Computers in Human Behavior: Artificial Humans 2 (2024) 100067

Contents lists available at ScienceDirect

Computers in Human Behavior: Artificial Humans

journal homepage: www.journals.elsevier.com/computers-in-human-behavior-artificial-humans

Artificial empathy in healthcare chatbots: Does it feel authentic?

Lennart Seitz *

Institute of Management & Organization, Leuphana University Lueneburg, Universitaetsallee 1, 21335, Lueneburg, Germany

A R T I C L E I N F O

Keywords:
Empathy
Chatbot
Healthcare
Authenticity
Anthropomorphism

A B S T R A C T

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.

1. Introduction

Driven by the rapid developments in AI and language processing systems, people increasingly interact with virtual assistants like chatbots (Araujo, 2018; Wirtz et al., 2018). 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 (Huang & Rust, 2018; Larivière et al., 2017).

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 (Huang & Rust, 2021). Therefore, chatbots are frequently equipped with social cues, e.g., by giving them names, avatars, or complex communication capabilities (Blut, Wang, Wünderlich, & Brock, 2021; Go & Sundar, 2019; Konya-Baumbach, Biller, & von Janda, 2023). 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 (Do, Gip, Guchait, Wang, & Baaklini, 2023; Seeger, Pfeiffer, & Heinzl, 2021). 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 (Chi, Jia, Li, & Gursoy, 2021; Pelau, Dabija, & Ene, 2021; Pepito, Ito, Betriana, Tanioka, & Locsin, 2020).

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) (Yalçın & DiPaola, 2020). 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

* Corresponding author.
E-mail address: lennart.seitz@leuphana.de.

https://doi.org/10.1016/j.chbah.2024.100067
Received 9 August 2023; Received in revised form 25 March 2024; Accepted 25 March 2024
Available online 26 March 2024

mindfulness and experiential capabilities as they describe the ability to feel, share, recognize, or understand the mental state of another (de Waal, 2008). These capabilities are, however, considered one of the key distinctions between humans and machines (Gray, Gray, & Wegner, 2007; Gray & Wegner, 2012). 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 (Do et al., 2023; Meng & Dai, 2021; Yu, Xiong, & Shen, 2022). 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., Crolic, Thomaz, Hadi, & Stephen, 2022; Kim, Chen, & Zhang, 2016).

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 (Jeffrey, 2016). Drawing on the concept of 'anthropomorphism' (Epley, Waytz, & Cacioppo, 2007), the related 'Social Response Theory' (Nass & Moon, 2000), and the 'Stereotype Content Model' (Fiske, Cuddy, Glick, & Xu, 2002), 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' (Gray et al., 2007) and the concept of authenticity (Wood, Linley, Maltby, Baliousis, & Joseph, 2008), 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 (Epley et al., 2010; Gray et al., 2007). 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 (Mayer, Davis, & Schoorman, 1995; Morhart, Malär, Guèvremont, Girardin, & Grohmann, 2015). 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 (Appel, Izydorczyk, Weber, Mara, & Lischetzke, 2020; Giger, Piçarra, Alves-Oliveira, Oliveira, & Arriaga, 2019; Mende, Scott, van Doorn, Grewal, & Shanks, 2019; Yu et al., 2022).

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 (Seitz, Bekmeier-Feuerhahn, & Gohil, 2022). Since it is not foreseeable that bots will be able to feel emotion soon (Do et al., 2023; Wirtz et al., 2018), they are sometimes imbued with artificial empathy to make conversations feel more human-like and to facilitate relationship-building (Liu & Sundar, 2018; Zhou, Gao, Li, & Shum, 2020). 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 (Brave, Nass, & Hutchinson, 2005; Liu & Sundar, 2018). Furthermore, users interacting with empathetic agents and AI show higher levels of satisfaction and usage persistence (Bickmore & Picard, 2005; Gelbrich, Hagel, & Orsingher, 2021; Lv, Yang, Qin, Cao, & Xu, 2022). Feeling a sense of empathy in bots can also enhance users' mood after social exclusion and even lead to a reduction in depressive symptoms (de Gennaro, Krumhuber, & Lucas, 2019; Fitzpatrick, Darcy, & Vierhile, 2017).

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 (Blut et al., 2021; Epley et al., 2007). 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 (Epley et al., 2007; Holtgraves, Ross, Weywadt, & Han, 2007). 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 (Krämer, Kopp, Becker-Asano, & Sommer, 2013). This general human tendency to anthropomorphize is also theorized in computer science and information systems research by the 'Social Response Theory' (Nass & Moon, 2000) and the related 'Computers Are Social Actors' paradigm (Reeves & Nass, 1996).

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 (de Waal, 2008; Kraft-Todd et al., 2017). 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 (Fiske et al., 2002). 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 (Chua, Ingram, & Morris, 2008; Mayer et al., 1995; McAllister, 1995). 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 (Borau, Otterbring, Laporte, & Fosso Wamba, 2021; Christoforakos, Gallucci, Surmava-Große, Ullrich, & Diefenbach, 2021; Gelbrich et al., 2021). A chatbot that provides a sense of warmth, e.g., by behaving empathetic, might therefore appear to be more trustworthy

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

Paper	Year	Study domain	Cue	Modality	Key findings
Klein et al. (2002)	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 (2005)	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 et al. (2005)	2005	Entertainment	Empathy, sympathy, facial expressions	Multimodal	Empathetic emotions enhance an agent's likeability, its trustworthiness, and felt support.
Nguyen et al. (2009)	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 (2018)	2018	Healthcare	Empathy (cognitive vs. affective), sympathy	Text	Expressions of affective empathy and sympathy (vs. cognitive empathy) enhance perceived support.
de Gennaro et al. (2019)	2020	Mood induction experiment	Sympathy, emojis	Text	Individuals who have experienced social exclusion report enhanced mood after interacting with an empathetic agent.
Gelbrich et al. (2021)	2021	Emotional support	Emotional support	Text	An emotionally supportive digital assistant enhances satisfaction and behavioral persistence.
Lv et al. (2022)	2022	Hospitality	Empathy, sympathy	Multimodal	Receiving a highly empathetic response from an AI after service failure enhances using intentions.

than a chatbot that responds mechanically (Brave et al., 2005; Pelau et al., 2021). 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 (Seitz et al., 2022). 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 (Chi et al., 2021; Gefen, Benbasat, & Pavlou, 2008; Glikson & Woolley, 2020).

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 (Cuff, Brown, Taylor, & Howat, 2016; Hall & Schwartz, 2019). 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 (Cuff et al., 2016). 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 (de Waal, 2008; Premack & Woodruff, 1978). 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 (Cuff et al., 2016; de Waal, 2008).

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 (de Waal, 2008; Frith, 2003). 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 (Cuff et al., 2016).

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" (Liu & Sundar, 2018). 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 (Dennis, Fuller, & Valacich, 2008). Affective empathetic reactions, in contrast, usually manifest in emotional responses that instinctively spill out, e.g., starting to cry when seeing a person cry (Frith, 2003). 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 (Cuff et al., 2016; de Waal, 2008; Jeffrey, 2016). Some scholars consider sympathy a part of empathy as it is a cognitive or affective response to another person's mental state or situation (de Waal, 2008). Researchers hence also refer to the term 'empathic concern' in describing sympathy (Hall & Schwartz, 2019). However, sympathy can also be an incongruent emotional state as it means feeling *for* another and not feeling *as* another (Escalas & Stern, 2003; Hein & Singer, 2008). 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 (Cuff et al., 2016; de Waal, 2008; Hall & Schwartz, 2019). In designing artificial empathy, *sympathetic* responses are typically implemented to chatbots by phrases like "I am sorry to hear that" (Bickmore & Picard, 2005; de Gennaro et al., 2019).

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 (Cuff et al., 2016). To worry about someone or to perceive suffering in another might trigger empathetic helping, i.e., helping someone to overcome a distressing event (de Waal, 2008). Theorists consider the emergence of empathetic helping to be either truly altruistic (i.e., helping as an expression of genuine concern and moral beliefs) (Batson, 1991; Batson, Duncan, Ackerman, Buckley, & Birch, 1981), or self-interest driven (i.e., helping with return on benefit expectations or to cope with own negative emotion) (Cialdini et al., 1987; de Waal, 2008). 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 (Lazarus & Folkman, 1984). 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 (Halpern, 2001; Jeffrey, 2016). 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?" (Leite, Castellano, Pereira, Martinho, & Paiva, 2014) or "Can you tell me more about how you feel?" (Klein, Moon, & Picard, 2002; Nguyen et al., 2009).

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 (Appel et al., 2020; Giger et al., 2019; Mori, MacDorman, & Kageki, 2012; Stein, Liebold, & Ohler, 2019). 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 (Mende et al., 2019) or by emphasizing and valuing human-unique attributes like creativity (Cha et al., 2020).

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 (Go & Sundar, 2019; Meng & Dai, 2021; Pitardi, Wirtz, Paluch, & Kunz, 2022). Schemas are cognitive frameworks that organize the knowledge we have about the attributes of certain objects (Fiske & Linville, 1980; Rouse & Morris, 1986). 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 (Grimes, Schuetzler, & Giboney, 2021). For instance, users might expect a chatbot to have lower problem-solving capabilities compared to a human (Belanche, Casaló, Flavián, & Schepers, 2020; Crolic et al., 2022; Yu et al., 2022), but to be able to respond immediately (Castelo, Boegershausen, Hildebrand, & Henkel, 2023). The perhaps most significant disparity between chatbots and humans lies in their incapacity to experience emotions (Do et al., 2023; Meng & Dai, 2021; Wirtz et al., 2018). According to 'Mind Perception Theory', the ability to think (*agency*) and the ability to feel (*experience*) are the two core dimensions of human mind (Gray et al., 2007). 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 (Gray & Wegner, 2012; Waytz & Norton, 2014). Chatbots are thus expected to provide a competent and fast service while lacking interpersonal warmth (Meng & Dai, 2021). 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 (Meng & Dai, 2021). Correspondingly, Klein et al. (2002) 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 (Heidegger, 1996; Wood et al., 2008). Authenticity is therefore related to dimensions like credibility, sincerity, and honesty and thus closely associated with someone's trustworthiness (Hennig-Thurau, Groth, Paul, & Gremler, 2006; Mayer et al., 1995; Morhart et al., 2015; Wood et al., 2008). 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 (Grandey, Fisk, Mattila, Jansen, & Sideman, 2005; Hennig-Thurau et al., 2006; Lechner, Mathmann, & Paul, 2022). 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)" (Morhart et al., 2015, 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 (Morhart et al., 2015; Moulard, Raggio, & Folse, 2016). 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 (Okubo, Kobayashi, & Ishikawa, 2012).

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 (Gray & Wegner, 2012; Meng & Dai, 2021). 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 (Meng & Dai, 2021). 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.

H4a. A healthcare chatbot responding (1) empathetic, or (2) sympathetic is perceived less authentic than a healthcare chatbot responding non-empathetic.

H4b. 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 (Crolic et al., 2022; Go & Sundar, 2019; Krämer et al., 2013). 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 (Crolic et al., 2022; Krämer et al., 2013). 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 (Schroll, 2023; Wang, Kim, & Zhou, 2023). 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 (Laughey et al., 2021). 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.

Fig. 1 summarizes all hypotheses in a holistic 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 Fig. 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' (SnatchBot, 2023). 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 (de Gennaro et al., 2019). 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.

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

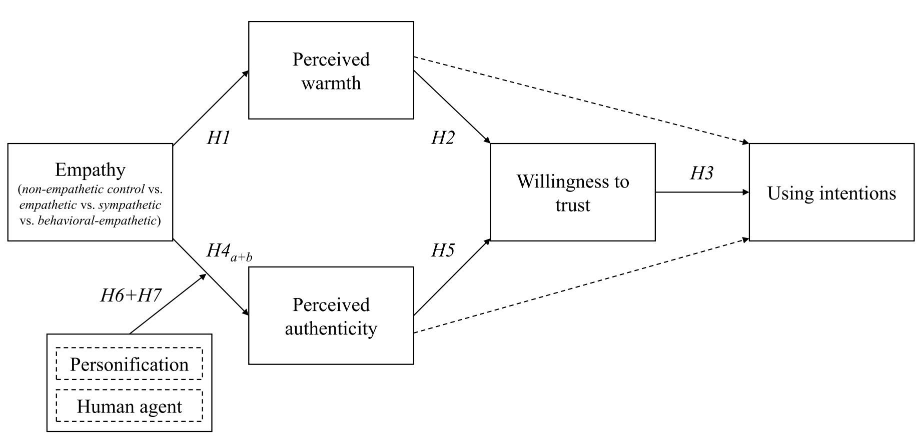


Fig. 1. Research model.

Table 2
Overview of conditions and exemplary responses.

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.

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

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. (2021)) 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 was a significant difference across groups ($F = 4.322$, $p < 0.01$). The score for the *non-empathetic control condition* (CC; $M = 3.67$) was lower than for the *empathetic* (EM; $M = 4.57$, $p < 0.02$), *sympathetic* (SY; $M = 4.45$, $p < 0.04$), and *behavioral-empathetic* (BE; $M = 4.95$, $p < 0.001$) conditions. Moreover, the empathy conditions did not differ significantly (all $ps > 0.17$). Next, three one-way ANOVAs on *perceived empathetic* ($\alpha = 0.93$), *sympathetic* ($\alpha = 0.89$), and *behavioral-empathetic* ($\alpha = 0.91$) 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 $ps < 0.01$), $F = 13.003$, $p < 0.001$. 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 $ps < 0.001$), $F = 27.539$, $p < 0.001$. 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 $ps < 0.05$), $F = 9.054$, $p < 0.001$. 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; $F = 0.004$, $p > 0.99$). Same applied for the *conversation's complexity*, that was on a low and comparable level across conditions ($M = 1.79$–2.53; $F = 2.475$, $p > 0.06$). 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 $f = 0.18$ (effect size; small to medium effect according to Cohen (1988), *power level* $= 0.80$, and *α error probability* $= 0.05$. 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. (2021) and Aaker, Vohs, & Mogilner (2010), *willingness to trust* by six items adapted from Söllner, Hoffmann, Hoffmann, Wacker, & Leimeister (2012) and McKnight, Choudhury, & Kacmar (2002) and *using intentions* by three items adapted from Venkatesh, Thong, & Xu (2012). As *perceived authenticity* has not been examined yet in comparable studies, the scale was based on conceptualizations of authentic personality (Wood et al., 2008). 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 (Seitz et al., 2022), two control variables were included: the participant's *general attitudes* towards using healthcare chatbots (adapted from Moon & Kim (2001)), 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 (Glikson & Woolley, 2020; Seitz et al., 2022). 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 (Mayer et al., 1995). 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 (Hayes, 2018). The model included *empathy* as independent variable (multicategorial; 0 = CC, 1 = BE, 2 = EM, 3 = SY), *perceived warmth* (M1, $\alpha = 0.77$) and *perceived authenticity* (M2, $\alpha = 0.84$) as first stage parallel mediators, the *willingness to trust* (M3, $\alpha = 0.90$) as second stage mediator, and *using intentions* ($\alpha = 0.93$) as dependent variable. The model also controlled for possible direct effects of *perceived warmth* and *perceived authenticity* on *using intentions* (see Fig. 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 ($M = 4.22$) ($M_{BE} = 4.78$, $b_{BE} = 0.56$, $p < 0.01$; $M_{EM} = 4.91$, $b_{EM} = 0.68$, $p < 0.001$; $M_{SY} = 4.94$, $b_{SY} = 0.71$, $p < 0.001$). *Perceived warmth*, subsequently, enhanced the *willingness to trust* the chatbot ($b = 0.36$, $p < 0.001$) hence lending credence for *H2*. Ultimately, supporting *H3*, the *willingness to trust* strongly predicted *using intentions* ($b = 0.87$, $p < 0.001$). 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* ($b_{BE} = 0.17$, [$CI = 0.06$; 0.30]; $b_{EM} = 0.21$, [$CI = 0.10$; 0.35]; $b_{SY} = 0.22$, [$CI = 0.11$; 0.36]). Moving towards the second mediator, *perceived authenticity* was lower for both the EM chatbot ($M = 4.42$, $b = -1.02$, $p < 0.001$), and the SY chatbot ($M = 4.55$, $b = -0.90$, $p < 0.001$) compared to the CC ($M = 5.45$) thus supporting *H4a*. Contradicting *H4b*, same applied for the BE chatbot, although the effect was weaker ($M = 4.90$, $b = -0.54$, $p < 0.01$). However, a one-factor ANOVA applying contrast analysis revealed that *perceived authenticity* was significantly higher for the BE chatbot compared to the EM chatbot ($b = 0.48$, $p =$

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

Predictor	Perceived warmth		Perceived authenticity		Willingness to trust		Using intentions	
	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	$R^2 = 0.07$ $F(3, 351) = 8.20, p < 0.001$		$R^2 = 0.09$ $F(3, 351) = 11.49, p < 0.001$		$R^2 = 0.26$ $F(2, 352) = 61.06, p < 0.001$		$R^2 = 0.58$ $F(3, 351) = 163.28, p < 0.001$	
With controls	$R^2 = 0.19$ $F(5, 349) = 16.60, p < 0.001$		$R^2 = 0.22$ $F(5, 349) = 19.70, p < 0.001$		$R^2 = 0.43$ $F(4, 350) = 64.83, p < 0.001$		$R^2 = 0.68$ $F(5, 349) = 151.05, p < 0.001$	

Notes: Significant effects (p < 0.05) are highlighted by **bold characters**. Parameters inside parentheses show effect sizes when including control variables.

0.014) and tendentially higher compared to the SY chatbot ($b = 0.35$, $p = 0.074$). Results further confirmed *H5* hypothesizing that *perceived authenticity* is positively related to the *willingness to trust* ($b = 0.28$, $p < 0.001$). 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 ($b_{\text{EM}} = -0.25$, [$CI = -0.38; -0.15$]; $b_{\text{SY}} = -0.22$, [$CI = -0.35; -0.12$]). Note that this – albeit smaller – effect was also observed unexpectedly for the BE chatbot ($b = -0.13$, [$CI = -0.25; -0.04$]). These opposing indirect effects resulted in an insignificant total effect of *empathy* on *trust* ($F = 0.202$, $p > 0.89$) and *using intentions* ($F = 0.478$, $p > 0.69$).

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 $ps > 0.08$). However, simple bivariate correlation analyses revealed significant correlations between both control variables and the *willingness to trust* ($r_{attitudes} = 0.59$, $p < 0.001$; $r_{risk} = -0.24$, $p < 0.001$). Both controls were therefore added to the model. The second model calculation showed robustness of the effects (see parameters in parentheses in Table 3).

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 (Blut et al., 2021; Gelbrich et al., 2021; Pelau et al., 2021). 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 (Gray et al., 2007; Waytz & Norton, 2014). 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 (Epley et al., 2007; Gray et al., 2007). 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 (Meng & Dai, 2021).

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. (2022) (α = 0.92; 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 ($M_{non\text{-}person} = 2.18$; $M_{person} = 5.33$, $p < 0.001$).

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* ($\alpha = 0.85$) remained significant for the EM chatbot ($M = 4.19$, $b = -0.78$, $p < 0.001$) and the SY chatbot ($M = 4.22$, $b = -0.75$, $p < 0.001$) when comparing to the CC ($M = 4.97$). However, for the BE chatbot, the unexpected negative effect observed in Study 1 disappeared ($M = 4.82$, $b = -0.15$, $p = 0.46$) thus partially supporting *H6*. Moreover, the positive effects on *perceived warmth* ($\alpha = 0.81$) were replicated for all chatbots providing a sense of empathy ($M_{BE} = 5.03$, $b_{BE} = 0.46$, $p < 0.01$; $M_{EM} = 5.06$, $b_{EM} = 0.50$, $p < 0.01$; $M_{SY} = 4.96$, $b_{SY} = 0.39$, $p < 0.03$) as they were perceived warmer than the CC ($M = 4.57$). Again, both *perceived warmth* ($b = 0.45$, $p < 0.001$), and *perceived authenticity* ($b = 0.25$, $p < 0.001$) were positively related to the *willingness to trust* ($\alpha = 0.91$) resulting in higher *using intentions* ($\alpha = 0.93$; $b = 0.88$, $p < 0.001$). 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 ($b_{BE} = 0.18$, [$CI = 0.04$; 0.35]; $b_{EM} = 0.20$, [$CI = 0.06$; 0.36]; $b_{SY} = 0.16$, [$CI = 0.03$; 0.30]). However, the negative indirect effect through the loss in *perceived authenticity* and *willingness to trust* only replicated for the EM chatbot ($b = -0.17$, [$CI = -0.29$; -0.07]) and the SY chatbot ($b = -0.16$, [$CI = -0.28$; -0.07]), but not the BE chatbot ($b = -0.03$, [$CI = -0.12$; 0.05]). 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* ($F = 0.559$, $p > 0.64$), and *using intentions* ($F = 1.979$, $p > 0.11$) was insignificant.

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* ($F(1, 692) = 5.364$, $p < 0.03$), and *empathy* ($F(3, 692) = 11.219$, $p < 0.001$), but no interaction effect, $F(3, 692) = 0.684$, $p = 0.56$. Continuing with *perceived authenticity*, there were significant main effects for *personification* ($F(1, 692) = 7.851$, $p < 0.01$), and *empathy* ($F(3, 692) = 18.603$, $p < 0.001$). Again, there was no interaction effect, $F(3, 692) = 0.673$, $p = 0.57$.

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 ($M_{Study\ 2} = 4.90$; $M_{Study\ 1} = 4.72$, $p < 0.04$). 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 ($M_{Study\ 2} = 4.54$; $M_{Study\ 1} = 4.82$, $p < 0.01$).

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 (Gray et al., 2007; Waytz & Norton, 2014). 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-

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

Predictor	Perceived warmth		Perceived authenticity		Willingness to trust		Using intentions	
	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	$R^2 = 0.03$		$R^2 = 0.07$		$R^2 = 0.30$		$R^2 = 0.62$	
	$F(3, 341) = 3.76$, $p = .01$		$F(3, 341) = 7.95$, $p < 0.001$		$F(2, 342) = 73.03$, $p < 0.001$		$F(3, 341) = 185.75$, $p < 0.001$	
With controls	$R^2 = 0.25$		$R^2 = 0.24$		$R^2 = 0.49$		$R^2 = 0.68$	
	$F(5, 339) = 22.89$, $p < 0.001$		$F(5, 339) = 21.39$, $p < 0.001$		$F(4, 340) = 80.38$, $p < 0.001$		$F(5, 339) = 141.98$, $p < 0.001$	

Notes: Significant effects (p < 0.05) are highlighted by **bold characters**. Parameters inside parentheses show effect sizes when including control variables.

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 (Castelo et al., 2023). 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 (Castelo et al., 2023). 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* ($\alpha = 0.78$) for none of the empathy conditions compared to the CC ($M = 5.08$) ($M_{BE} = 5.02$, $b_{BE} = -0.06$, $p = 0.74$; $M_{EM} = 5.20$, $b_{EM} = 0.12$, $p = 0.53$; $M_{SY} = 4.96$, $b_{SY} = -0.11$, $p = 0.54$). However, the positive effect of empathy on *perceived warmth* ($\alpha = 0.88$) remained significant ($M_{BE} = 5.54$, $b_{BE} = 0.37$, $p < 0.03$; $M_{EM} = 5.59$, $b_{EM} = 0.42$, $p < 0.02$; $M_{SY} = 5.71$, $b_{SY} = 0.54$, $p < 0.001$) 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* ($b = 0.60$, $p < 0.001$) and *perceived authenticity* ($b = 0.34$, $p < 0.001$) were positively related to the *willingness to trust* the human agent ($\alpha = 0.95$) ultimately facilitating *using intentions* ($\alpha = 0.94$; $b = 0.90$, $p < 0.001$). 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 ($b_{BE} = 0.20$, [$CI = 0.02$; 0.39]; $b_{EM} = 0.23$, [$CI = 0.04$; 0.42]; $b_{SY} = 0.29$, [$CI = 0.12$; 0.48]). 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* ($F = 0.199$, $p > 0.89$), and *using intentions* ($F = 0.957$, $p > 0.41$) was insignificant.

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 (Meng & Dai, 2021). 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 (Gray et al., 2007; Premack & Woodruff, 1978), *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.

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

Predictor	Perceived warmth		Perceived authenticity		Willingness to trust		Using intentions	
	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 < .01		F(3, 357) = 0.57, p = .63		F(2, 358) = 139.81, p < 0.001		F(3, 357) = 282.19, p < 0.001	
With controls	$R^2 = 0.31$		$R^2 = 0.22$		$R^2 = 0.61$		$R^2 = 0.74$	
	F(4, 356) = 39.47, p < 0.001		F(4, 356) = 24.49, p < 0.001		F(3, 357) = 186.23, p < 0.001		F(4, 356) = 258.81, p < 0.001	

Notes: Significant effects (p < 0.05) are highlighted by **bold characters**. Parameters inside parentheses show effect sizes when including control variables.

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' (Nass & Moon, 2000) and the 'Stereotype Content Model' (Fiske et al., 2002) 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 (Brave et al., 2005), using intentions (Bickmore & Picard, 2005; Lv et al., 2022), and behavioral persistence (Gelbrich et al., 2021; Klein et al., 2002). 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 (Gray et al., 2007; Waytz & Norton, 2014). *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 (Hennig-Thurau et al., 2006) or brands (Morhart et al., 2015), 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 (Morhart et al., 2015). 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' (Appel et al., 2020; Giger et al., 2019; Mori et al., 2012; Stein et al., 2019) or unrealistic high expectations humanized chatbots might elicit in consumers (Crolic et al., 2022). 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 (Castelo et al., 2023; Efendić, Van De Calseyde, & Evans, 2020).

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 (Blut et al., 2021). 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 (Seeger et al., 2021). 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 (Epley et al., 2007) 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 (Lv et al., 2022). 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 (Seeger et al., 2021). 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) (Diederich et al., 2019). 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 (Appel et al., 2020; Crolic et al., 2022; Mende et al., 2019; Yu et al., 2022). 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 (Pitardi et al., 2022).

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" (Huang & Rust, 2018). 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 (Gambino, Fox, & Ratan, 2020; Rouse & Morris, 1986). Anthropomorphism and social responses towards computers are considered cognitive biases that are more likely to occur when people have little knowledge about an agent (Epley et al., 2007; Nass & Moon, 2000). 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 revealed 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.

Funding

This work was supported by the Federal Ministry of Education and Research (16SV8303). The funder had no role in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

CRediT authorship contribution statement

Lennart Seitz: Conceptualization, Data curation, Formal analysis, Methodology, Project administration, Software, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

I would like to thank Anne Diedrich for her support in literature research, her inspiring ideas, and all the fruitful discussions on artificial empathy. I would further like to thank my supervisor Prof. Dr. Sigrid Bekmeier-Feuerhahn for her helpful comments during the development of the manuscript.

Appendices.

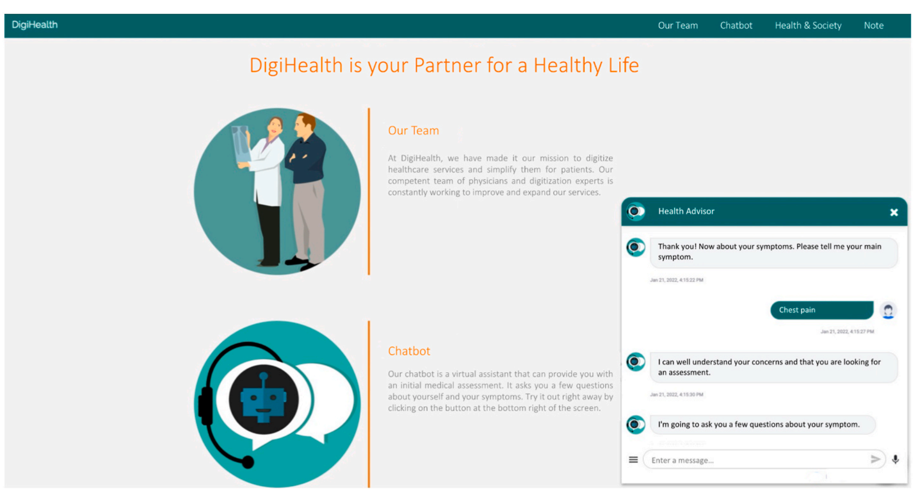
Appendix A. Measurements of pre-test (Study 1)

Measurement/Items	Cronbach's Alpha
Perceived empathetic responses (self-developed based on empathy theories)	0.93
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)	0.89
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)	0.91
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. (2021))	0.93
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).

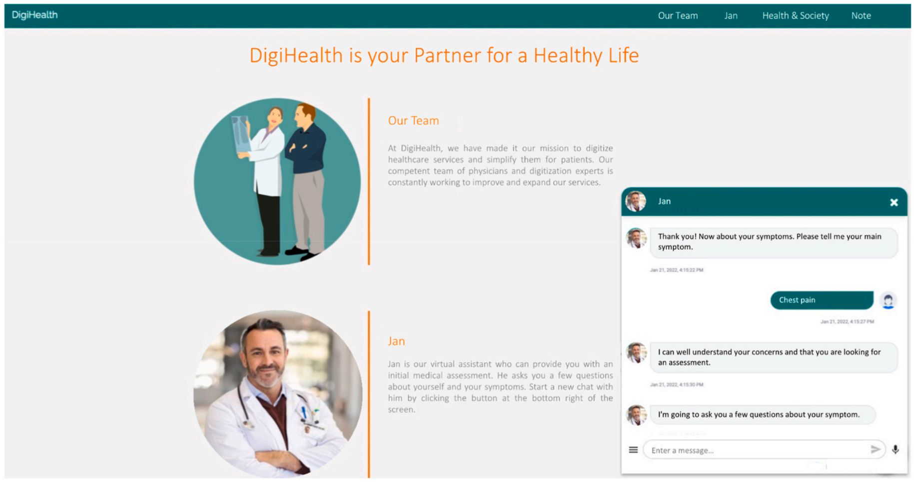
Appendix B. Screenshots of websites and chatbot pop-ups (Study 1 and Study 2)

Study 1



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

Study 2



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

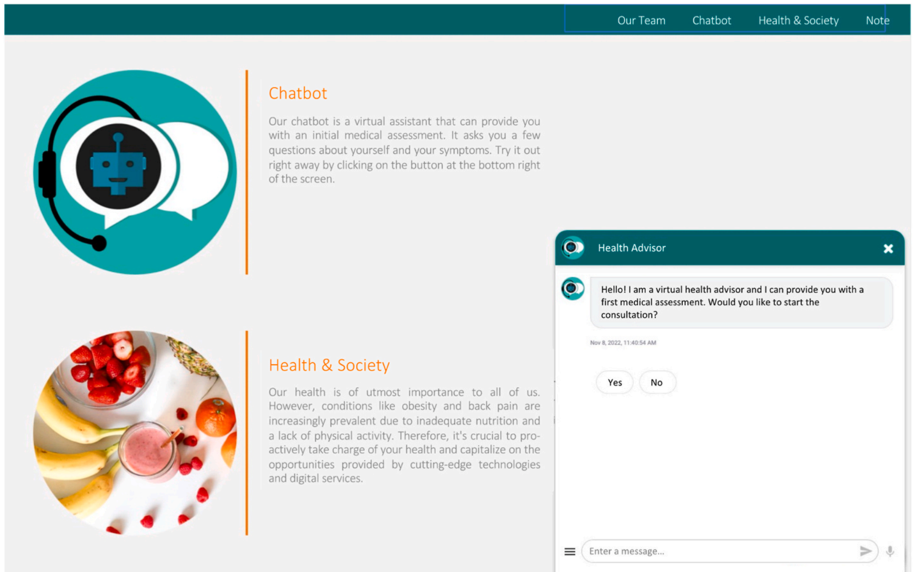
Appendix C. Measurements of all main studies (Study 1, Study 2, Study 3)

Measurement/Items	Cronbach's Alpha
Perceived warmth (adapted from Gelbrich et al. (2021) and Aaker et al. (2010))	0.77/.81/.88
The chatbot (physician) was …	
warm.	
kind.	
friendly.	
Perceived authenticity (self-developed based on theories on authentic personality)	0.84/.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. (2012) and McKnight et al. (2002))	0.90/.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. (2012))	0.93/.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 (2001))	0.91/.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)	0.92/.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.	

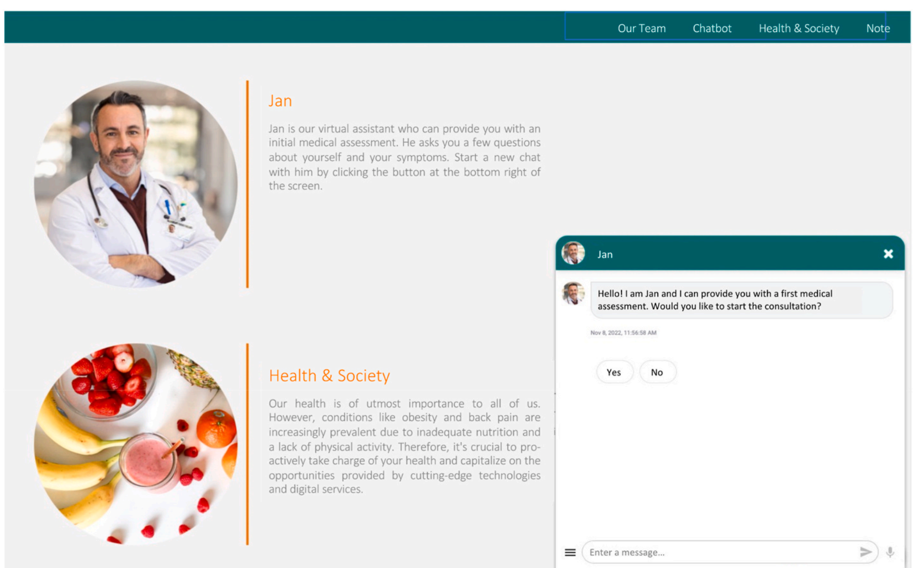
Note: Seven-point Likert scales with 1 = "strongly disagree" and 7 = "strongly agree" (if not indicated otherwise).

Appendix D. Screenshots of pre-test (Study 2)

Non-personified chatbot



Personified chatbot

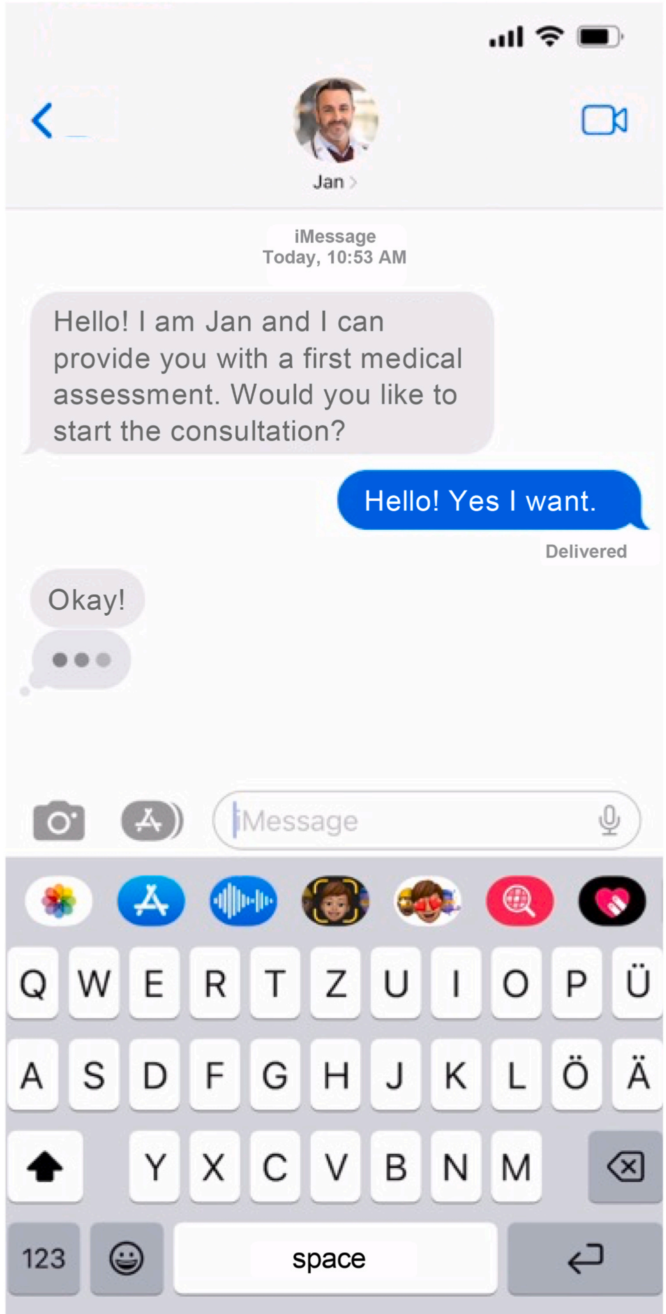


Appendix E. Measurements of pre-test (Study 2)

Measurement/Items	Cronbach's Alpha
Personification (adapted from Crolic et al. (2022))	0.92
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)



References

Aaker, J. L., Vohs, K. D., & Mogilner, C. (2010). Nonprofits are seen as warm and for-profits as competent: Firm stereotypes matter. *Journal of Consumer Research, 37*(2), 224–237. https://doi.org/10.1086/651566

Appel, M., Izydorczyk, D., Weber, S., Mara, M., & Lischetzke, T. (2020). The uncanny of mind in a machine: Humanoid robots as tools, agents, and experiencers. *Computers in Human Behavior, 102*, 274–286. https://doi.org/10.1016/j.chb.2019.07.031

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

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

Batson, C. D., Duncan, B. D., Ackerman, P., Buckley, T., & Birch, K. (1981). Is empathic emotion a source of altruistic motivation? *Journal of Personality and Social Psychology, 40*(2), 290–302. https://doi.org/10.1037/0022-3514.40.2.290

Belanche, D., Casaló, L. V., Flavián, C., & Schepers, J. (2020). Robots or frontline employees? Exploring customers' attributions of responsibility and stability after service failure or success. *Journal of Service Management, 31*(2), 267–289. https://doi.org/10.1108/JOSM-05-2019-0156

Bickmore, T. W., & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction, 12*(2), 293–327. https://doi.org/10.1145/1067860.1067867

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
Brave, S., Nass, C., & Hutchinson, K. (2005). Computers that care: Investigating the effects of orientation of emotion exhibited by an embodied computer agent. *International Journal of Human-Computer Studies, 62*(2), 161–178. https://doi.org/10.1016/j.ijhcs.2004.11.002
Castelo, N., Boegershausen, J., Hildebrand, C., & Henkel, A. P. (2023). Understanding and improving consumer reactions to service bots. *Journal of Consumer Research, ucad023*. https://doi.org/10.1093/jcr/ucad023
Cha, Y.-J., Baek, S., Ahn, G., Lee, H., Lee, B., Shin, J., et al. (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
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*, Article 106700. https://doi.org/10.1016/j.chb.2021.106700
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
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
Cialdini, R. B., Schaller, M., Houlihan, D., Arps, K., Fultz, J., & Beaman, A. L. (1987). Empathy-based helping: Is it selflessly or selfishly motivated? *Journal of Personality and Social Psychology, 52*, 749–758. https://doi.org/10.1037/0022-3514.52.4.749
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale: Lawrence Erlbaum Associates.
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
Cuff, B. M. P., Brown, S. J., Taylor, L., & Howat, D. J. (2016). Empathy: A review of the concept. *Emotion Review, 8*(2), 144–153. https://doi.org/10.1177/1754073914558466
de Gennaro, M., Krumhuber, E. G., & Lucas, G. (2019). Effectiveness of an empathic chatbot in combating adverse effects of social exclusion on mood. *Frontiers in Psychology, 10*(3061), 1–14. https://doi.org/10.3389/fpsyg.2019.03061
de Waal, F. B. M. (2008). Putting the altruism back into altruism: The evolution of empathy. *Annual Review of Psychology, 59*, 279–300. https://doi.org/10.1146/annurev.psych.59.103006.093625
Dennis, A. R., Fuller, R. M., & Valacich, J. S. (2008). Media, tasks, and communication processes: A theory of media synchronicity. *MIS Quarterly, 32*(3), 575–600. https://doi.org/10.2307/25148857
Diederich, S., Janßen-Müller, M., Brendel, A. B., & Morana, S. (2019). Emulating empathetic behavior in online service encounters with sentiment-adaptive responses: Insights from an experiment with a conversational agent. In *ICIS 2019 Proceedings*. Atlanta: Association for Information Systems.
Do, K. T., Gip, H., Guchait, P., Wang, C.-Y., & Baaklini, E. S. (2023). Empathetic creativity for frontline employees in the age of service robots: Conceptualization and scale development. *Journal of Service Management, 34*(3), 433–466. https://doi.org/10.1108/JOSM-09-2021-0352
Efendić, E., Van De Calseyde, P. P. F. M., & Evans, A. M. (2020). Slow response times undermine trust in algorithmic (but not human) predictions. *Organizational Behavior and Human Decision Processes, 157*, 103–114. https://doi.org/10.1016/j.obhdp.2020.01.008
Epley, N., & Waytz, A. (2010). Mind perception. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (pp. 498–541). Hoboken: John Wiley & Sons, Inc.. https://doi.org/10.1002/9780470561119.socpsy001014
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
Escalas, J. E., & Stern, B. B. (2003). Sympathy and empathy: Emotional responses to advertising dramas. *Journal of Consumer Research, 29*(4), 566–578. https://doi.org/10.1086/346251
Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82*(6), 878–902. https://doi.org/10.1037/0022-3514.82.6.878
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
Frith, U. (2003). *Autism: Explaining the enigma* (2nd ed.). Oxford: Blackwell.
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
Giger, J., Piçarra, N., Alves-Oliveira, P., Oliveira, R., & Arriaga, P. (2019). Humanization of robots: Is it really such a good idea? *Human Behavior and Emerging Technologies, 1*(2), 111–123. https://doi.org/10.1002/hbe2.147
Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *The 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
Grandey, A. A., Fisk, G. M., Mattila, A. S., Jansen, K. J., & Sideman, L. A. (2005). Is "service with a smile" enough? Authenticity of positive displays during service encounters. *Organizational Behavior and Human Decision Processes, 96*(1), 38–55. https://doi.org/10.1016/j.obhdp.2004.08.002
Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science, 315*(5812). https://doi.org/10.1126/science.1134475, 619–619.
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
Grimes, G. M., Schuetzler, R. M., & Giboney, J. S. (2021). Mental models and expectation violations in conversational AI interactions. *Decision Support Systems, 144*, Article 113515. https://doi.org/10.1016/j.dss.2021.113515
Hall, J. A., & Schwartz, R. (2019). Empathy present and future. *The Journal of Social Psychology, 159*(3), 225–243. https://doi.org/10.1080/00224545.2018.1477442
Halpern, J. (2001). *From detached concern to empathy: Humanizing medical practice*. New York: Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780195111194.001.0001
Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (2nd ed.). New York: Guilford Press.
Heidegger, M. (1996). *Being and time*. Albany: State University of New York Press.
Hein, G., & Singer, T. (2008). I feel how you feel but not always: The empathic brain and its modulation. *Current Opinion in Neurobiology, 18*(2), 153–158. https://doi.org/10.1016/j.conb.2008.07.012
Hennig-Thurau, T., Groth, M., Paul, M., & Gremler, D. D. (2006). Are all smiles created equal? How emotional contagion and emotional labor affect service relationships. *Journal of Marketing, 70*(3), 58–73. https://doi.org/10.1509/jmkg.70.3.058
Holtgraves, T. M., Ross, S. J., Weywadt, C. R., & Han, T. L. (2007). Perceiving artificial social agents. *Computers in Human Behavior, 23*(5), 2163–2174. https://doi.org/10.1016/j.chb.2006.02.017
Huang, M.-H., & Rust, R. T. (2018). Artificial intelligence in service. *Journal of Service Research, 21*(2), 155–172. https://doi.org/10.1177/1094670517752459
Huang, M.-H., & Rust, R. T. (2021). Engaged to a robot? The role of AI in service. *Journal of Service Research, 24*(1), 30–41. https://doi.org/10.1177/1094670520902266
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
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
Klein, J., Moon, Y., & Picard, R. W. (2002). This computer responds to user frustration: Theory, design, and results. *Interactive Computing, 14*(2), 119–140. https://doi.org/10.1016/S0953-5438(01)00053-4
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
Kraft-Todd, G. T., Reinero, D. A., Kelley, J. M., Heberlein, A. S., Baer, L., & Riess, H. (2017). Empathic nonverbal behavior increases ratings of both warmth and competence in a medical context. *PLoS One, 12*(5), Article e0177758. https://doi.org/10.1371/journal.pone.0177758
Krämer, N., Kopp, S., Becker-Asano, C., & Sommer, N. (2013). Smile and the world will smile with you - the effects of a virtual agent's smile on users' evaluation and behavior. *International Journal of Human-Computer Studies, 71*(3), 335–349. https://doi.org/10.1016/j.ijhcs.2012.09.006
Larivière, B., Bowen, D., Andreassen, T. W., Kunz, W., Sirianni, N. J., Voss, C., et al. (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
Laughey, W. F., Brown, M. E. L., Dueñas, A. N., Archer, R., Whitwell, M. R., Liu, A., et al. (2021). How medical school alters empathy: Student love and break up letters to empathy for patients. *Medical Education, 55*(3), 394–403. https://doi.org/10.1111/medu.14403
Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping*. New York: Springer.
Lechner, A. T., Mathmann, F., & Paul, M. (2022). Frontline employees' display of fake smiles and angry faces: When and why they influence service performance. *Journal of Service Research, 25*(2), 211–226. https://doi.org/10.1177/1094670520975148
Leite, I., Castellano, G., Pereira, A., Martinho, C., & Paiva, A. (2014). Empathic robots for long-term interaction evaluating social presence, engagement and perceived support in children. *International Journal of Social Robotics, 6*(3), 329–341. https://doi.org/10.1007/s12369-014-0227-1

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

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

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

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., 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

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

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

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

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

Nass, C. I., & 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

Nguyen, H., & Masthoff, J. (2009). 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, Article 7*. New York: Association for Computing Machinery. https://doi.org/10.1145/1541948.1541958.

Okubo, M., Kobayashi, A., & Ishikawa, K. (2012). A fake smile thwarts cheater detection. *Journal of Nonverbal Behavior, 36*(3), 217–225. https://doi.org/10.1007/s10919-012-0134-9

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

Pepito, J. A., Ito, H., Betriana, F., Tanioka, T., & Locsin, R. C. (2020). Intelligent humanoid robots expressing artificial humanlike empathy in nursing situations. *Nursing Philosophy, 21*(4), Article e12318. https://doi.org/10.1111/nup.12318

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

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

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

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

Schroll, R. (2023). "Ouch!" When and why food anthropomorphism negatively affects consumption. *Journal of Consumer Psychology, 33*(3), 561–574. https://doi.org/10.1002/jcpy.1316

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., & 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*, Article 102848. https://doi.org/10.1016/j.ijhcs.2022.102848

SnatchBot, SnatchBot.(2023). https://snatchbot.me. (Accessed 22 March 2024).

Söllner, M., Hoffmann, A., Hoffmann, H., Wacker, A., & Leimeister, J. M. (2012). Understanding the formation of trust in IT artifacts. In *ICIS 2012 Proceedings*. Atlanta: Association for Information Systems.

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

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

Wang, L., Kim, S., & Zhou, X. (2023). Money in a "safe" place: Money anthropomorphism increases saving behavior. *International Journal of Research in Marketing, 40*(1), 88–108. https://doi.org/10.1016/j.ijresmar.2022.02.001

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

Wirtz, J., Patterson, P. G., Kunz, W. H., Gruber, T., Lu, V. N., Paluch, S., et al. (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

Yalçın, Ö. N., & DiPaola, S. (2020). Modeling empathy: Building a link between affective and cognitive processes. *Artificial Intelligence Review, 53*(4), 2983–3006. https://doi.org/10.1007/s10462-019-09753-0

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, jcpy1330*. https://doi.org/10.1002/jcpy.1330

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