travel bookings faster. Using the chatbot would increase my productivity regarding travel bookings. Using the chatbot would increase the efficiency of my travel bookings. Using the chatbot would simplify the process of searching for a travel. Overall, I would find the chatbot useful for travel bookings.	Davis 1989; Venkatesh, Thong, and Xu 2012
Anthropomorphism	I believe chatbots can have intentions. I believe chatbots are capable of emotion. I believe chatbots can have a free will. I believe chatbots can be conscious. I believe chatbots can have complex feelings. I believe chatbots can have a mind on their own.	Kozak and Marsh 2006; Epley, Waytz, and Cacioppo 2007

Web Appendix B: Study 2

Measures

Concept	Items	Origin
Expected usefulness	I expect a chatbot/a human advisor… …to help me accomplish travel bookings faster. …to increase my productivity regarding travel bookings. …to increase the efficiency of my travel bookings. …to simplify the process of searching for a travel. …to be overall useful for travel bookings.	Davis 1989; Venkatesh, Thong, and Xu 2012
Familiarity	I use chatbots/digital consultancies regularly. I am experienced in using chatbots/digital consultancies. I am skilled in using chatbots/digital consultancies.	Self-developed

Web Appendix C: Study 3

C1 Scenario

Now, please imagine that you want to book a weekend trip for yourself and a friend to Hamburg. The trip should take place from September 23, 2022, to September 25, 2022. You plan to spend a maximum of €300, and it would be great if breakfast is included. Otherwise, you have no other requirements.

At the beginning of your search, you come across the website of a travel booking agency called "TravelVista" specializing in city trips. On the website, you notice a chatbot used for customer support. You decide to use this chatbot.

To access the "TravelVista" website, please click on the link below. The website will open in a separate browser window, allowing you to switch back and forth between the website and this survey if you need to review your travel information.

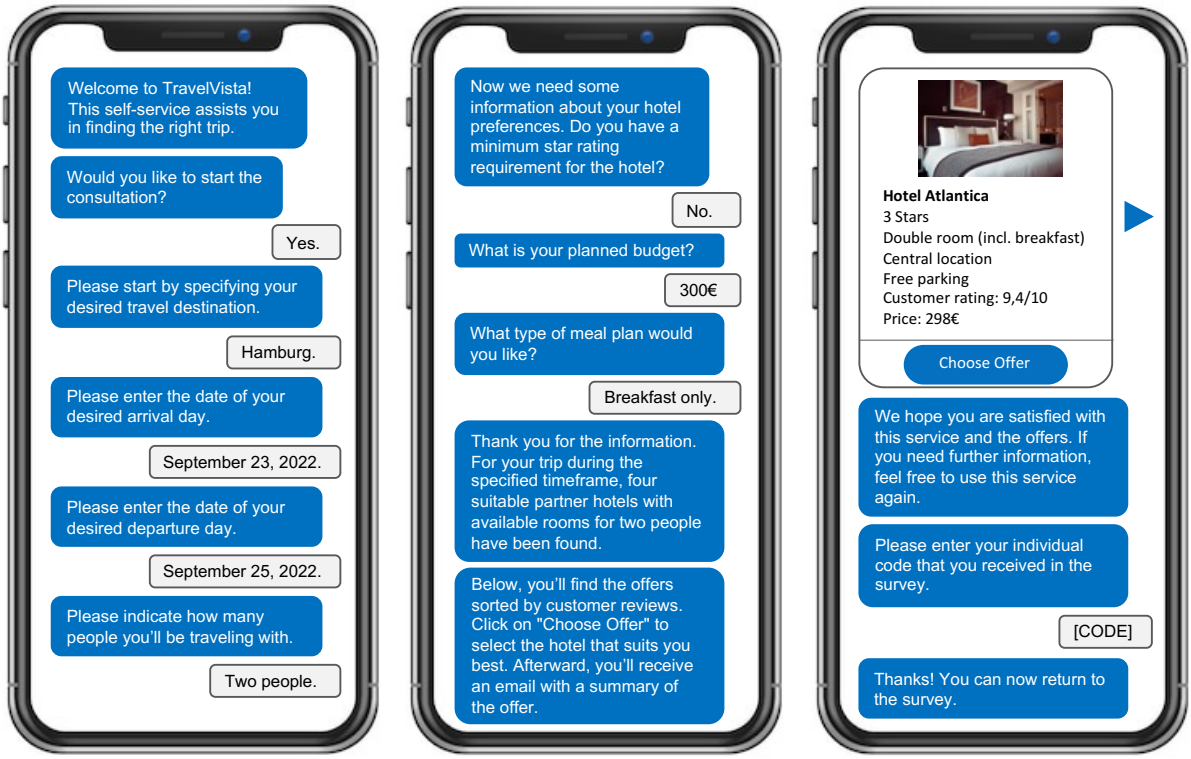
On the website, a chatbot icon will appear in the lower right corner. Please select the icon and wait for the chatbot to initiate the conversation and start the consultation. The chatbot will then ask you about your preferences in the conversation and, in the end, suggest several hotels for your trip.

Attention: Please complete the consultation fully by selecting your preferred hotel. After that, wait for a moment until the chatbot asks you for your individual code and provide it. Only after completing the conversation with the chatbot, please return to the survey. You will proceed in the survey only after interacting with the chatbot. At that point, a "Continue" button will appear.

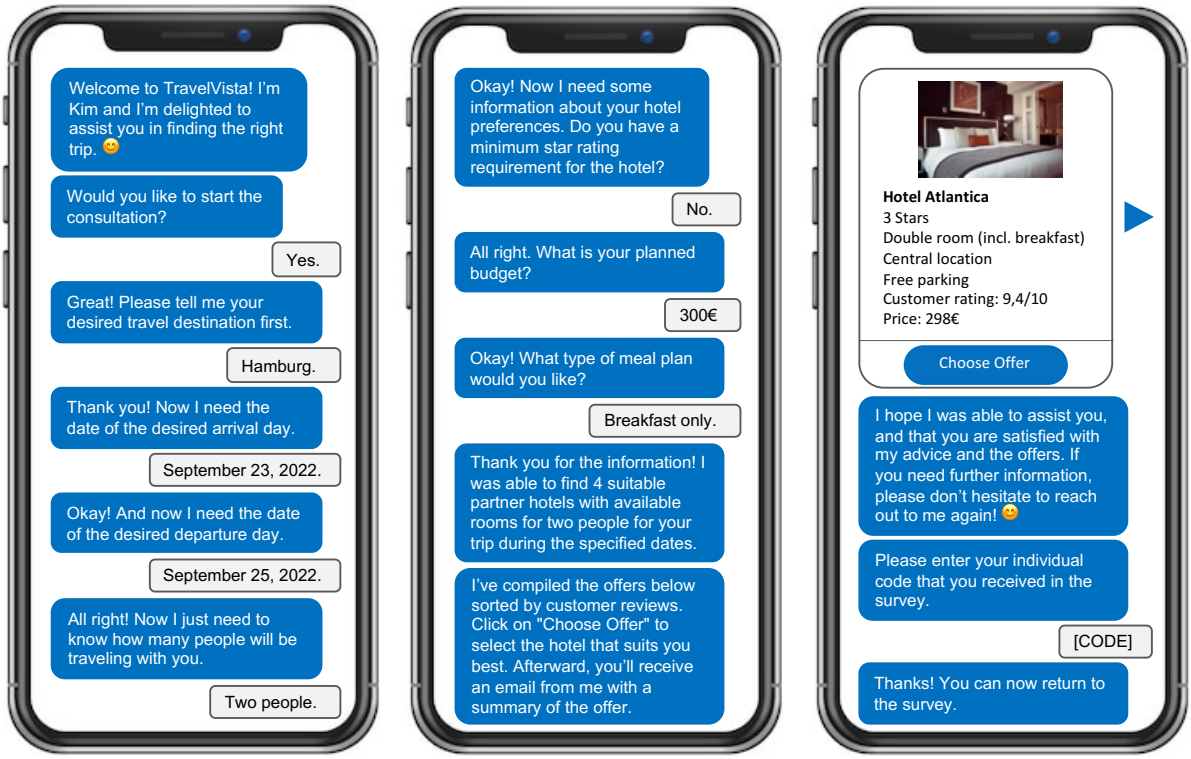
Please remind your individual code: [INDIVIDUAL CODE]

C2 Conversation Scripts

C2.1 Non-Social



C2.2 Social and Social Delays



C3 Measures

Concept	Items	Origin
Service outcome satisfaction	The hotel recommendations were (1) very dissatisfying vs. (7) very satisfying. The hotel recommendations (1) did not meet my expectations vs. (7) met my expectations. The hotel recommendations were (1) not attractive vs. (7) attractive.	Self-developed

Web Appendix D: Study 4

D1 Pre-Test Scenarios

Dear participant,

Thanks for your support in this study. This survey is about your perception and evaluation of a service situation.

Below, we will describe a [travel booking situation/ticket booking situation/situation from the healthcare context], followed by a few questions. Please answer these questions honestly and spontaneously. There are no right or wrong answers; we are interested in your very own opinion.

The Situation

Now, please imagine [that you want to book a trip/that you want to book train tickets/that you are sick and need a medical assessment on your symptoms]. To what extent do you agree with the following statements regarding [travel booking/train ticket booking/a medical assessment]?

[ITEMS ON SERVICE TASKS COMPLEXITY]

Thank you! More and more companies and service providers are digitizing their customer service by offering chats on their websites. These can include both fully automated digital assistants (e.g., chatbots) and live chats with human assistants. Now, we would like to ask you a few questions about using such a chat-based system for [travel booking/train ticket booking/a medical assessment].

In the first question, we will inquire whether you consider [travel booking/train ticket booking/a medical assessment] to be a computer-like or a human-like task. The more you believe that [travel booking/train ticket booking/a medical assessment] can be performed by a computer-assisted algorithm, the more you should place your rating towards "computer-like task". The more you believe that [travel booking/train ticket booking/a medical assessment] can only be

performed by a human assistant, the more you should place your rating towards "human-like task".

[ITEMS ON SERVICE TASK'S HUMAN-LIKENESS]

D2 Measures Pre-Test

Concept	**Items**	**Origin**
Service task's complexity	Accomplishing [SERVICE]… (1) can be done by myself vs. (7) requires assistance. For me, accomplishing [SERVICE] is… (1) very easy vs. (7) very difficult.	Self-developed
Service task's human-likeness	[SERVICE] is… (1) a computer-like task vs. (7) a human-like task. To accomplish [SERVICE]… (1) a chatbot is sufficient vs. (7) it needs a human agent.	Self-developed

D3 Scenarios

D3.1 Train Ticket Booking

Now, please imagine that you need to travel from Berlin to Munich for an appointment in early January 2023. You decide to take the train and plan to commence your journey on January 3, 2023, returning on January 6, 2023. You will be travelling alone.

As you search for train ticket options, you come across the website of "TravelTrain", a company specializing in long-distance train travel. On the website, you notice a chatbot that "TravelTrain" uses for customer support. You decide to use this chatbot.

To access the "TravelTrain" website, please click on the link below. The website will open in a separate browser window, allowing you to switch back and forth between the website and the survey if you need to review your travel preferences.

On the website, a chatbot icon will appear in the lower right corner. Please select the icon and wait for the chatbot to initiate the conversation. Ensure that you complete the conversation by clicking on "Select Ticket" at its conclusion. Finally, the chatbot will provide you with a code that is required to proceed in the survey.

Upon completing the conversation, please return to the survey. You will proceed in the survey only after interacting with the chatbot as you need the code. At that point, a "Continue" button will appear.

D3.2 Medical Assessment

Now, please imagine that you have been experiencing pressing chest pains for a few days, which are of moderate intensity and radiate to surrounding areas of your body. The pain increases when you move your upper body and when you inhale.

As you search for possible medical conditions, you come across the website of "DigiHealth", a company specializing in the digitization of healthcare services. On the website, you notice a chatbot that "DigiHealth" uses for patient consultation. You decide to use this chatbot.

To access the "DigiHealth" website, please click on the link below. The website will open in a separate browser window, allowing you to switch back and forth between the website and the survey if you need to review the information about your symptoms.

On the website, a chatbot icon will appear in the lower right corner. Please select the icon and wait for the chatbot to initiate the conversation. Ensure that you complete the conversation by clicking on "End Consultation" at its conclusion. Finally, the chatbot will provide you with a code that is required to proceed in the survey.

Upon completing the conversation, please return to the survey. You will proceed in the survey only after interacting with the chatbot as you need the code. At that point, a "Continue" button will appear.

D4 Conversation Scripts

D4.1 Train Ticket Booking

D4.1.1 Non-Social

D4.1.2 Social and Social Delays

D4.2 Medical Assessment

D4.2.1 Non-Social

D4.2.2 Social and Social Delays

Web Appendix E: Study 5

Measures

Concept	Items	Origin
Service provider evaluation	TravelVista is… (1) unfavorable vs. (7) favorable (1) negative vs. (7) positive (1) bad vs. (7) good (1) unprofessional vs. (7) professional	Crolic et al. 2022