RESEARCH PAPER

Seeking empathy or suggesting a solution? Effects of chatbot messages on service failure recovery

Martin Haupt[1,2] · Anna Rozumowski[3] · Jan Freidank[2] · Alexander Haas[1]

Abstract

Chatbots as prominent form of conversational agents are increasingly implemented as a user interface for digital customer-firm interactions on digital platforms and electronic markets, but they often fail to deliver suitable responses to user requests. In turn, individuals are left dissatisfied and turn away from chatbots, which harms successful chatbot implementation and ultimately firm's service performance. Based on the stereotype content model, this paper explores the impact of two universally usable failure recovery messages as a strategy to preserve users' post-recovery satisfaction and chatbot re-use intentions. Results of three experiments show that chatbot recovery messages have a positive effect on recovery responses, mediated by different elicited social cognitions. In particular, a solution-oriented message elicits stronger competence evaluations, whereas an empathy-seeking message leads to stronger warmth evaluations. The preference for one of these message types over the other depends on failure attribution and failure frequency. This study provides meaningful insights for chatbot technology developers and marketers seeking to understand and improve customer experience with digital conversational agents in a cost-effective way.

Responsible Editor: Martin Adam

✉ Martin Haupt
martin.haupt@w.thm.de

Anna Rozumowski
anna.rozumowski@fhnw.ch

Jan Freidank
jan.freidank@w.thm.de

Alexander Haas
alexander.haas@wirtschaft.uni-giessen.de

1   Department of Marketing and Sales Management, Justus-Liebig University, Licher Strasse 66, 35394 Giessen, Germany

2   THM Business School, Technische Hochschule Mittelhessen, Wiesenstrasse 14, 35390 Giessen, Germany

3   School of Business, Institute for Competitiveness and Communication, FHNW University of Applied Sciences and Arts Northwestern Switzerland, Riggenbachstrasse 16, 4600 Olten, Switzerland

Introduction

Driven by innovative technological advancements such as artificial intelligence or machine learning, chatbots are widely used nowadays and provide customer service on digital platforms such as social media, enterprise messengers, or websites (Pizzi et al., 2021; Stoeckli et al., 2020). These agents increasingly substitute for human staff in electronic markets (van Pinxteren et al., 2020) and the global chatbot market is predicted to rise substantially from $17 billion in 2020 to over $102 billion in 2026 (Mordor Intelligence, 2021). As a remarkable and most recent example, Open AI's "ChatGPT" has attracted over 1 million users in 5 days, and is sought to disrupt numerous tasks in marketing, law, or journalism and might even threaten Google by offering more humanlike answers and a smoother experience (Olson, 2022).

However, despite these technological advancements and considerable market potential, chatbots often fail in practice to deliver satisfactory responses to users' requests (Adam et al., 2021; Brandtzaeg & Følstad, 2018; Seeger & Heinzl, 2021). Customers are often left dissatisfied

after receiving a response failure message from chatbots, which leads firms to risk negative consequences such as usage discontinuance and a decrease in firm performance (Diederich et al., 2020; Weiler et al., 2022). According to a recent survey from the banking industry, four out of five consumers are dissatisfied with chatbot interactions and almost 75% of consumers confirm that chatbots are often unable to provide correct answers (Sporrer, 2021). Concerning the consequences, about one-third of consumers (30%) stated that they would turn away from the company or spread negative word of mouth after just one negative experience with the chatbot. Due to that threat and high levels of service failures, numerous companies including Facebook or SAP shut down chatbots on their digital platforms (Dilmegani, 2022; Thorbecke, 2022).

A chatbot response failure refers to an inadequate answer or no answer at all, which is sometimes also labelled as conversational breakdown (Benner et al., 2021; Weiler et al., 2022). Chatbot response failures reflect a service failure for the company, as the digital agent was unable to deliver satisfying information to support users' goals. Users evaluate chatbot's response failures as an insufficient service offer, comparable to response failures from a human frontline service employee, service robots, or other digital self-service technologies (Sungwoo Choi et al., 2021; Collier et al., 2017). Service failures have serious impacts on firms as they harm favorable customer reactions such as satisfaction, loyalty, or positive word-of-mouth (Roschk & Gelbrich, 2014). According to research from Qualtrics and ServiceNow (2021), almost half of the respondents consider switching brands already after a single negative customer service interaction, and US companies risk losing around $1.9 trillion of customer spending annually due to such poor experiences. This is particularly relevant for digital platforms (e.g., Airbnb, Uber), as their major basis for value creation resides in providing "efficient and convenient facilitation of transactions" (Hein et al., 2020, p. 91). In contrast to other industries, these platforms highly depend on their service offer (instead of products) and positive user experiences. Thus, providing the option to use chatbots offers large efficiencies for them, but at the same time also pose a threat in case of insufficient implementation.

Failure recovery strategies are therefore urgently needed to mitigate negative user responses and financial losses. In this regard, recovery messages are suggested as a viable option for chatbots to mitigate negative responses after self-inflicted response failure by the chatbots (Ashktorab et al., 2019; Benner et al., 2021). Such recovery messages aim to increase the chatbot's response capabilities to address the response failure, but also to mitigate negative user reactions to reduce the impacts of the perceived service failure.

Yet, relatively little is known about the impact of recovery strategies in chatbot conversations on customer responses.

Recent studies rather focused on reasons for response failures (Janssen et al., 2021; Reinkemeier & Gnewuch, 2022), identified different recovery strategies (Benner et al., 2021), or assessed user preferences for diverse recovery strategies (Ashktorab et al., 2019). These scholars asserted the potential of such strategies to prevent negative reactions following a failure (Benner et al., 2021; Reinkemeier & Gnewuch, 2022), but analyses of the effectiveness of recovery messages or comparisons of different types remain scarce. In contrast to such post-failure messages, Weiler et al. (2022) investigated how ex-ante messages (i.e., at the beginning of the chatbot interaction) influence users discontinuance of the chatbot interaction.

Based on the stereotype content model (Cuddy et al., 2008) and the results of a pilot study which assessed chatbot's failure recovery strategies in real life, this study investigates the effects of two fundamental recovery message types—namely, seeking user's empathy versus suggesting a solution—on user's perceived warmth and competence, as well as on post-recovery responses. Although research about recovery messages remains largely unexplored, these two message types reflect two relevant recovery strategies as presented in the literature-based analysis of Benner et al. (2021). To receive initial empirical evidence, the pilot study assessed the failure recovery messages of 101 chatbots from business, education, and public administration and revealed that these two message types (along with a simple error message) reflect the mainly used recovery attempts in business practice.

Furthermore, this research aims to understand under which circumstances which message type is advantageous regarding user satisfaction. Therefore, it considers two situational factors, in particular failure attribution and failure frequency. Results of three experimental studies show that both messages (i.e., empathy and solution) trigger specific social cognitions, more precisely either higher warmth or competence perceptions. In turn, these perceptions were found to influence people's post-recovery satisfaction and re-use intentions—but they do so to different degrees depending on the context.

This research contributes to the growing literature in information systems (IS) related to chatbots as digital conversational agents and offers relevant implications for firms on how their chatbots should respond to a response failure in different contexts. Thereby, we integrate the technological (and IS) perspective related to chatbots' limited functionality and response failures with the service-oriented (and consumer psychology) perspective of recovery attempts to the service failure occurred. Our findings highlight the possibility to use recovery messages as low-cost, easy to program, and universally usable strategy. Furthermore, they reveal the need to design a chatbot conversation carefully, and that the choice of an effective recovery message depends

on situational factors. Recommendations for chatbot software developers and chatbot-employing firms are provided.

Conceptual background

Chatbots as digital conversational agents

Chatbots are text-based digital conversational agents that use natural language processing (NLP) to interact with users (Gnewuch et al., 2017; Wirtz et al., 2018). These features lead to higher interaction and intelligence levels compared to other IS technologies (Maedche et al., 2019). Chatbots are a cost-effective tool for companies to automate customer-firm interactions while maintaining value and personalized service for their clients. Due to the convenient, easy, and fast service and their 24/7 availability, the integration of chatbots is growing exponentially in various industries such as service, hospitality, healthcare, or education (van Pinxteren et al., 2020). With the rise of chatbots, research increased tremendously in the last years, and scholars mainly investigated chatbot interactions from three perspectives, namely digital agent's design elements (Diederich et al., 2020; Gnewuch et al., 2018; Kull et al., 2021), consumer responses to the digital interaction (Mozafari et al., 2022), and consumer responses to chatbot failures (see Sands et al., 2022 for an overview). Among these research fields, finding appropriate solutions for recovery of chatbot failures is particularly relevant, as it determines consumers' continuance decisions and ultimately a chatbot's success (Adam et al., 2021; Lv et al., 2022a, 2022b; Song et al., 2022). This is because, despite continuous development and the promising advantages for both customers and companies in service encounters, chatbots often do not live up to customer expectations and fail to understand or process user enquiries (Lv et al., 2022a, 2022b; Weiler et al., 2022; Xu & Liu, 2022).

Chatbot response failures

Lately, scholars have started to analyze the impacts of chatbot response failures. For instance, Seeger and Heinzl (2021) showed that digital agent's failures harm customer trust and stimulate negative word-of-mouth. Chatbot response failures also increase people's frustration and anger (Gnewuch et al., 2017; Mozafari et al., 2022; van der Goot et al., 2021), and create skepticism and reluctance to follow the bot's instructions (Adam et al., 2021). As a consequence, users frequently quit the conversation (Akhtar et al., 2019) and might even reject future chatbot interactions (Benner et al., 2021; van der Goot et al., 2021).

Chatbots fail frequently, because the processing of natural language input was found to be a complex task for machines due to unpredictable entries (Brendel et al., 2020). Moreover, chatbots are often integrated on digital platforms in wrong use cases and are not connected to relevant data sources (Janssen et al., 2021; Mostafa & Kasamani, 2022). In addition, users were found to have exaggerated expectations of chatbots due to their humanlike design. According to the "computers are social actors" (CASA) paradigm, people ascribe social rules, norms, and expectations to interactions with computers although they are aware that they are interacting with a machine (Nass et al., 1996). As such, people expect a chatbot to understand their request and respond with a suitable answer, just as they would expect of a human (Wirtz et al., 2018).

Parallel to the increased interest in chatbot technology, research on chatbot failure recovery strategies has gained traction in recent years (see Table 1 for an overview). This literature stream can be divided into three major sub-divisions. First, some scholars reviewed the literature or conducted expert interviews to derive critical success factors for chatbot interactions (Janssen et al., 2021) or categories of recovery strategies (Benner et al., 2021; Poser et al., 2021). The second body of research empirically examines how chatbot interaction could be designed pre-failure in order to mitigate negative consumer perceptions due to failures. For instance, research results indicate that higher chatbot anthropomorphism (Seeger & Heinzl, 2021; Sheehan et al., 2020) or specific message techniques (Weiler et al., 2022) positively influence consumer responses before the failure occurs. Third, and contrasting this, other scholars investigated the effects of post-failure recovery strategies. As one of the first studies, Ashktorab et al. (2019) compared user preferences of eight different recovery strategies and found that providing explanations or options of answers are favored as they display chatbot initiative. Mozafari et al. (2022) assessed that the mere disclosure of the chatbot (vs. human) identity has already a mitigating effect following failure. Further scholars found that chatbots are preferred over human agents after a functional failure (but not after a non-functional failure) (Xing et al., 2022), and chatbot self-recovery (vs. human agent recovery) leads to more positive user reactions (Song et al., 2022).

Scholars have also started to investigate effects of post-failure messages and discovered for instance that some communication patterns (e.g., chatbot as "victim" or "helper") lead to more positive responses than other patterns (e.g., "persecutor") (Brendel et al., 2020). Other studies revealed that cute or empathic responses (Lv et al., 2021; Lv et al., 2022a, 2022b), expressions of gratitude or apology (Lv et al., 2022a, 2022b), or self-depreciating humor (Xu & Liu, 2022; Yang et al., 2023) lead to more positive consumer reactions. Moreover, messages highlighting the human-chatbot relationship (i.e., appreciation message) were found to be more effective to increase post-recovery satisfaction compared to apology-related messages (Song et al., 2023).

Table 1 Literature review—Chatbots service failure and recovery

Research streams	Source	Topic	Literature field(s)	Mediator(s), moderator(s)	Main findings
Conceptual/qualitative	Benner et al. (2021)	Literature review on chatbot failure recovery strategies	IS, HCI	-	Literature analysis presents six categories of chatbots' conversational breakdown recovery strategies
	Janssen et al. (2021)	Reasons of chatbot failures and critical factors for success	Design science, IS, HCI		Based on literature review and expert interviews, major reasons for chatbot failures and 12 critical success factors are identified
	Poser et al. (2021)	Design principles for handover from chatbot to human agent after failure	Design science, HCI	-	Experts confirmed positive perception of real-time hand-over to human agent when chatbots fail to resolve requests. The process of instant hand-over is proposed. Experts confirmed a positive perception of this hybrid approach for service recovery
Pre-failure interaction strategies (to mitigate failure severity perception)	Sands et al. (2022)	Effects of agent type (human vs. chatbot) related to failure type and magnitude	Customer service, IS	Mod: Service failure magnitude	Consumers respond less negative to a process failure when it occurs in human agent interaction (vs. chatbot as virtual agent). In case of a large outcome failure, consumers respond more positively when the agent is virtual (vs. human)
	Seeger and Heinzl (2021)	Anthropomorphic design to protect trust levels from damages due to service failure	IS, HCI	Med: Perceived norm violation	Anthropomorphic design of a chatbot mitigates negative effect of a failure on trust and Word-of-mouth
	Sheehan et al. (2020)	Effects of clarification technique on anthropomorphism and adoption intention	HCI, customer service	Med: Anthropomorphism Mod: Need for human interaction	Chatbot interactions using clarification technique are similarly effective regarding chatbot adoption compared to a perfectly understanding (i.e. error-free) chatbot, and both types are significantly more effective than a chatbot which failed to understand a user input. These effects are mediated by higher anthropomorphism compared to a failing chatbot
	Weiler et al. (2022)	Effects of ex-ante messages to reduce discontinuance after chatbot failures	Customer service	Mod: Performance level (low vs. high), Linguistic form (quantitative vs. qualitative)	Inoculation messages (i.e., preparing users for possible failures) reduce user's decision to discontinue chatbot after service failure. Quantitative form of the message moderated the effect of perceived performance on discontinuance, whereas a qualitative form moderated the effect of trusting beliefs

Table 1 (continued)

Research streams	Source	Topic	Literature field(s)	Mediator(s), moderator(s)	Main findings
Post-failure recovery strategies	Ashktorab et al. (2019)	User preferences of different chatbot failure repair strategies	HCI, Conversational design	-	User preferences of eight chatbot notification strategies after a chatbot response failure. Notifications which provide options or explanations are favored as they reflect a chatbot's initiative and are actionable to recover from failures
	Brendel et al. (2020)	Effects of three communication patterns on user's emotions and responses	IS, HCI	-	After a service failure, a "helping" communication style user leads to low levels of harassment, whereas both a "persecutor" and a "victim" style lead to significantly higher harassment. User frustration is not different across communication styles, but post-failure satisfaction is significantly higher for "helping" and "victim" (vs. "persecutor")
	Lv (Linxiang)et al. (2022a)	Effects of communication recovery strategies for AI device failures	HCI, Customer service	Med: Relational needs and efficacy needs	When chatbot message expresses gratitude (vs. apology), consumers are more likely to forgive a failure if service failures includes being ignored (vs. being ignored)
	Lv (Xingyang) et al. (2021)	Effects of cuteness after AI service failure	Design science, customer service	Med: tenderness and performance expectancy Mod: failure severity, time pressure	Cuteness increases consumers' tolerance of AI service failure, mediated by tenderness and performance expectancy
	Lv (Xingyang) et al. (2022b)	Effects of empathic response on users continuance intentions	HCI, Customer service	Med: Psychological distance, trust	Highly empathic AI response increases continuance intentions. Multisensory (i.e., text and audio) interactions amplify the effect, compared to text-only interactions
	Mozafari et al. (2022)	Effects of chatbot disclosure depending on service-related factors	Service	Med: Trust Mod: Service criticality, Failure vs. no failure	In case of service failure, chatbot identity disclosure leads to positive effect on retention. Without a failure, disclosure negatively affects customer retention through reduced trust for services with high levels of criticality
	Song et al. (2022)	Effect of chatbot self-recovery vs. human support on recovery satisfaction	IS, Customer service	Mod: Chatbot intelligence	Chatbot self-recovery has a more positive effect on post-recovery satisfaction compared to human-supported recovery. Effect is moderated by chatbot intelligence
	Song et al. (2023)	Effects of two politeness strategies on post-recovery satisfaction	Customer service	Med: Face concern, Mod: Time pressure	Focusing on human-chatbot relationship with an appreciation message led to more post-recovery satisfaction compared to an admittance of the limited abilities with an apology message
	Xing et al. (2022)	Effects of service failure types on consumers' recovery strategy choices	HCI, conversational design	Mod: Chatbot intelligence	After a functional (nonfunctional) failure, consumers tend to use the chatbot (a human) for service recovery. Chatbot intelligence moderates this choice. Chatbots (vs. human employees) have higher perceived governance
	Xu & Liu (2022)	Effects of humor in messages after AI service failure	HCI, Conversational design	Med: Perceived warmth and competence Mod: Time pressure, customer inoculation	Humorous response of chatbot after service failure increases consumer tolerance, mediated by warmth and competence
	Yang et al. (2023)	Impact of AI's self-deprecating humor on service recovery satisfaction	Customer service, conversational design	Med: Perceived intelligence, perceived sincerity Mod: Failure experience, Sense of power	Self-deprecating humor improves consumers' willingness to tolerate failure and recovery satisfaction, mediated by higher perceptions of intelligence and sincerity. Effect of humor was present in process-related failures, but not in outcome-related failures
	Our Study	Effectiveness of two failure-recovery message strategies depending on situational factors	Customer service, Conversational design	Med: Perceived warmth and competence	Using an empathy-seeking (solution-suggesting) message as recovery strategy increases post-recovery satisfaction and re-use intention, mediated by warmth (competence), respectively. Moreover, message preference depends on failure attribution (user vs. chatbot) and failure frequency

Chatbots and the stereotype content model

Following related studies about human–machine interactions (i.e., robots or chatbots), people quickly draw inferences about a bot's personality as interaction partner similarly as they would evaluate a human frontline employee (Belanche et al., 2021; Choi et al., 2021). For example, following the "computers are social actors" (CASA) paradigm (Nass et al., 1996), consumers are expected to evaluate a chatbot as digital interaction partner similarly as they would evaluate a human conversation partner—for instance by assessing its warmth and competence.

According to the stereotype content model (Fiske et al., 2007) as one of the most established frameworks regarding social cognitions, people use *warmth* and *competence* as two universal dimensions of social perception when judging others. Thereby, warmth covers aspects like honesty, kindness, or trustworthiness, while competence perceptions reflect capability, confidence, intelligence, and skillfulness (Dubois et al., 2016; Fiske et al., 2007; Judd et al., 2005). Taken together, these dimensions are suggested to "account almost entirely how people characterize others" (Fiske et al., 2007, p. 77). Originally, this system of social judgment was applied to explain perceptions of social groups (Fiske et al., 2007) or individuals (Judd et al., 2005). Since then, scholars have extended its use to brands (Aaker et al., 2010) and more recently to service interactions with humans (Scott et al., 2013) or non-human entities (i.e., robots or virtual agents) (Choi et al., 2021; Kull et al., 2021; Xu & Liu, 2022).

Judgments of warmth and competence influence how people interact with others, as well as how people feel and behave (Cuddy et al., 2008; Marinova et al., 2018). Warmth is generally linked to cooperative intentions and prosocial behavior, whereas competence is associated with the power and ability to realize one's goals (Cuddy et al., 2008). Inferred warmth and competence assessments enhance customer- and service-related outcomes such as satisfaction, trust, or brand admiration, and they influence downstream behaviors like purchase intentions and retention (Aaker et al., 2010; Cuddy et al., 2008; Marinova et al., 2018; Scott et al., 2013).

Recently, scholars have increasingly investigated the impact of social cognitions, i.e., warmth and competence perceptions, on various outcomes in the field of digital agents (Belanche et al., 2021; Choi et al., 2021; Kull et al., 2021; McKee et al., 2022; Xu & Liu, 2022). These studies mainly focus on anthropomorphism effects. For instance, Sungwoo Choi et al. (2021) found that people perceive humanoid (vs. nonhumanoid) service robots as warmer but not as more competent. In turn, higher warmth influences satisfaction after a failure and supports recovery effectiveness. In contrast, Belanche et al. (2021) revealed that both dimensions of warmth and competence indicate a robot's level of "humanness," and both dimensions positively influence customers'

loyalty. Warmth and competence perceptions are also found to influence human-digital agent collaboration. More precisely, perceptions of these social cognitions predict people's choice of a particular agent, irrespective of the agent's objective performance level (McKee et al., 2022). Moreover, Xu and Liu (2022) found that humorous chatbot answers increase consumers' tolerance after a service failure, mediated by higher warmth and competence. In general, Han et al. (2021) assessed that chatbot service failures trigger consumers' reactance (i.e., anger and negative cognitions). In turn, these negative cognitions reduce competence perceptions and subsequently decrease service quality and satisfaction. Finally, Kull et al. (2021) found that when chatbots use a warm (vs. competent) initial message, people's brand engagement increased, because they feel closer to the brand in that condition. Despite these initial insights, however, little is known about effects of message-related cues on respondents' warmth or competence evaluations and subsequent service assessments. This gap is relevant because many chatbots are text-based agents, and thus, users mainly have to rely on the chatbot's (text-based) messages as cues to, e.g., evaluate the chatbot's warmth and competence (van Pinxteren et al., 2020). Moreover, although chatbot service failures are common (Seeger & Heinzl, 2021), scholars confirm that there is still a lack of scientific knowledge about chatbot service recovery and its effectiveness (Xu & Liu, 2022). Therefore, this study evaluates how two distinct chatbot messages increase perceptions of social cognitions and enhance subsequent recovery responses.

Recovery strategies for chatbot failure

As chatbot response failures seem inevitable and lead to severe negative outcomes, firms are well advised to consider failure recovery strategies (Benner et al., 2021; Janssen et al., 2021). Thereby, a recovery strategy refers to an "effort [that] mitigates the previous negative effect of the failure" (Roschk & Gelbrich, 2014, p. 196). Scholars have revealed a wide range of such strategies as organizational responses, mainly with regard to service failures (for an overview, see van Vaerenbergh et al. (2019)). There are two basic dimensions of such failure recovery responses, namely (1) tangible compensation, such as monetary refunds, and (2) psychological compensations, including positive service employee behavior (Roschk & Gelbrich, 2014; van Vaerenbergh et al., 2019). (1) Tangible compensations mainly include financial and process-related efforts within a firm. A common approach to tangible compensation in chatbot failures is to hand over the conversation to a human employee to manage the problem and to prevent negative experiences (Ashktorab et al., 2019; Janssen et al., 2021). However, this solution comes with additional costs and reduces the level of automation (Reinkemeier & Gnewuch, 2022). In contrast, (2)

psychological compensations generally come without costs and could be executed by the service encounter agent (i.e., frontline employee or chatbot) directly. Prominent examples are apologies from the service employee or expressions of regret for the occurred failure (van Vaerenbergh et al., 2019). This research focuses on psychological compensations, as this is of interest for both research and management: Scientifically, this study complements initial research which evaluates effects of different message elements (such as expressions of humor, cuteness, apology, or gratitude) (Lv et al., 2022a, 2022b; Lv et al., 2021; Xu & Liu, 2022; Yang et al., 2023). Managerially, this type of compensation requires fewer resources (vs. human recovery) and can be integrated directly into the conversational process. In fact, a textual addition is all that is required to deliver these types of psychological compensation.

As gestures and nonverbal behaviors do not exist in chatbot conversations, people judge the chatbot conversation mainly based on written messages (van Pinxteren et al., 2020). We, therefore, analyze how different messages trigger social cognitions. As the study's outcome, post-recovery satisfaction and re-use intentions were chosen to evaluate recovery effectiveness. Post-recovery satisfaction represents one of the most widely used metrics to indicate successful recovery efforts (Song et al., 2022; Worsfold et al., 2007; Yang et al., 2023). Re-use intentions indicate continued acceptance of a chatbot and are relevant for its long-term success on digital platforms (Adam et al., 2021; Lv et al., 2022a, 2022b; Weiler et al., 2022).

Different messages as failure recovery strategies

Even small changes in the framing of communication messages were found to influence people's judgments and behaviors (You et al., 2020). Regarding chatbot conversations, different messages could be used in response to a service failure. In this research, two distinct message types labeled as an empathy-seeking message or solution-oriented message were deliberately chosen as they (1) represent the most common failure recovery strategies as revealed by our pilot study (see below in the empirical studies section) and (2) are thought to influence warmth and competence perceptions, respectively. Both types express a request from the chatbot. As first type, a chatbot might ask for a user's empathy and understanding regarding its limited abilities. This request for understanding is sought to elicit empathic concern for the chatbot's "infancy" and difficulties in handling requests. Scholars also refer to this message as "social" recovery strategy, which reflects apologizing for the failure "to appeal to the users' empathy and understanding similar to that which is shown in human–human conversations" (Benner et al., 2021, p. 9).

Empathy is defined as a person's intellectual or imaginative understanding of another person's condition or state (Hogan, 1969). Related to service, empathic customers were found to be less angry and more forgiving when they encounter a service failure (Wieseke et al., 2012). In a study with "classic" human frontline employees, customer empathy towards an employee was found to enhance social interactions, foster supportive attitudes, and create a more satisfying experience (Lazarus, 1991; Wieseke et al., 2012). Scholars in the field of social service research support this, showing that empathy-related expressions are often beneficial to build or strengthen social bonds between interaction partners (Gerdes, 2011), which in turn increase warmth perceptions (Cuddy et al., 2008; Judd et al., 2005).

These well-documented effects could also be observed in human interactions with digital agents. As a chatbot reflects a digital version of a service employee, a chatbot message that evokes empathy (e.g., asking for patience and to hold on to the joint interaction) should trigger these warmth perceptions. Scholars consistently demonstrated that humans can feel empathy with inanimate objects such as chatbots or robots (Misselhorn, 2009). Related to the adjacent field of service robots, Wirtz et al. (2018) concludes that a bot's social-emotional and relational elements (e.g., social interactivity) increase warmth. As the chatbot's empathy message contains mainly such social-emotional and relational elements (e.g., asking for patience and to hold on to the joint interaction), we propose:

H1. The message type empathy increases consumer-perceived chatbot warmth.

As an alternative option, a chatbot could request the user to adapt the input to the chatbot's abilities, e.g., by rephrasing the input in short and simple words. Input from users was often found to be complex, and a shorter and more precise input has a higher probability of being processed correctly (Ashktorab et al., 2019; Luger & Sellen, 2016). Indeed, conversational agents were found to respond more successfully when the input was rather simple, short, and unambiguous (Luger & Sellen, 2016). This type of request could be labeled as a solution-oriented message, as the chatbot tries to solve the failure actively. Related IS research has already used the solution-oriented message (i.e., "please rephrase your inquiry and try again") to encourage users to continue with the chatbot (Benner et al., 2021; Weiler et al., 2022). While Weiler et al. (2022) use this message as ex-ante strategy at the beginning of the interaction, this study employs it as ex-post strategy to address the chatbot response failure directly when it occurred.

This concept has also been observed in human service interactions. When a frontline employee focuses on the task (vs. social components) as the "core" of the service delivery and offers a possible solution to make the interaction more successful and convenient, this task-related behavior increases the perceived competence of this employee (Marinova et al.,

2018). Several scholars support this argumentation, and acknowledge that competence-oriented messages imply that service providers are "very capable in providing consumers with solutions" (Huang & Ha, 2020, p. 620).

Related to the chatbot, the message-type solution focuses on the task, that is, to make the interaction with the customer effective. As consumers perceive digital assistants such as chatbots as social actors (Nass et al., 1996; van Pinxteren et al., 2020), this solution-oriented message should increase chatbot competence perceptions (Marinova et al., 2018). Moreover, the solution message indicates that the chatbot is aware of the linguistic complexity of user input and of options to improve the quality of the chatbot's answer (Weiler et al., 2022). Both aspects (i.e., awareness of a problem, and presentation of a possible solution) indicate a kind of skillfulness or intelligence, two key items reflecting competence (Cuddy et al., 2008; Xu & Liu, 2022). In addition, related service robot literature proposed that when a bot can serve a user's functional needs (e.g., offering a solution to a request), this service enhances perceptions of its usefulness and competence (Wirtz et al., 2018). Therefore, we hypothesize:

H2. The message type solution increases consumer-perceived chatbot competence.

According to scholars, warmth and competence perceptions can serve as underlying mechanisms that explain how consumers respond to technology infusion in service (Belanche et al., 2021; van Doorn et al., 2016). According to van Doorn et al. (2016), warmth and competence perceptions elicited by digital service technology both enhance consumers' satisfaction and loyalty intentions. A chatbot-study found that if chatbots could elicit warmth perceptions within human-chatbot interactions, chatbot use is rising (Mozafari et al., 2021). Supporting this, research from the related field of service robots found that warmth perceptions significantly increased post-failure satisfaction and loyalty (Choi et al., 2021). Similarly, research confirmed that consumers' competence perceptions (e.g., the belief that chatbots are capable to fulfill a task or enable successful service recovery) increase their interaction satisfaction and re-use intentions (Lv et al., 2022a, 2022b; Mozafari et al., 2022). Further studies about human (Babbar & Koufteros, 2008; Güntürkün et al., 2020; Habel et al., 2017) and digital service agents (Belanche et al., 2021; Han et al., 2021) support that higher warmth and competence perceptions drive consumers' service value perceptions, satisfaction, and loyalty. Thus:

H3. Stronger consumer-perceived (a) warmth and (b) competence increase consumers' post-recovery satisfaction.

H4. Stronger consumer-perceived (a) warmth and (b) competence increase consumers' chatbot re-use intentions.

Factors influencing the perception of recovery messages

Research has shown that situational factors regarding chatbot interactions influence user perceptions and responses (Gnewuch et al., 2017; Janssen et al., 2021; Pizzi et al., 2021). Therefore, we identified two relevant factors, namely *failure attribution* and *failure frequency*, which are thought to impact users' reactions and preference for one of the recovery messages. Both factors were found to be important elements in the failure and recovery literature (Choi & Mattila, 2008; Collier et al., 2017; Ozgen & Duman Kurt, 2012; van Vaerenbergh et al., 2019).

Failure frequency In chatbot conversations, users regularly need to make multiple attempts to enter a request in a way that the chatbot will understand (Ashktorab et al., 2019). That means that many initial service failures are not recovered adequately but lead to a second service failure—a situation also labeled as *double deviation* (Johnston & Fern, 1999; van Vaerenbergh et al., 2019). Such double deviations were found to reinforce negative customer responses that were caused by the first failure, such as customer dissatisfaction, anger, or churn (Ozgen & Duman Kurt, 2012; van Vaerenbergh et al., 2019). Furthermore, people were found to prefer different recovery strategies for a single vs. double deviation, leading to the conclusion that the service provider should adequately account for the failure frequency in choosing the appropriate recovery strategy (Pacheco et al., 2019). Therefore, chatbot creators need to identify the best-possible "match" for the response to the failure (Roschk & Gelbrich, 2014). After a first failure, both response messages are expected to mitigate negative consequences via the paths of warmth and competence as proposed above. Yet, when users re-enter their request and the chatbot fails again to deliver an appropriate answer, this represents a new situation with (potential) implications for the effectiveness of both message types after the second failure.

The empathy-related message seeks to evoke understanding and empathy and create feelings of warmth and mutual connection (Cuddy et al., 2008; Lazarus, 1991; Wieseke et al., 2012). Asking for understanding regarding the chatbot's limited abilities is possible at any interaction stage or situation, as the chatbot refers to its own lack of abilities (vs. the user). Therefore, an empathy message is assumed to create warmth perceptions irrespective of the failure frequency. Related to the message type solution, as argued above, people are expected to accept the request to re-phrase
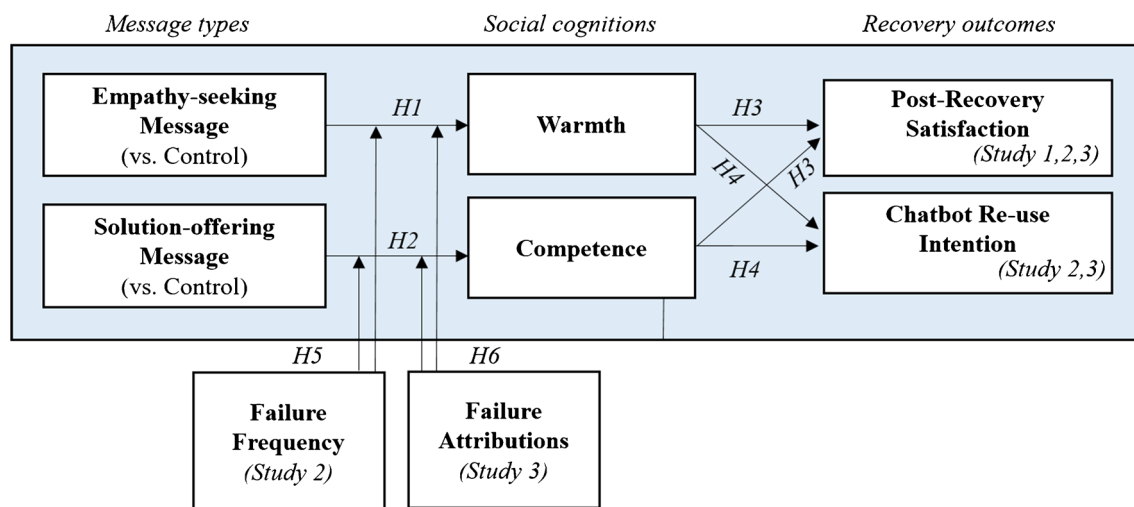
Fig. 1 Overview of research studies

their input to better adhere to a chatbot's needs after a first failure and even perceive that chatbot as competent (Chong et al., 2021; Marinova et al., 2018). However, after re-phrasing the request and being confronted with a second service failure, this competence perception is assumed to be negatively affected as the chatbot was again not able to provide a solution. As Johnston and Fern's (1999) study showed, more than half of the respondents lost confidence in a service agent's competence after a double deviation. Taken together, after a double deviation, empathy-seeking message should be more effective than solution-oriented messages. Formally:

> H5. After a double deviation, an empathy message is more effective than a solution message in that the effect of the empathy message on consumer-perceived chatbot warmth is stronger than the effect of the solution message on consumer-perceived chatbot competence.

Failure attribution Following attribution theory (Weiner, 1985, 2012), particularly in its application to service failures, customers seek to attribute the responsibility for the occurrence of a negative incident to some person or thing as a way to understand the situation and regain control over their environment. Thereby, people mainly differentiate between two dimensions of a so-called locus of control—either they blame others (i.e., external attribution) or they blame themselves (i.e., internal attribution) for the failure that has occurred (Weiner, 1985). Previous research showed that customers respond differently to service failures depending on which party they believe to be responsible for the failure (Choi & Mattila, 2008; Collier et al., 2017). For instance, when people assign the firm or its service agent

as responsible for the failure, people react more negatively than when they perceive that they are (at least partially) responsible for the failure as well (Choi & Mattila, 2008). Consequently, people which respond more positively to service failures that are self-attributed (versus firm-attributed) remain more satisfied with the firm and are more likely to forgive such failures (Choi & Mattila, 2008; Gelbrich, 2010).

When considering which chatbot recovery message should be employed (i.e., solution or empathy), failure attributions are supposed to differentiate its effectiveness. More precisely, we expect that the failure attribution and the recovery message should match the failure type to create positive outcomes. Recovery research has shown that matching the recovery strategy with the failure type (e.g., monetary compensation for monetary failure) is more effective than a non-match (Roschk & Gelbrich, 2014). Related to chatbot interaction, when users attribute the failure to the chatbot (i.e., blame it for the failure), an empathy (vs. solution) message should be a better match, as in that case the attributed party "takes the blame" by asking for empathy and understanding. Scholars have established that such messages send cues that clarify and acknowledge blame attributions, and they help users to understand the possible reason for the failure (e.g., the "infancy" of the chatbot). In turn, these cues work as a coping mechanism to handle the negative consumer reactions caused by the failed service (Gelbrich, 2010). In line with that, an empathy message as response to a chatbot-caused failure is supposed to match, while the solution message expresses that the user is also part of the failure—a message cue which does not match the responsibility perception of the user.

Vice versa, the solution message matches a user-attributed failure because it offers guidance for the user to tailor the request to the chatbot. When a user acknowledges to

be (at least partly) responsible for the failure or is unsure about who to blame, a solution message (vs. empathy) should better match this perception. To put it differently, users are supposed to accept a request to rephrase their entry when they admit to be part of the problem (Choi & Mattila, 2008), and they might even be thankful for guidance on how to react in the interrupted process. Yet, when a chatbot is believed to be the responsible party, a solution message that expresses a user action to resolve the situation is expected to be perceived as less appropriate, and should therefore affect consumers' competence perceptions to a smaller extent. Thus:

> H6a. An empathy message leads to higher consumer-perceived warmth in the case of a chatbot-attributed failure (match) than in the case of a user-attributed failure (mismatch).
>
> H6b. A solution message leads to higher consumer-perceived competence in the case of a user-attributed failure (match) than in the case of a chatbot-attributed failure (mismatch).

Empirical studies

Pilot study

As initial pilot study, chatbots from different companies and across industries in the DACH-region (Germany, Austria, and Switzerland) were analyzed to assess which recovery strategy they used after a service failure. A service failure reflects that a chatbot did not understand the user's request and was provoked by entering some random letters as incomprehensible input. The final sample resulted in 101 chatbots from business, education, and public administration. Almost a third of these bots (i.e., 27) did not allow any free-text entry but only a set of options to choose, and consequently no "failure" in communication could occur when engaging with them. Out of the remaining 74 (free-text processing) chatbots, 34 ask the user to reformulate their request, reflecting the solution message type. Users were asked to use short sentences, simple words, and to be as precise as possible in their wording. Furthermore, 12 chatbots appealed to the user's empathy and understanding. Lastly, no clear strategy was identified for 28 chatbots, and most of these chatbots just replied with a simple error feedback message. That means the chatbot just sends short messages like "Sorry I did not understand that."

In sum, the pilot study revealed that four major message-based recovery strategies are prominent in chatbot conversations: (1) pre-defined answers, (2) a solution-oriented message, (3) an empathy-seeking message, and (4) a simple error feedback message. As pre-defined answers limit the variety of

entries, they are generally less flexible. Therefore, this message type was omitted and the latter three types were analyzed.

Study 1

Study design To investigate the influence of type of the recovery messages on users' post-recovery satisfaction, Study 1 applies a one-factorial between-subjects experiment with three cases (message type: empathy vs. solution vs. control) (Fig. 1). Participants were recruited from two European universities through email distribution lists and randomly assigned to one of the scenarios (see Fig. 2 for detailed scenarios). After excluding four participants who failed in the attention check (i.e., "If you read this, please press button 1"), our sample resulted in 178 participants (M_{Age}: 24 years, SD: 18.34, 56.2% females). Participants had to imagine that they interact with a chatbot of an electronics provider, as electronic retail and service offers are nowadays vastly provided by digital platforms and electronic markets, and prior research considered e-commerce as prevalent field of chatbot service (Adam et al., 2021; Alt, 2020; Gnewuch et al., 2017).

As for the conversation, three questions about a camera were asked; two of which the chatbot answers correctly and the last one where the chatbot mentioned a non-understanding of the user request (i.e., response failure, see Table 2). As a manipulation, we varied the failure responses: The chatbot either asked the user to have empathy with its limited abilities and to try again (i.e., type empathy) or to adapt and simplify the input (i.e., type solution). As control case, the chatbot just replied, "Sorry, I did not understand your request." As manipulation checks, we relied to Hosseini and Caragea (2021) as they described empathy-seeking behavior: People in the empathy message scenario perceived more strongly that the chatbot had "asked for their empathy and understanding" ($M_{Empathy}$: 6.31, $M_{Solution}$: 2.88, $M_{Control}$: 3.02; $F = 90.14$, $p < 0.001$) compared to the other scenarios. Likewise, for the solution case, we relied on Marinova et al. (2018) to describe problem-solving behavior: Respondents of the solution message perceived more strongly that the chatbot "has asked to rephrase my request" ($M_{Empathy}$: 2.53, $M_{Solution}$: 5.98, $M_{Control}$: 2.32, $F_{(2175)} = 94.94$, $p < 0.001$). Thus, the manipulation was effective. Moreover, the scenarios were perceived as realistic (i.e., "The scenario is realistic" and "I can imagine a chatbot interaction happening like this in real life.") (α: 0.89, M: 5.71, SD: 1.41 (on a 7-point Likert scale)).

Measures For all three studies, reflective multi-item measures with 7-point Likert scales (1 = strongly disagree and 7 = strongly agree) from the extant literature were used and adapted to the study context. Post-recovery satisfaction was captured with three items from Agustin and Singh (2005). Perceived competence and warmth of the chatbot were captured by three-item scales each from Aaker et al. (2010), followed by some demographics (i.e., age, gender). Reliability and validity values were all above the thresholds (see Table 3). Cronbach's alpha and composite reliability values are above the cut-off value of 0.70, indicating

construct-level reliability (Hulland et al., 2018). Second, the average variance extracted (AVE) for every multiple-item construct exceeded 0.50, showing appropriate convergent validity. Third, the AVE values were found to be larger than the shared variance of any other remaining construct, indicating discriminant validity (Hulland et al., 2018). All items and factor loadings are illustrated in Table 3, and means and standard deviations for the main variables are provided in Table 4.

Results An ANOVA revealed significant effects of the three message types on post-recovery satisfaction $(F(2,175) = 15.97, p < 0.001)$. Post-hoc tests (Bonferroni) showed that both the empathy message and the solution message led to significantly higher post-recovery satisfaction than the control message $(M_{Solution}: 2.93$ vs. $M_{Control}: 1.79$, $p < 0.001; M_{Empathy}: 2.60$ vs. $M_{Control}: 1.79, p < 0.001)$. In contrast, the empathy and solution messages did not lead to significantly different satisfaction $(p = 0.36)$. Thus, both messages enhance satisfaction compared to control—but not to a different degree when compared to each other.

To test H1 to H3, a mediation analysis was conducted with PROCESS Model 4 using 5,000 bootstrapping samples and 95% confidence intervals (CIs) (Hayes, 2018). The message types were used as a multicategorical independent variable, warmth and competence served as parallel mediators, satisfaction was the outcome, and age and gender were covariates.

As hypothesized, the empathy message (vs. control) increased warmth $(b = 1.01, p < 0.001)$, and the solution message (vs. control) led to higher competence perceptions $(b = 0.88, p < 0.001)$, supporting H1 and H2, respectively. The empathy message (vs. control) did not increase competence perceptions $(p = 0.94)$, while the solution message also increased warmth $(b = 0.54, p < 0.05)$. In turn, both warmth $(b = 0.12, p < 0.05)$ and competence $(b = 0.42, p < 0.001)$ had a positive effect on satisfaction, supporting H3 (a and b). The indirect effects of the empathy message on satisfaction were significant via warmth $(b = 0.12, [0.01, 0.27])$, and they were significant for the solution message on satisfaction via competence $(b = 0.37, [0.17, 0.60])$.[1]

---

[1] We also conducted a study (students from two European universities, $n = 270$, $M_{Age} = 27$ years, 52% female) with the same measures based on a further scenario (i.e., a chatbot as pizza delivery agent as food delivery represents another common field for digital platforms (e.g., Uber eats, Deliveroo, HelloFresh) and for chatbot services (Li et al. (2020); van Pinxteren et al. (2020))). Results of a mediation analysis (PROCESS model 4) showed that the empathy message (vs. control) led again to higher warmth perceptions $(b = 1.16, p < 0.001)$ while the solution message (vs. control) did not $(p = 0.45)$. The solution message (vs. control) led to higher competence perceptions $(b = 0.56, p < 0.05)$, whereas the empathy message (vs. control) did not $(p = 0.56)$. In turn, satisfaction was influenced by warmth $(b = 0.30, p < 0.001)$ and competence $(b = 0.48, p < 0.001)$. Neither message influenced satisfaction directly. In sum, the results also provide support for H1–H3 again and add further validity to Study 1.

Study 2—Failure frequency

Design and procedure Study 2 examined the effect of the recovery messages on post-recovery satisfaction and re-use intentions under different failure recovery conditions (i.e., success vs. second failure after the recovery). Two hundred fifty-eight respondents were recruited via the online platform Prolific (US participants with 95% former tasks approval ratio). Participants were randomly assigned to a 3 (message type: empathy vs. solution vs. control) × 2 (recovery outcome: success vs. second failure) between-subjects experiment and had to imagine a chatbot interaction for a table reservation in a restaurant (see Table 2). The chatbot did not understand the initial user request and responded with one of the three message types from Study 1. After reading the recovery message, respondents had to rate their warmth, competence, anger, and satisfaction and enter an individual input as response. On the next page, the survey tool illustrates the interaction including the individual user input and adds either a success message (i.e., "I successfully booked a table") or a second failure message. In case of the second failure, one of the three message types (i.e., empathy-seeking, solution-oriented, control) was displayed (again); with a slightly adapted text for the solution-message to fit the context. After these messages, respondents again rated their perceptions (i.e., warmth, competence, satisfaction, re-use intentions, anger). On average, respondents needed 8 min to complete the survey. To increase realism and the fit of user-entry and message, we excluded fourteen participants in the recovery success condition who entered nonsensical input. Furthermore, we excluded seven participants who failed the attention check (i.e., participants who agreed to the false statement "the chatbot has forwarded me to a human service employee"); the final sample consisted of 237 respondents $(M_{Age}: 45$ years, $SD_{Age}: 14.56, 49\%$ female).

Measures Scales were identical to those used in Study 1. Chatbot re-use intentions were measured with the scale from Wallenburg and Lukassen (2011). As control variable, we assessed participant's anger with three items from Xie et al. (2015), as this emotional response could influence user reactions in chatbot interactions (Crolic et al., 2021). All scales displayed adequate validity and reliability (see Table 3). Moreover, scenarios were perceived as realistic $(M: 5.41, SD: 1.45)$ and the manipulation checks were effective. The empathy message was perceived as stronger for seeking empathy and understanding $(M_{Solution}: 3.26, M_{Empathy}: 5.70, M_{Control}: 2.25, F(2234) = 71.16, p < 0.001)$, and respondents in the solution message scenario agreed

Fig. 2 Exemplary scenarios

Solution message   Empathy message
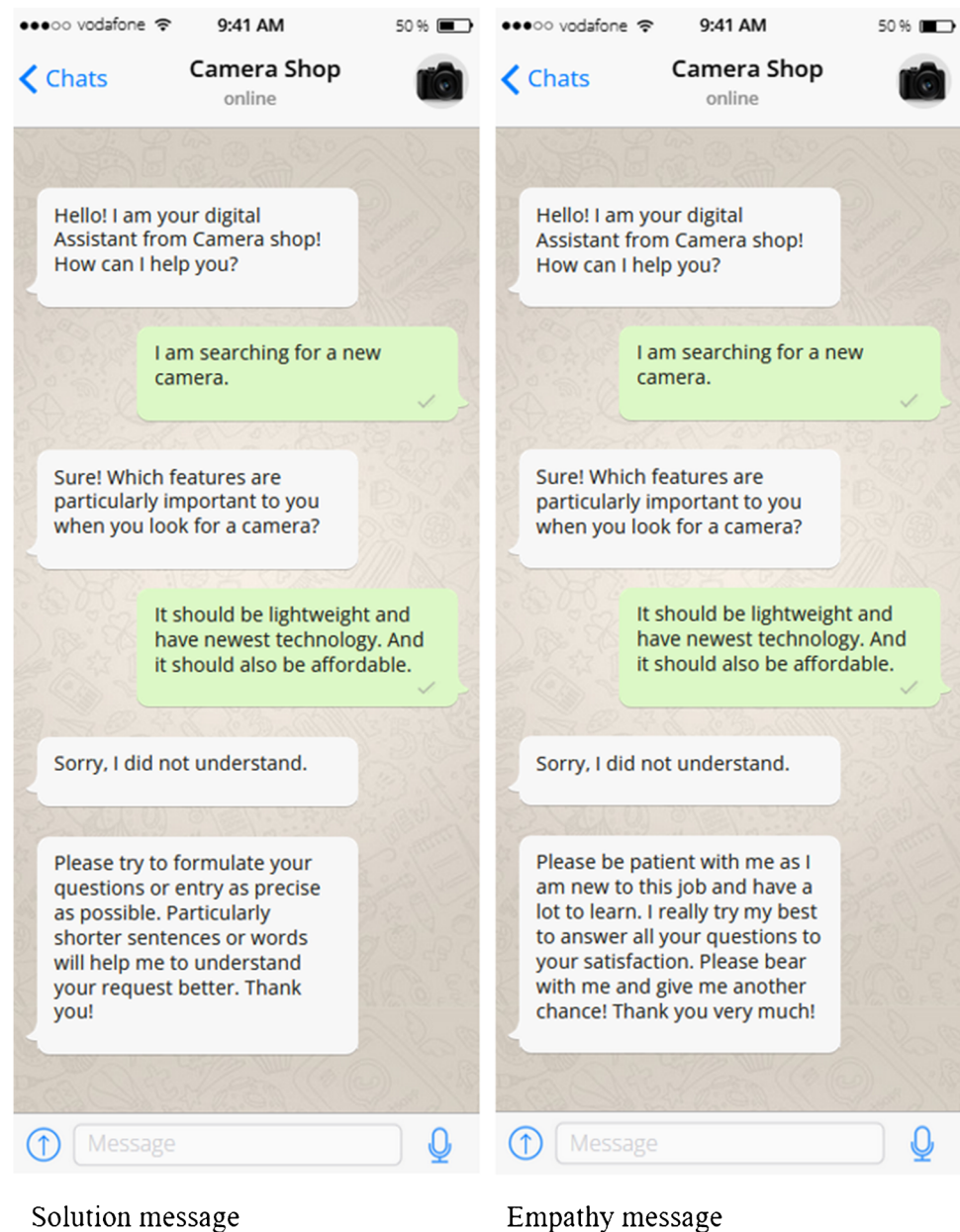
more that the chatbot has asked to rephrase the input as possible solution ($M_{Solution}$: 6.62, $M_{Empathy}$: 2.45, $M_{Control}$: 2.56, $F(2234) = 155.62$, $p < 0.001$). Regarding the recovery success manipulation, participants in the success scenarios (vs. second failure) rated significantly stronger that the chatbot "has successfully reserved a table" ($M_{Success}$: 6.67, $M_{Second-Failure}$: 1.28, $t(235) = 50.82$, $p < 0.001$). Moreover, Table 4 provides descriptives for the main variables.

Results To test H1 to H3 in one comprehensive model, we again conducted a mediation analysis (PROCESS Model 4, Hayes (2018) with 5000 bootstrap samples and 95% CIs) with the same setup as in Study 1. Anger, age, and gender were added as covariates. Consumer perceptions were evaluated after the first failure. The empathy message (vs. control) increased perceived warmth ($b = 1.22$, $p < 0.001$), and the solution message (vs. control) led to higher competence perceptions ($b = 0.49$, $p < 0.05$), supporting H1 and H2. The solution message also increased perceived warmth ($b = 0.50$, $p < 0.05$), whereas the empathy message did not increase competence ($p = 0.84$). Satisfaction was influenced by warmth ($b = 0.26$, $p < 0.001$) and competence ($b = 0.51$, $p < 0.001$), supporting H3. The indirect effect of the empathy message on satisfaction via warmth was significant ($b = 0.31$, [0.15, 0.51]) and the indirect effect of the solution message on satisfaction via competence was significant ($b = 0.25$, [0.02, 0.49]). In our analysis, neither of the two message types had a direct impact on satisfaction.

Table 2 Scenario descriptions

Solution message scenarios	Empathy message scenarios	Control
Study 1 (*Electronics provider*)	***Study 1*** (*Electronics provider*)	***Study 1*** (*Electronics provider*)
[Chatbot]: Hello! I am your digital Assistant from [Camera shop]! How can I help you?	[Chatbot]: Hello! I am your digital Assistant from [Camera shop]! How can I help you?	[Chatbot]: Hello! I am your digital Assistant from [Camera shop]! How can I help you?
[User:] I am searching for a new camera	[User:] I am searching for a new camera	[User:] I am searching for a new camera
[Chatbot]: Sure! Which features are particularly important to you when you look for a camera?	[Chatbot]: Sure! Which features are particularly important to you when you look for a camera?	[Chatbot]: Sure! Which features are particularly important to you when you look for a camera?
[User:] It should be lightweight and have newest technology. And it should also be affordable	[User:] It should be lightweight and have newest technology. And it should also be affordable	[User:] It should be lightweight and have newest technology. And it should also be affordable
[Chatbot]: Sorry, I did not understand	[Chatbot]: Sorry, I did not understand	[Chatbot]: Sorry, I did not understand
[Chatbot]: Please try to formulate your questions or entry as precise as possible. Particularly shorter sentences or words will help me to understand your request better. Thank you!	[Chatbot]: Please be patient with me as I am new to this job and have a lot to learn. I really try my best to answer all your questions to your satisfaction. Please bear with me and give me another chance! Thank you very much!	
Study 2 (*Table reservation*)	***Study 2*** (*Table reservation*)	***Study 2*** (*Table reservation*)
[Chatbot]: Hello! I am your digital Assistant from Pizza House! How can I help you?	[Chatbot]: Hello! I am your digital Assistant from Pizza House! How can I help you?	[Chatbot]: Hello! I am your digital Assistant from Pizza House! How can I help you?
[User:] I would like to reserve a table for next week Tuesday between 12 and 13 h for me and my partner	[User:] I would like to reserve a table for next week Tuesday between 12 and 13 h for me and my partner	[User:] I would like to reserve a table for next week Tuesday between 12 and 13 h for me and my partner
[Chatbot]: Sorry, I did not understand	[Chatbot]: Sorry, I did not understand	[Chatbot]: Sorry, I did not understand
[Chatbot]: Please try to formulate your questions or entry as precise as possible. Particularly shorter sentences or precise words will help me to understand your request better. Thank you!	[Chatbot]: Please be patient with me as I am new to this job and have a lot to learn. I really try my best to answer all your questions to your satisfaction. Please bear with me and give me another chance. Thank you very much!	[User:] [individual entry]
[User:] [individual entry]	[User:] [individual entry]	[Chatbot]: Sorry, I did not understand you again
[Chatbot]: Sorry, I did not understand you again. Please formulate your entry as precise as possible. For example, please enter only the date, arrival time and the number of people for the reservation. [2nd failure] / Thank you! I booked a table for Tuesday for flexible arrival between 12:00 and 13:00 for 2 people. Your booking code is 2553. See you soon! [Recovery success]	[Chatbot]: Sorry, I did not understand you again Please be patient with me again as I am new to this job and have a lot to learn. I really try my best to answer all your questions to your satisfaction. Please bear with me and give me another chance. Thank you very much! [2nd failure] / Thank you! I booked a table for Tuesday for flexible arrival between 12:00 and 13:00 for 2 people. Your booking code is 2553. See you soon! [Recovery success]	[2nd failure] / Thank you! I booked a table for Tuesday for flexible arrival between 12:00 and 13:00 for 2 people. Your booking code is 2553. See you soon! [Recovery success]
Study 3 (*Pizza delivery*)	***Study 3*** (*Pizza delivery*)	***Study 3*** (*Pizza delivery*)
[Chatbot]: Hello! I am your digital Assistant from Pizza House! How can I help you?	[Chatbot]: Hello! I am your digital Assistant from Pizza House! How can I help you?	[Chatbot]: Hello! I am your digital Assistant from Pizza House! How can I help you?
[User:] I would like to order a Pizza Salami to my home	[User:] I would like to order a Pizza Salami to my home	[User:] I would like to order a Pizza Salami to my home
[Chatbot]: Sure! To which address may I deliver your Pizza?	[Chatbot]: Sure! To which address may I deliver your Pizza?	[Chatbot]: Sure! To which address may I deliver your Pizza?
[User:] To Schlösschen Street 12, please [Chatbot-attributed failure] / "To my home please" [User-attributed failure]	[User:] To Schlösschen Street 12, please [Chatbot-attributed failure] / "To my home please" [User-attributed failure]	[User:] To Schlösschen Street 12, please [Chatbot-attributed failure] / "To my home please" [User-attributed failure]
[Chatbot]: Sorry, I did not understand. I do not know the address "To Schlösschen Street 12 please" [Chatbot-attributed failure] / "To my home please" [User-attributed failure]	[Chatbot]: Sorry, I did not understand. I do not know the address "To Schlösschen Street 12 please" [Chatbot-attributed failure] / "To my home please" [User-attributed failure]	[Chatbot]: Sorry, I did not understand
[Chatbot]: Please try to formulate your questions as precise as possible. Particularly shorter sentences or words will help me to understand your request better. Thank you!	[Chatbot]: Please be patient with me as I am new to this job and have a lot to learn. I really try my best to answer all your questions to your satisfaction. Please bear with me and give me another chance! Thank you very much!	

Table 3 Scale items and statistics

Construct name and items	Factor loading		
	Study 1	Study 2a/b	Study 3
Warmth (Study 1/2a and b/3: α = .80/.95 and .96/.87; CR = .87/.92 and .93/.91; AVE = .69/.80 and .81/.76)			
I perceive the chatbot as …			
• warm	.80	.90/.91	.84
• kind	.84	.91/.92	.88
• generous	.85	.87/.87	.91
Competence (Study 1/2a and b/3: α = .87/.95 and .98/.90; CR = .88/.90 and .89/.88; AVE = .71/.75 and .72/.71)			
I perceive the chatbot as …			
• competent	.77	.84/.86	.77
• effective	.87	.88/.84	.90
• efficient	.89	.88/.85	.86
Post-recovery satisfaction (Study 1/2a and b/3: α = .74/.93 and .98/.90; CR = .80/.83 and .86/.82; AVE = .58/.62 and .67/.60)			
The interaction with the chatbot service was …			
• satisfying	.79	.74/.84	.73
• pleasant	.64	.81/.79	.81
• good	.83	.81/.83	.79
Re-use intentions (Study 2b/3: α = .96/.96, CR = .85/.87; AVE = .65/.70)			
• I would use this chatbot again		- /.79	.79
• I would use this chat service in my daily life		- /.83	.89
• I would order my pizza again with this chatbot		- /.80	.83

Study 2a: values after first failure; 2b: values after second response

Next, to examine effects of responses to the second failure (H3, H4, and H5), we used a moderated mediation analysis (PROCESS Model 8, Hayes (2018) with 5000 bootstrap samples and 95% CIs) and compared the different messages after the second response of the chatbot. The response condition (i.e., second failure vs. successful chatbot answer) was used as moderator, and anger, age, and gender were covariates again. Related to the effects of the mediators on the dependent variables (i.e., H3, H4), results of the mediation model with satisfaction showed that warmth and competence significantly increased post-recovery satisfaction ($b_{warmth} = 0.16$, $p < 0.001$; $b_{competence} = 0.72$, $p < 0.001$), supporting H3. Similarly, when using re-use intentions as dependent variable, warmth and competence significantly increased re-use intentions ($b_{warmth} = 0.24$, $p < 0.005$; $b_{competence} = 0.50$, $p < 0.001$), supporting H4.

Results of the messages after the second failure on the mediators (H5) show that the empathy message still led to perceived warmth ($b = 0.94$, $p < 0.01$), whereas the solution message did not lead to higher competence perceptions ($p = 0.21$). Thus, H5 could be supported. Correspondingly, the indirect effect of the empathy message on satisfaction via warmth was significant ($b = 0.15$, [0.04;

0.31]), whereas the indirect effect of the solution message on satisfaction via competence was not significant ($b = 0.25$, [− 0.19; 0.70]). Similarly, the indirect effect of the empathy message on re-use intentions via warmth was significant ($b = 0.23$, [0.05; 0.47]), whereas the indirect effect of the solution message on re-use intentions via competence was not significant ($b = 0.15$, [− 0.11; 0.44]). Thus, as hypothesized, the empathy message was found to be more effective than the solution message after the second failure.

In case of a successfully resolved second attempt, the indirect effect of the empathy message on satisfaction via warmth was not significant ($b = 0.11$, [− 0.004; 0.28]). Similarly, the indirect effect of solution on satisfaction via competence was not significant ($b = 0.23$, [− 0.15; 0.60]). Thus, message effects dissolve when the chatbot solved the user's request.

Study 3—Failure attributions

Design and procedure Study 3 aimed to examine the effect of the recovery messages on post-recovery satisfaction and re-use intentions under different failure attribution conditions, i.e., either chatbot or user was responsible for the failure.

Table 4 Descriptives for studies 1, 2, and 3

		Dependent variables			
		Warmth	Competence	Post-recovery satisfaction	Re-use intentions
Study 1					
	Empathy	4.40 (1.43)	2.66 (1.27)	2.11 (1.03)	-
	Solution	3.87 (1.30)	3.19 (1.29)	2.50 (1.19)	-
	Control	3.53 (1.33)	2.66 (1.35)	1.78 (0.79)	-
Study 2					
First failure					
	Empathy	4.35 (1.51)	2.88 (1.61)	3.08 (1.48)	-
	Solution	3.64 (1.49)	3.51 (1.52)	3.18 (1.56)	-
	Control	3.04 (1.61)	2.78 (1.52)	2.41 (1.39)	-
Second failure					
	Empathy	3.59 (1.64)	1.80 (1.11)	1.95 (1.14)	1.86 (1.25)
	Solution	3.28 (1.54)	2.66 (1.75)	2.66 (1.79)	2.23 (1.87)
	Control	2.57 (1.49)	1.90 (1.32)	1.77 (1.18)	2.17 (1.34)
Success					
	Empathy	4.67 (1.73)	5.17 (1.50)	5.10 (1.65)	4.63 (1.81)
	Solution	3.69 (1.55)	5.02 (1.03)	4.64 (1.22)	4.81 (1.36)
	Control	3.85 (1.84)	4.76 (1.61)	4.81 (1.61)	4.68 (1.69)
Study 3					
User-attributed failure					
	Empathy	4.63 (1.41)	3.66 (1.47)	4.49 (1.62)	4.37 (1.89)
	Solution	3.70 (1.16)	4.54 (1.49)	4.31 (1.53)	4.13 (1.70)
	Control	3.48 (1.36)	3.79 (1.34)	4.27 (1.65)	4.02 (1.78)
Chatbot-attributed failure					
	Empathy	4.28 (1.30)	3.21 (1.15)	3.63 (1.22)	3.33 (1.70)
	Solution	3.53 (1.53)	3.49 (1.72)	3.15 (1.39)	3.46 (1.99)
	Control	3.06 (1.35)	2.75 (1.67)	2.72 (1.22)	2.38 (1.46)

Numbers represent means (standard deviations)

Respondents from a German university were recruited via E-Mail distribution lists and randomly assigned to a 3 (message type: empathy vs. solution vs. control) × 2 (user fault vs. chatbot fault) between-subjects experiment. After excluding eight participants who failed the attention check (i.e., if you read this, please press button 1), the final sample consisted of 249 respondents (M_{Age}: 27 years, SD_{Age}: 14.24, 63% female). As scenario, a pizza delivery case was used (see Table 2), as this case represents another common field for digital platforms (e.g., Uber eats, Deliveroo, HelloFresh) and for chatbot services (Li et al., 2020; van Pinxteren et al., 2020). As user-fault scenario, the user entered "to my home" as the delivery address, which obviously could not be found in a database. As chatbot-fault scenario, the user entered an address "to Schlösschen Street 12," which a chatbot would be supposed to find in a location database. Recovery messages were taken from Study 1 and slightly adapted to fit the failure situation.

Measures After reading the scenario, participants rated their post-recovery satisfaction, followed by demographics and manipulation and realism checks. Scales were identical to the ones used in Study 1 and Study 2. All scenarios were perceived as realistic ($\alpha = 0.81$; M: 5.71, SD: 1.26). As manipulation check for failure attribution, respondents rated "who was responsible for the failure," anchored at "user (1)" up to "chatbot (7)." People in the chatbot-fault scenario held the chatbot more responsible for the failure compared to the user-fault scenario ($M_{Chatbot-fault}$: 4.69; SD: 1.89 vs. $M_{User-fault}$: 3.56, SD: 2.27, $t(247) = -4.08$, $p < 0.001$). Moreover, for the message types, respondents of the empathy scenario rated significantly stronger that the chatbot asked for their empathy and understanding ($M_{Empathy}$: 5.33, $M_{Solution}$: 2.90, $M_{Control}$: 2.25, $F(2246) = 130.93$, $p < 0.001$). Similarly, respondents in the solution message scenario perceived more strongly that the chatbot has suggested a solution ($M_{Solution}$: 4.34, $M_{Empathy}$: 3.10, $M_{Control}$: 2.39, $F(2246) = 26.62$, $p < 0.001$). Again, all scales exhibited adequate validity and reliability (see Table 3). In addition, Table 4 shows the means and standard deviations of the key variables.

Results To test H1 to H3 in one comprehensive model, we used a moderated mediation analysis (PROCESS Model 8, Hayes (2018) with 5000 bootstrap samples and 95% CIs) with the same setup as in the studies above, including age and gender as covariates.

Regarding H1 and H2, results confirmed Study 1 and Study 2. Again, the empathy message (vs. control) increased perceived warmth ($b = 1.18$, $p < 0.01$), and the solution message (vs. control) led to higher competence perceptions ($b = 1.22$, $p < 0.01$), supporting H1 and H2. In addition, results showed that the solution message (vs. control) did not increase warmth ($p = 0.51$) and the empathy message did not increase competence ($p = 0.50$). Satisfaction was influenced by warmth ($b = 0.16$, $p < 0.01$) and competence ($b = 0.58$, $p < 0.001$), supporting H3 again. Both message types had no direct impact on satisfaction.

Regarding H6a, the interaction of empathy message×failure attribution had no significant impact on warmth ($p = 0.64$). The indirect effect of the empathy message (vs. control) on satisfaction via warmth was significant in the case of a user-attributed failure ($b = 0.19$; [0.03, 0.41]) and in the case of a chatbot-attributed failure ($b = 0.15$; [0.04, 0.31]). Subsequently, the moderated mediation effect was not significant ($b = -0.04$; [−0.23, 0.14]). This means, irrespective of the failure attribution, there is a mediation effect of empathy on satisfaction via warmth. As a consequence, 6a could not be supported.

However, the situation changes when considering the solution message (H6b). In this case, the interaction of the solution message×failure attribution had a negative impact on competence ($b = -1.23$, $p < 0.05$). The indirect effect of the solution message (vs. control) on satisfaction via competence was significant in the case of a user-attributed failure ($b = 0.71$; [0.32, 1.14]), but not significant in the case of a chatbot-attributed failure ($b = 0.06$; [−0.31, 0.45]). The index of moderated mediation was significant and negative ($b = -0.65$; [−1.22, −0.12]). This indicates that the positive (mediated) effect of the solution message through competence on satisfaction is only supported when the failure is attributed to the user. When the chatbot is responsible for the failure, the positive effect diminishes. In sum, H6b could be supported.

Finally, to test H4 (a and b), we applied the same moderated mediation model (Model 8) and replaced satisfaction with re-use intentions. Results are comparable to those above. Empathy led to warmth ($b = 1.18$, $p < 0.01$) and solution increased competence ($b = 1.22$, $p < 0.01$). Moreover, solution did not lead to warmth ($p = 0.51$) and empathy did not lead to competence ($p = 0.50$). "In turn, chatbot re-use intentions were influenced by warmth ($b = 0.16$, $p < 0.05$) and competence ($b = 0.69$, $p < 0.001$), supporting H4 (a and b)." The effects of moderated mediation remained comparable to those above: The indirect effects of empathy via warmth on re-use intentions were significant irrespective of failure attribution ($b_{\text{user-attribution}} = 0.19$, [0.01; 0.46] and $b_{\text{chatbot-attribution}} = 0.15$, [0.01; 0.33]; index = −0.04; [−0.25;0.16]), while the indirect effects of solution via competence on re-use intentions were only significant in case of user-attribution (and not for chatbot-attribution) (i.e., $b_{\text{user-attribution}} = 0.84$, [0.37; 1.37] and $b_{\text{chatbot-attribution}} = 0.07$, [−0.38; 0.54]; index = −0.77; [−1.48; −0.13]).

Discussion

As response failures occur frequently during chatbot interactions, recovery strategies are greatly needed to mitigate negative user reactions, avoid financial losses, and assure re-use intentions. This is especially relevant for electronic markets and digital platforms such as Airbnb, Booking, or Uber, as service provision and customer-facing support are part of their key assets. To help answer the question of whether and how recovery messages might support these goals, the present research investigated how people respond to two characteristic recovery messages in chatbot conversations and focused on the mediating role of social cognition. Three experiments in two contexts compared the two characteristic messages empathy and solution and identified that these messages trigger social cognitions of warmth or competence (*H1* and *H2*)—which positively influence post-recovery satisfaction and chatbot re-use intentions (*H3 and H4*). Furthermore, the impacts of situational factors on message effectiveness were analyzed. First, failure frequency influences which message should be preferred (*H5*). More precisely, after a double deviation, only an empathy message has a significantly positive effect on warmth, whereas the solution message had no significant effect on competence any more. In contrast, when the chatbot solved the user request successfully after an initial failure, effects of different recovery messages dissolved. Thus, the final success of a chatbot interaction shifts post-hoc perceptions of the previous recovery messages.

Second, integrating the factor of failure attribution (*H6a/b*) showed that a solution message is particularly detrimental to user satisfaction with a chatbot-attributed failure (i.e., a mismatch). In this situation, the solution message did not lead to higher satisfaction (mediated via competence). In contrast, in a user-attributed failure situation, people seemed to accept a solution message more, as this message type led to higher post-recovery satisfaction via increased competence perceptions. An empathy message was found to be acceptable for both user- and chatbot-failure attributions. This indicates that an apology and request for understanding is "always possible" and a less critical approach compared to the solution message, and rather preferable when failure attribution remains unclear.

Theoretical contributions

This research responds to scholarly calls for further user-centered investigation of chatbot response failures (Diederich et al., 2020) and provides several theoretical contributions. First, we add to the growing body of research regarding digital agents' conversational design (Crolic et al., 2021; Sands et al., 2021; Song et al., 2022; Weiler et al., 2022). Interactions in electronic markets and particularly digital platforms (e.g., Airbnb, eBay) rise continuously, leading to a parallel increase in demand for effective and efficient customer service (Hein et al., 2020; Suta et al., 2020). Next to such user-facing platforms, chatbots are also increasingly implemented in corporate applications (e.g., Slack or Microsoft Teams) to facilitate processes and information access (Stoeckli et al., 2020). Thus, as chatbots are increasingly taking over tasks in the digital surrounding and are a major service innovation, an appropriate design of chatbot responses is key for positive customer experiences and firm profitability (Mozafari et al., 2022). This study proposes that message types, when used as a psychological recovery attempt, should be carefully chosen depending on situational factors like failure frequencies or failure attribution. These results offer a more nuanced view on the effectiveness of recovery messages—and confirm former studies that stated that chatbot designs should follow human service chat interactions in order to be successful (Belanche et al., 2021; Gnewuch et al., 2018; van Pinxteren et al., 2020).

Second, this research adds to the literature of service failures and recovery, particularly in the domain of digital agents (Chong et al., 2021; Mozafari et al., 2022). With this study, we respond to scholars who have called for an examination of effective recovery strategies to improve users' service experience after chatbot failures (Benner et al., 2021; Janssen et al., 2021; van der Goot et al., 2021). We also complement the findings of Weiler et al. (2022) who examined ex-ante strategies by showing that messages directly after the failure (ex-post) also have a positive effect on re-use intention and thus reduced discontinuance. Moreover, this research complements studies which consider the impact of recovery messages of digital agents (L. Lv et al., 2022a, 2022b; Song et al., 2023). As service delivery by chatbots becomes more widespread, understanding how people respond to chatbot recovery attempts is of crucial relevance to secure service quality and consumer loyalty (Mozafari et al., 2022; Sands et al., 2021). Supporting findings from related studies (such as Xu and Liu (2022), our study results show that messages could trigger different social cognitions and achieve their goal of increasing post-recovery satisfaction via different paths. In addition, this study examines several conditions that influence the effectiveness of a particular message. By including failure frequency (i.e., double deviation) and failure attributions in the research design, we illustrated that such dimensions indeed play a role for the optimal message choice. As such, this paper also adds to the scant research around double deviations (Pacheco et al., 2019) and to knowledge of the effects of failure attributions in the field of human–computer interaction. Additionally, results might encourage future-related work to incorporate these factors into their research as well.

Third, this study adds to research assessing social cognitions. Only recently have scholars started to assess the perceptions of warmth or competence in relation to digital (conversational) agents (Choi et al., 2021; Han et al., 2021; Kull et al., 2021; McKee et al., 2022; Xu & Liu, 2022). As new technology, such as artificial intelligence or machine learning, further develops, digital agents will interact in more humanlike service interactions and will increasingly imitate human behavior in order to create more favorable user responses. While many studies in this field concentrate on anthropomorphism as visual cues for warmth or competence (e.g., Choi et al., 2021), our research extends insights about text-related cues (Han et al., 2021; Kull et al., 2021). While prior studies focused on effects due to people's reactance (Han et al., 2021) or the impacts of an initial warmth- or competence-related message at the beginning of a conversation (Kull et al., 2021), this study considers post-failure messages and examines how two prototypical messages trigger social cognitions. Warmth and competence perceptions were found to be the underlying mechanisms of the respective messages on users' post-recovery responses. More precisely, message elements requesting a person's understanding are social-oriented and were perceived as warm, whereas a message which presents a possible solution is task-oriented and was perceived as competent. In turn, both perceptions increased post-recovery satisfaction and re-use intentions. This supports the "computers are social actors" (CASA) paradigm (Nass et al., 1996) and shows that chatbot responses are processed and perceived like human service-agent messages. However, the study also shows that the mediation through social perceptions could be eliminated by external circumstances. For instance, a double deviation (i.e., a chatbot's second non-understanding) removed the mediated effect of solution-oriented messages via competence.

Managerial implications

Results of the three studies provide guidance to both software designers and companies employing chatbots on how to implement chatbot recovery messages as cost-effective and universally usable tool to mitigate negative service experiences. First, using a dedicated recovery message is beneficial to mitigate negative users' responses after a chatbot failure with only marginal costs for software programming. This research revealed that each message follows a distinct path

to increase post-recovery satisfaction—either by driving competence-perceptions or warmth-perceptions of users. Uncovering these underlying mechanisms helps managers to understand how consumers' responses are formed. In particular, software designers can now formulate precise warmth- or competence-related messages as effective response to service failures.

Second, across the studies, competence perceptions generally exerted a stronger total effect on satisfaction than warmth. As the solution message fosters competence perceptions, this message type could therefore be considered a more effective strategy for both product- and service-related contexts. Using the solution message also allows chatbot designers to employ corrective measures to successfully conclude the conversation. However, if the recovery process was successful after the initial failure (i.e., the chatbot successfully resolved the request), the impact of the recovery messages dissolved, as consumers do not seem to care (posthoc) how they got to this point. Nevertheless, as likelihood of failure is high, managers and chatbot developers should be encouraged to incorporate one of the two message forms to safeguard against negative effects in case of failure without risking negative effects in case of success.

Third, the analysis of situational factors revealed several insights. When failing twice, the empathy message led to warmth (which acts as a mediator between message and satisfaction), while the solution message did not increase competence (and subsequently did not mediate between message and satisfaction). Regarding the final outcomes, however, the solution message generated higher means for post-recovery satisfaction and re-use intentions than the empathy message (see Table 4). This should be considered by managers when deciding on which message to use. Moreover, when people attribute the chatbot as responsible for the failure, only the empathy message is preferable. In that case, the solution message had no indirect effect on satisfaction (via competence), while the empathy message had a positive indirect effect on satisfaction. When managers are in doubt about whether the chatbot or user is responsible for the failure, the empathy message reflects a rather uncritical choice. In sum, our results show that the "solution" message is more effective than the "empathy" message in some situations, while it is the other way round in other situations. Therefore, managers need to be aware of the type of failure to evaluate failure attributions, and about the failure frequency, in order to adapt the recovery messages accordingly.

More generally, with the fast-paced developments in the field of deep learning and large language models, managers might be tempted to integrate chatbots such as "ChatGPT" in their service processes (Dwivedi et al., 2023). However, unlike most current chatbots (based on natural language processing or simple decision trees), which respond generally with some sort of error message (e.g., "Sorry, I don't

know"), ChatGPT generally responds with a text expressing the most likely answer. Based on a vast amount of available text, the algorithm aims to anticipate the highest likelihood of an answer by forecasting what a human would use to reply to the specific request. Thus, instead of acknowledging failure, ChatGPT often "hallucinates," meaning that this kind of chat tools produce information that may be nonsensical, untrue, or inconsistent with the content of the source input (Dwivedi et al., 2023; Ji et al., 2023). In the context of diverse service interactions, such hallucinated responses to user queries pose a significant threat, as service activities are often associated with actions (e.g., customer data, confirmations, bookings, and returns). Therefore, while integrating language processing models such as ChatGPT may be beneficial for service interactions, failure acknowledgment and recovery attempts (e.g., via messages) remain highly relevant for digital service interactions.

Limitations and future research

Although this research offers valuable insights, it also has some limitations.

First, our study relied on screenshots of chat conversations to ensure high internal validity. To add validity to our findings, future research could investigate our framework in the field. In this vein, scholars could also analyze if new and more sophisticated bots such as ChatGPT are less prone to service failures, and whether these bots could also integrate more context-aware information to create a more personalized and failure-congruent recovery message. Moreover, longitudinal designs would provide a fuller perspective on the chatbot recovery process and allow to investigate possible long-term effects of chatbot messages.

Second, this study considered failure frequency and failure attribution as two situational factors. Future studies could include additional factors such as the type of product or service associated with the chatbot service. While our studies used a product-related and two service-related cases to somehow include this situational factor, future studies might investigate product- or service-specific features (e.g., simple vs. complex; hedonic vs. utilitarian; low- vs. high-risk). Next to this, future research might also explore the effects of other design elements, such as different message tonalities or recovery feedback elements, in combination with the two message types. For instance, a chatbot could present a message and ask if the information was helpful. Related chatbot studies revealed that already minor adaptations in the conversational design (e.g., response delays or chatbot- vs. user-initiation) have effects on user's satisfaction with the chatbot (Gnewuch et al., 2018; Pizzi et al., 2021).

Third, while our research did not focus on the role of emotions in chatbot failure and recovery, prior research

found emotions to influence consumers' reactions in chatbot interactions (Crolic et al., 2021). Future studies should therefore investigate the role of emotions such as anger, frustration, and helplessness in human-chatbot interactions.

Fourth, we used two prototypical messages to measure their effects precisely, neglecting other possible forms or mixtures of messages, or even the combination with other forms of compensation such as vouchers or human interaction, leaving open a fruitful field for future research related to digital agents' conversational design.

References

Aaker, J., 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

Adam, M., Wessel, M., & Benlian, A. (2021). AI-based chatbots in customer service and their effects on user compliance. *Electronic Markets, 31*(2), 427–445. https://doi.org/10.1007/s12525-020-00414-7

Agustin, C., & Singh, J. (2005). Curvilinear effects of consumer loyalty determinants in relational exchanges. *Journal of Marketing Research, 42*(1), 96–108. https://doi.org/10.1509/jmkr.42.1.96.56961

Akhtar, M., Neidhardt, J., & Werthner, H. (2019). The potential of chatbots: Analysis of chatbot conversations. In M. Akhtar, J. Neidhardt, & H. Werthner (Eds.), *Potential of Chatbots: Analysis of Chatbot Conversations* (pp. 397–404). IEEE. https://doi.org/10.1109/CBI.2019.00052

Alt, R. (2020). Evolution and perspectives of electronic markets. *Electronic Markets, 30*(1), 1–13. https://doi.org/10.1007/s12525-020-00413-8

Ashktorab, Z., Jain, M., Liao, Q. V., & Weisz, J. D. (2019). Resilient chatbots. In S. Brewster, G. Fitzpatrick, A. Cox, & V. Kostakos (Eds.), *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19* (pp. 1–12). ACM Press. https://doi.org/10.1145/3290605.3300484

Babbar, S., & Koufteros, X. (2008). The human element in airline service quality: Contact personnel and the customer. *International Journal of Operations & Production Management, 28*(9), 804–830. https://doi.org/10.1108/01443570810895267

Belanche, D., Casaló, L. V., Schepers, J., & Flavián, C. (2021). Examining the effects of robots' physical appearance, warmth, and competence in frontline services: The Humanness-Value-Loyalty model. *Psychology & Marketing, 38*(12), 2357–2376. https://doi.org/10.1002/mar.21532

Benner, D., Elshan, E., Schöbel, S., & Janson, A. (2021). What do you mean? A review on recovery strategies to overcome conversational breakdowns of conversational agents. *ICIS 2021 Proceedings.* 13.

Brandtzaeg, P. B., & Følstad, A. (2018). Chatbots: Changing user needs and motivations. *Interactions, 25*(5), 38–43. https://doi.org/10.1145/3236669

Brendel, A. B., Greve, M., Diederich, S., Bührke, J., & Kolbe, L. M. (2020). 'You are an Idiot!' – How Conversational Agent Communication Patterns Influence Frustration and Harassment. *AMCIS 2020 Proceedings.* 13.

Choi, S [Sunmee]., & Mattila, A. S. (2008). Perceived controllability and service expectations: Influences on customer reactions following service failure. *Journal of Business Research, 61*(1), 24–30. https://doi.org/10.1016/j.jbusres.2006.05.006

Choi, S [Sungwoo]., Mattila, A. S., & Bolton, L. E. (2021). To Err Is Human(-oid): How Do Consumers React to Robot Service Failure and Recovery? *Journal of Service Research, 24*(3), 354–371. https://doi.org/10.1177/1094670520978798

Chong, T., Yu, T., Keeling, D. I., & de Ruyter, K. (2021). AI-chatbots on the services frontline addressing the challenges and opportunities of agency. *Journal of Retailing and Consumer Services, 63*, 102735. https://doi.org/10.1016/j.jretconser.2021.102735

Collier, J. E., Breazeale, M., & White, A. (2017). Giving back the "self" in self service: Customer preferences in self-service failure recovery. *Journal of Services Marketing, 31*(6), 604–617. https://doi.org/10.1108/JSM-07-2016-0259

Crolic, C., Thomaz, F., Hadi, R., & Stephen, A. T. (2021). Blame the bot: Anthropomorphism and anger in customer-chatbot interactions. *Journal of Marketing, 86*(1), 132–148. https://doi.org/10.1177/00222429211045687

Cuddy, A. J., Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology. Advances in experimental social psychology* (Vol. 40, pp. 61–149). Elsevier. https://doi.org/10.1016/S0065-2601(07)00002-0

Diederich, S., Brendel, A. B., & Kolbe, L. M. (2020). Designing anthropomorphic enterprise conversational agents. *Business & Information Systems Engineering, 62*(3), 193–209. https://doi.org/10.1007/s12599-020-00639-y

Dilmegani, C. (2022). *Chatbot: 9 Epic Chatbot/Conversational Bot Failures*. Retrieved September, 27, 2023 from https://research.aimultiple.com/chatbot-fail/

Dubois, D., Rucker, D. D., & Galinsky, A. D. (2016). Dynamics of communicator and audience power: The persuasiveness of competence versus warmth. *Journal of Consumer Research, 43*(1), 68–85. https://doi.org/10.1093/jcr/ucw006

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D.,`Carter, L., Chowdhury, S., Crick, T., Cunningham, S. W., Davies, G. H., Davison, R. M., Dé, R., Dennehy, D., Duan, Y., Dubey, R., Dwivedi, R., Edwards, J. S., Flavián, C., Gauld, R., Grover, V., Hu, M.-C., Janssen, M., Jones, P., Junglas, I., Khorana, S., Kraus, S., Larsen, K. R., Latreille, P., Laumer, S., Malik, F. T., Mardani, A., Mariani, M., Mithas, S., Mogaji, E., Nord, J. H., O'Connor, S., Okumus, F., Pagani, M., Pandey N., Papagiannidis, S., Pappas, I. O., Pathak, N., Pries-Heje, J., Raman, R., Rana, N. P., Rehm, S.-V., Ribeiro-Navarrete, S., Richter, A., Rowe, F., Sarker, S., Stahl, B. C., Tiwari, M. K., van der Aalst, W., Venkatesh, V.,

Viglia, G., Wade, M., Walton, P., Wirtz, J., & Wright, R. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, *71*, 102642.

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences, 11*(2), 77–83. https://doi.org/10.1016/j.tics.2006.11.005

Gelbrich, K. (2010). Anger, frustration, and helplessness after service failure: Coping strategies and effective informational support. *Journal of the Academy of Marketing Science, 38*(5), 567–585. https://doi.org/10.1007/s11747-009-0169-6

Gerdes, K. E. (2011). Empathy, sympathy, and pity: 21st-century definitions and implications for practice and research. *Journal of Social Service Research, 37*(3), 230–241. https://doi.org/10.1080/01488376.2011.564027

Gnewuch, U., Morana, S., & Maedche, A. (2017). Towards designing cooperative and social conversational agents for customer service. *Proceedings of the International Conference on Information Systems*, *38*.

Gnewuch, U., Morana, S., Adam, M. T., & Maedche, A. (2018). Faster is not always better: Understanding the effect of dynamic response delays in human-chatbot interaction. *ECIS Proceedings Research Papers.* 113. https://www.semanticscholar.org/paper/Faster-is-Not-Always-Better%3A-Understanding-the-of-Gnewuch-Morana/22cbf658ea99b2901b3f6f649e21ef8a3c7a590d

Güntürkün, P., Haumann, T., & Mikolon, S. (2020). Disentangling the Differential Roles of Warmth and Competence Judgments in Customer-Service Provider Relationships. *Journal of Service Research, 23*(4), 476–503. https://doi.org/10.1177/1094670520920354

Habel, J., Alavi, S., & Pick, D. (2017). When serving customers includes correcting them: Understanding the ambivalent effects of enforcing service rules. *International Journal of Research in Marketing, 34*(4), 919–941. https://doi.org/10.1016/j.ijresmar.2017.09.002

Han, E., Yin, D., & Zhang, H. (2021). Interruptions during a service encounter: Dealing with imperfect chatbots. *International Conference on Interaction Sciences (ICIS) Proceedings*. 6. https://aisel.aisnet.org/icis2021/ai_business/ai_business/6

Hayes, A. F. (2018). *Introduction to mediation, moderation and conditional process analysis A Regression-Based Approach* (2nd ed.). The Guilford Press.

Hein, A., Schreieck, M., Riasanow, T., Setzke, D. S., Wiesche, M., Böhm, M., & Krcmar, H. (2020). Digital platform ecosystems. *Electronic Markets, 30*(1), 87–98. https://doi.org/10.1007/s12525-019-00377-4

Hogan, R. (1969). Development of an empathy scale. *Journal of Consulting and Clinical Psychology, 33*(3), 307–316. https://doi.org/10.1037/h0027580

Hosseini, M., & Caragea, C. (2021). It takes two to empathize: One to seek and one to provide. *Proceedings of the AAAI Conference on Artificial Intelligence, 35*(14), 13018–13026. https://doi.org/10.1609/aaai.v35i14.17539

Huang, R., & Ha, S. (2020). The effects of warmth-oriented and competence-oriented service recovery messages on observers on online platforms. *Journal of Business Research, 121*, 616–627. https://doi.org/10.1016/j.jbusres.2020.04.034

Hulland, J., Baumgartner, H., & Smith, K. M. (2018). Marketing survey research best practices: Evidence and recommendations from a review of JAMS articles. *Journal of the Academy of Marketing Science, 46*(1), 92–108. https://doi.org/10.1007/s11747-017-0532-y

Janssen, A., Grützner, L., & Breitner, M. H. (2021). Why do chatbots fail? A critical success factors analysis. *ICIS 2021 Proceedings*. 6.

Ji, Z., Lee, N., Frieske, R., Yu, T [Tiezheng]., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys, 55*(12), 1–38. https://doi.org/10.1145/3571730

Johnston, R., & Fern, A. (1999). Service recovery strategies for single and double deviation scenarios. *The Service Industries Journal, 19*(2), 69–82. https://doi.org/10.1080/02642069900000019

Judd, C. M., James-Hawkins, L., Yzerbyt, V., & Kashima, Y. (2005). Fundamental dimensions of social judgment: Understanding the relations between judgments of competence and warmth. *Journal of Personality and Social Psychology, 89*(6), 899–913. https://doi.org/10.1037/0022-3514.89.6.899

Kull, A. J., Romero, M., & Monahan, L. (2021). How may I help you? Driving brand engagement through the warmth of an initial chatbot message. *Journal of Business Research, 135*, 840–850. https://doi.org/10.1016/j.jbusres.2021.03.005

Lazarus, R. S. (1991). *Emotion and Adaptation*. Oxford University Press.

Li, C [Charlene]., Mirosa, M., & Bremer, P. (2020). Review of online food delivery platforms and their impacts on sustainability. *Sustainability, 12*(14), 5528. https://doi.org/10.3390/su12145528

Luger, E., & Sellen, A. (2016). "Like having a really bad PA". In J. Kaye, A. Druin, C. Lampe, D. Morris, & J. P. Hourcade (Eds.), *#chi4good: Chi 2016 : San Jose, CA, USA, May 7–12 : Proceedings : The 34th Annual CHI Conference on Human Factors in Computing Systems : San Jose Convention Center* (pp. 5286–5297). The Association for Computing Machinery. https://doi.org/10.1145/2858036.2858288

Lv, X., Liu, Y [Yue]., Luo, J., Liu, Y [Yuqing]., & Li, C [Chunxiao]. (2021). Does a cute artificial intelligence assistant soften the blow The impact of cuteness on customer tolerance of assistant service failure. *Annals of Tourism Research, 87*, 103114. https://doi.org/10.1016/j.annals.2020.103114

Lv, L., Huang, M., Guan, D., & Yang, K. (2022a). Apology or gratitude? The effect of communication recovery strategies for service failures of AI devices. *Journal of Travel & Tourism Marketing, 39*(6), 570–587. https://doi.org/10.1080/10548408.2022.2162659

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

Maedche, A., Legner, C., Benlian, A., Berger, B., Gimpel, H., Hess, T., Hinz, O., Morana, S., & Söllner, M. (2019). AI-based digital assistants. *Business & Information Systems Engineering, 61*(4), 535–544. https://doi.org/10.1007/s12599-019-00600-8

Marinova, D., Singh, S. K., & Singh, J. (2018). Frontline problem-solving effectiveness: A dynamic analysis of verbal and nonverbal cues. *Journal of Marketing Research, 55*(2), 178–192. https://doi.org/10.1509/jmr.15.0243

McKee, K. R., Bai, X., & Fiske, S. T. (2022). Warmth and competence in human-agent cooperation. *AAMAS Proceedings.*898–907

Misselhorn, C. (2009). Empathy with inanimate objects and the uncanny valley. *Minds and Machines, 19*(3), 345–359. https://doi.org/10.1007/s11023-009-9158-2

Mordor Intelligence. (2021). *Chatbot market: growth, trends, COVID-19 impact, and forecasts*. Retrieved September, 27, 2023 from https://www.mordorintelligence.com/industry-reports/chatbot-market

Mostafa, R. B., & Kasamani, T. (2022). Antecedents and consequences of chatbot initial trust. *European Journal of Marketing, 56*(6), 1748-1771. https://doi.org/10.1108/EJM-02-2020-0084

Mozafari, N., Weiger, W. H., & Hammerschmidt, M. (2021). That's so embarrassing! When not to design for social presence in human–chatbot. *ICIS 2021 Proceedings*, *5*. https://aisel.aisnet.org/icis2021/hci_robot/hci_robot/5

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

Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates? *International Journal of Human-Computer Studies, 45*(6), 669–678. https://doi.org/10.1006/ijhc.1996.0073

Olson, P. (2022). *Google Faces a Serious Threat From ChatGPT*. Retrieved September, 27, 2023 from https://www.washingtonpost.com/business/energy/google-faces-a-serious-threat-from-chatgpt/2022/12/07/363d2440-75f5-11ed-a199-927b334b939f_story.html

Ozgen, O., & Duman Kurt, S. (2012). Pre-recovery and post-recovery emotions in the service context: A preliminary study. *Managing Service Quality: An International Journal, 22*(6), 592–605. https://doi.org/10.1108/09604521211287561

Pacheco, N. A., Pizzutti, C., Basso, K., & van Vaerenbergh, Y. (2019). Trust recovery tactics after double deviation: Better sooner than later? *Journal of Service Management, 30*(1), 2–22. https://doi.org/10.1108/JOSM-02-2017-0056

Pizzi, G., Scarpi, D., & Pantano, E. (2021). Artificial intelligence and the new forms of interaction: Who has the control when interacting with a chatbot? *Journal of Business Research, 129*, 878–890. https://doi.org/10.1016/j.jbusres.2020.11.006

Poser, M., Singh, S., & Bittner, E. (2021). Hybrid service recovery: Design for seamless inquiry handovers between conversational agents and human service agents. In T. Bui (Ed.), *HICSS Proceedings*. 54. https://doi.org/10.24251/HICSS.2021.144

Qualtrics and ServiceNow. (2021). *Customer Experience*. Retrieved September, 27, 2023 from https://www.qualtrics.com/blog/qualtrics-servicenow-customer-service-research/

Reinkemeier, F., & Gnewuch, U. (2022). Designing effective conversational repair strategies for chatbots. *ECIS Proceedings*, 30.

Roschk, H., & Gelbrich, K. (2014). Identifying appropriate compensation types for service failures. *Journal of Service Research, 17*(2), 195–211. https://doi.org/10.1177/1094670513507486

Sands, S., Ferraro, C., Campbell, C., & Tsao, H.-Y. (2021). Managing the human–chatbot divide: How service scripts influence service experience. *Journal of Service Management, 32*(2), 246–264. https://doi.org/10.1108/JOSM-06-2019-0203

Sands, S., Campbell, C., Plangger, K., & Pitt, L. (2022). Buffer bots: The role of virtual service agents in mitigating negative effects when service fails. *Psychology & Marketing, 39*(11), 2039–2054. https://doi.org/10.1002/mar.21723

Scott, M. L., Mende, M., & Bolton, L. E. (2013). Judging the book by its cover? How consumers decode conspicuous consumption cues in buyer–seller relationships. *Journal of Marketing Research, 50*(3), 334–347. https://doi.org/10.1509/jmr.11.0478

Seeger, A.-M., & Heinzl, A. (2021). Chatbots often fail! Can anthropomorphic design mitigate trust loss in conversational agents for customer service? *ECIS Proceedings,* 12.

Sheehan, B., Jin, H. S., & Gottlieb, U. (2020). Customer service chatbots: Anthropomorphism and adoption. *Journal of Business Research, 115*, 14–24. https://doi.org/10.1016/j.jbusres.2020.04.030

Song, M., Du, J., Xing, X., & Mou, J. (2022). Should the chatbot "save itself" or "be helped by others"? The influence of service recovery types on consumer perceptions of recovery satisfaction. *Electronic Commerce Research and Applications, 55*, 101199. https://doi.org/10.1016/j.elerap.2022.101199

Song, M., Zhang, H., Xing, X., & Duan, Y [Yucong]. (2023). Appreciation vs. apology: Research on the influence mechanism of chatbot service recovery based on politeness theory. *Journal of Retailing and Consumer Services, 73*, 103323. https://doi.org/10.1016/j.jretconser.2023.103323

Sporrer, T. (2021). *Your Chatbot Is Frustrating Your Customers: Improve the Experience with Guidance*. Retrieved September, 27, 2023 from https://customerthink.com/your-chatbot-is-frustrating-your-customers-improve-the-experience-with-guidance/

Stoeckli, E., Dremel, C., Uebernickel, F., & Brenner, W. (2020). How affordances of chatbots cross the chasm between social and traditional enterprise systems. *Electronic Markets, 30*(2), 369–403. https://doi.org/10.1007/s12525-019-00359-6

Suta, P., Lan, X., Wu, B., Mongkolnam, P., & Chan, J. H. (2020). An overview of machine learning in chatbots. *International Journal of Mechanical Engineering and Robotics Research*, 502–510. https://doi.org/10.18178/ijmerr.9.4.502-510

Thorbecke, C. (2022). *Chatbots: A long and complicated history*. CNN. Retrieved September, 27, 2023 from https://edition.cnn.com/2022/08/20/tech/chatbot-ai-history/index.html

van der Goot, M. J., Hafkamp, L., & Dankfort, Z. (2021). Customer service chatbots: A qualitative interview study into customers' communication journey. In A. Følstad, T. Araujo, S. Papadopoulos, E.L.-C. Law, E. Luger, M. Goodwin, & P. B. Brandtzaeg (Eds.), *Chatbot Research and Design* (pp. 190–204). Springer International Publishing.

van Doorn, J., Mende, M., Noble, S. M., Hulland, J., Ostrom, A. L., Grewal, D., & Petersen, J. A. (2016). Domo Arigato Mr. *Roboto. Journal of Service Research, 20*(1), 43–58. https://doi.org/10.1177/1094670516679272

van Pinxteren, M. M., Pluymaekers, M., & Lemmink, J. G. (2020). Human-like communication in conversational agents: A literature review and research agenda. *Journal of Service Management, 31*(2), 203–225. https://doi.org/10.1108/JOSM-06-2019-0175

van Vaerenbergh, Y., Varga, D., de Keyser, A., & Orsingher, C. (2019). The service recovery journey: Conceptualization, integration, and directions for future research. *Journal of Service Research, 22*(2), 103–119. https://doi.org/10.1177/1094670518819852

Wallenburg, C. M., & Lukassen, P. (2011). Proactive improvement of logistics service providers as driver of customer loyalty. *European Journal of Marketing, 45*(3), 438–454. https://doi.org/10.1108/03090561111107267

Weiler, S., Matt, C., & Hess, T. (2022). Immunizing with information - Inoculation messages against conversational agents' response failures. *Electronic Markets, 32*(1), 239–258. https://doi.org/10.1007/s12525-021-00509-9

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review, 92*(4), 548–573. https://doi.org/10.1037/0033-295X.92.4.548

Weiner, B. (2012). An attribution theory of motivation. In P. A. Van-Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (1st ed., pp. 135–155). Sage.

Wieseke, J., Geigenmüller, A., & Kraus, F. (2012). On the role of empathy in customer-employee interactions. *Journal of Service Research, 15*(3), 316–331. https://doi.org/10.1177/1094670512439743

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

Worsfold, K., Worsfold, J., & Bradley, G. (2007). Interactive effects of proactive and reactive service recovery strategies: The case of rapport and compensation. *Journal of Applied Social Psychology, 37*(11), 2496–2517. https://doi.org/10.1111/J.1559-1816.2007.00267.X

Xie, C., Bagozzi, R. P., & Grønhaug, K. (2015). The role of moral emotions and individual differences in consumer responses to corporate green and non-green actions. *Journal of the Academy of Marketing Science, 43*(3), 333–356. https://doi.org/10.1007/s11747-014-0394-5

Xing, X., Song, M., Duan, Y [Yucong]., & Mou, J. (2022). Effects of different service failure types and recovery strategies on the

consumer response mechanism of chatbots. *Technology in Society, 70*, 102049. https://doi.org/10.1016/j.techsoc.2022.102049

Xu, X., & Liu, J. (2022). Artificial intelligence humor in service recovery. *Annals of Tourism Research, 95*, 103439. https://doi.org/10.1016/j.annals.2022.103439

Yang, Z., Zhou, J., & Yang, H. (2023). The impact of AI's response method on service recovery satisfaction in the context of service failure. *Sustainability, 15*(4), 3294. https://doi.org/10.3390/su15043294

You, Y., Yang, X., Wang, L., & Deng, X. (2020). When and Why Saying "Thank You" Is Better Than Saying "Sorry" in Redressing Service Failures: The Role of Self-Esteem. *Journal of Marketing, 84*(2), 133–150. https://doi.org/10.1177/0022242919889894