European Journal of Marketing

Relief through digital modalities: evaluating text-based chatbots and voicebots in service failures

Mai Nguyen
Department of Marketing, Griffith Business School, Griffith University, Queensland, Australia, and School of Languages and Tourism, Hanoi University of Industry, Hanoi, Vietnam

Luis Arango
Department of Marketing, The University of Auckland, Auckland, New Zealand

Yunen Zhang
School of Business Administration, Jiangxi University of Finance and Economics, Nanchang, China, and

Nicole Hartley and Felix Septianto
UQ Business School, The University of Queensland, Brisbane, Australia

Received 27 August 2024
Revised 23 April 2025
17 September 2025
19 November 2025
Accepted 2 February 2026

Abstract

Purpose – This study aims to examine the comparative effect that two widespread technologies, text-based chatbots and voicebots, have on perceived relief and subsequent customer outcomes in service failure settings. The moderating effect of service failure severity on this mediation relationship is also evaluated.

Design/methodology/approach – This research adopts an experiential design, consisting of an initial exploratory field study (Study 1), to examine how AI bot modality influences real-world customer evaluations in a naturalistic service failure setting. This was followed by two between-subjects experimental studies (Studies 2 and 3), which sought to validate the underlying mechanism (i.e. perceived relief) and test the moderating role of service failure severity under controlled conditions.

Findings – Study 1 provided initial empirical evidence that, in a real-world service failure context, customers who interacted with a chatbot (vs a voicebot) gave significantly higher satisfaction ratings. Study 2 provided evidence of the positive indirect effect of chatbots on customer reviews via perceived relief. Study 3 supported this mediation relationship while also demonstrating that such a relationship is moderated by the severity of the service failure.

Research limitations/implications – Based on the findings of this study, it would be a stretch to conclude that chatbots are invariably better. The dynamic nature of customer preferences and technological advancements necessitates continuous evaluation and adaptation of customer service strategies. Managers should regularly review the performance of both chatbots and voicebots, incorporating customer feedback and technological improvements to optimize their service delivery approaches. By understanding and applying these insights, managers can better navigate service failures, ultimately leading to improved customer outcomes.

European Journal of Marketing
Vol. 60 No. 13, 2026
pp. 325-352
Emerald Publishing Limited
0309-0566
DOI 10.1108/EJM-08-2024-0693

Practical implications – This study used an overall measure of satisfaction as the dependent variable. Future research might examine similar effects with other dependent variables, such as customer loyalty, trust or perceived value. Investigating these additional dimensions could provide a more comprehensive understanding of how chatbots and voicebots influence various aspects of customer experience and behavior.

Originality/value – This study represents one of the first to empirically examine the differential effects of text-based chatbots and voicebots (bot modality) on salient customer outcomes. To the best of the authors' knowledge, it is the first study to specifically examine their comparative effects in response to service failure settings where these AI-driven bots are predominantly used.

Keywords Service failure, AI, Chatbot, Modality, Voicebot

Paper type Research paper

Customer interactions are increasingly mediated by AI-powered robots (Belanche *et al.*, 2020). These human–robot interactions are rapidly growing across a wide variety of digital consumer settings. In online contexts, text-based chatbots, which enable real-time written word interactions, have been used in sectors such as healthcare (Holland *et al.*, 2021; Valtolina *et al.*, 2020), retail (Jiang *et al.*, 2022), education (Pérez *et al.*, 2020) or entertainment (García-Méndez *et al.*, 2021) to provide customer service support and enhance customer experience. A number of factors, including advances in generative AI that power "smarter" chat exchanges and well-known economic efficiencies (e.g. reduced human labor costs), indicate that this market is likely to continue expanding at a high pace (Dinh and Park, 2023). Parallel to the rise of text-based chatbots, voicebots are also becoming prevalent (Hoy, 2018; Klaus and Zaichkowsky, 2020). Voicebots, which use natural language processing and speech recognition technologies (Kaur *et al.*, 2020), provide an intuitive and accessible interface for users by allowing interactions through spoken language. This technology has also seen significant adoption across a wide variety of sectors, with the market expected to reach almost 100 billion in size by 2027 (Begum, 2024). Currently, global market players such as Amazon, Marriott, Microsoft and Domino's are leveraging the power of both chatbots and voicebots (Bezrukava, 2023).

While consumer preference for either chatbot or voicebot interaction exchanges is partly driven by contextual factors (e.g. some contexts might favor the hands-free modality provided by voicebots; e.g. Zierau *et al.*, 2023), frequently customers face a choice between the chat and voice modalities when interacting with a bot. For instance, with the recent launch of GPT-4-o ("omni"), customers can interact with this state-of-the-art LLM through both text and conversation. Considering the prevalence of chatbots and voicebots and their high rate of adoption, and the choice presented to consumers in these interactions in relation to the type of bot to interact with, a question arises as to the effects of these different modalities on customer service outcomes.

In this work, we advance the consumer literature on bot modality (text versus voice) in online customer service contexts, specifically focusing on their use in human–robot interactions pertaining to service failures. Despite an emerging wealth of research on chatbots and voicebots (for reviews and research agendas, see Klaus and Zaichkowsky, 2020; Mohamad Suhaili *et al.*, 2021; Schöbel *et al.*, 2024), only a few studies have directly compared these two modalities (e.g. Rohit *et al.*, 2024; Rzepka *et al.*, 2022; Zierau *et al.*, 2023). Moreover, in service failure settings, to the extent of our knowledge, there remains a lack of scholarly research comparing the differential effects of chatbots and voicebots on customer service outcomes. Given the widespread adoption and predicted growth of AI-driven customer service tools (McKinsey and Company, 2024; Salesforce, 2024), this lack of

comparative research in service failure settings represents a significant gap. Thus, this research seeks to answer the following key research question:

RQ1. How do chatbots and voicebots differentially impact customer evaluations following a service failure?

Because service failures trigger emotional responses in consumers (Balaji *et al.*, 2017; Harrison-Walker, 2012, 2019) and vary in severity (Leo and Huh, 2020; Weun *et al.*, 2004), refining our understanding of *RQ1* requires examining the mechanisms and contexts that potentially drive differential consumer evaluations of chatbots and voicebots. Hence, we also ask:

RQ2. Do chatbot and voicebot interactions elicit distinct emotional responses from customers during service failures?

RQ3. Does service failure severity interact with bot modality to influence customer responses to service failures?

Tackling the gap constituted by these research questions is essential due to the increasing prevalence of AI-driven service interactions and the risk that companies may inadvertently destroy value by using a less effective modality in service failure settings – or use it in an inappropriate context, thereby failing to promote positive customer outcomes. Our predictions and the broad rationale behind them are as follows. First, chatbot interactions, which are reliant on text responses, introduce elements such as delays or higher cognitive demands that facilitate emotion regulation mechanisms (e.g. affective fading and attentional deployment), allowing negative emotions evoked by service failures the opportunity to subside or downregulate (Skowronski *et al.*, 2014). Because relief is directly linked to more favorable customer evaluations (Reeck and Onuklu, 2022), the use of chatbots in service failure interactions could be expected to lead to higher customer review ratings following a service failure.

Also, when service failure severity is considered to be minor, customers have the motivation and cognitive resources to regulate their emotions effectively through mechanisms such as the ones mentioned, making chat-based interactions more effective at providing relief (Berger *et al.*, 2022; Gross, 1998). However, in high-severity failures, the urgency of problem resolution takes precedence over emotion regulation, as customers shift their focus toward obtaining immediate solutions rather than managing their emotions (Gross, 2015; Matthews *et al.*, 2021). In high-severity failures, both chatbots and voicebots present themselves as equally efficient problem-solving devices, and then no differences are expected.

Building primarily on insights from emotion regulation theory (Gross, 1998) while also integrating perspectives from other paradigms (e.g. extended model of emotion regulation; Gross, 2015), this research experimentally demonstrates that text-based chatbots are generally preferable in service failure settings (Studies 1 and 2). Demonstrating this baseline effect is important because, despite the widespread adoption of both chatbots and voicebots in customer service, little is known about their relative effectiveness in emotionally charged contexts such as service failures. Establishing this foundational difference sets the stage for deeper inquiry into why and when such differences emerge, particularly under varying levels of the severity of the service failure (Study 3). It is shown that the mechanism driving the positive effect of text-based chatbots on customer outcomes such as customer review (our dependent variable, defined as an overall assessment of the service interaction; Crolic *et al.*, 2021) in instances of low-severity-failure scenarios is increased perceived relief (Studies 2 and 3), defined as the easing and smoothing of tension and negative feelings (Kobel and Groeppel-Klein, 2021).

This work makes several theoretical and managerial contributions. First, it extends the literature on AI-driven service failures by systematically comparing chatbot and voicebot interactions, an area largely neglected in prior research (Rohit *et al.*, 2024; Rzepka *et al.*, 2022). While chatbots and voicebots have been extensively studied in isolation (Klaus and Zaichkowsky, 2020; Mohamad Suhaili *et al.*, 2021), little research has examined how these modalities differentially affect customer outcomes in service failure scenarios, where managing negative emotions is critical. Second, this research contributes to the literature on human–robot interaction by demonstrating that more humanlike AI interactions, such as voicebots, may not always lead to better customer service outcomes (Crolic *et al.*, 2021; Valenzuela and Hadi, 2018). Specifically, in service failures, the increased emotional engagement associated with voicebot interactions may amplify distress, whereas chatbot interactions facilitate perceived relief. Third, we build on emotion regulation theory (Gross, 1998, 2015) by introducing a new perspective on how AI-mediated customer interactions influence emotion regulation. While prior research has focused on consumer strategies for managing emotions (Berger *et al.*, 2022), our study shows that modality itself shapes regulatory processes, offering a novel lens through which to understand consumer responses to service failures. Fourth and relatedly, it is demonstrated that chatbot and voicebot modalities have distinct impacts on relief. Prior work on relief has related this construct to, for example, humor (Kobel and Groeppel-Klein, 2021), but research remains highly limited. Finally, this work offers important managerial insights by helping firms optimize their AI-driven service strategies using human–robot interaction technologies. Firms can tailor their bot modality choices based on service failure characteristics, ensuring that AI tools are leveraged in ways that maximize customer relief and satisfaction while minimizing unintended negative consequences.

Theoretical development

Chatbots vs voicebots

Chatbots are disembodied online AI-driven conversational agents that provide assistance to individuals through text (Liu *et al.*, 2023). Voicebots share most of these features, except that, assistance is voice-based (Klaus and Zaichkowsky, 2020). Although there are obvious technological differences between chatbots and voicebots, such as voicebots using speech recognition, they share many features, including AI and machine learning algorithms (or, now less frequently, rules-based algorithms), integration capabilities and data collection. Both technologies are widely used in various sectors, with common applications in customer service across retail service firms (Klaus and Zaichkowsky, 2020; Rapp *et al.*, 2021), but also with applications in sectors such as healthcare (Sezgin *et al.*, 2020; Singh *et al.*, 2024), as well as emerging sectors such as education (Sharma and Singh, 2021; Tärning and Silvervarg, 2019). Factors like their instantaneous and permanent availability, scalability, intuitive interfaces and potential for a high return on investment are strong selling points (Huang and Dootson, 2022; McLean and Osei-Frimpong, 2019).

In customer settings, a copious body of research has emerged on chatbots and voicebots. Both chatbots and voicebots have been associated with positive customer outcomes such as positive word-of-mouth (Mishra *et al.*, 2022; Yun and Park, 2022), positive feelings (Bilquise *et al.*, 2022; Hernandez-Ortega and Ferreira, 2021) and loyalty (Jenneboer *et al.*, 2022; Shahzad *et al.*, 2024). However, these outcomes, not surprisingly, depend on the particular design of the bots (e.g. language style: Jenneboer *et al.*, 2022). Also, some negative outcomes associated with the use of chatbots and voicebots loom large, for instance, privacy concerns (Agnihotri and Bhattacharya, 2024).

While both chatbots and voicebots have been widely studied individually, and to a much lesser extent comparatively, their relative impact in service failure settings remains largely unexplored (Rohit *et al.*, 2024; Zierau *et al.*, 2023). This oversight is notable given the growing use of both technologies in service contexts (Begum, 2024; Dinh and Park, 2023), the cost implications of adopting one over the other (Akanksha, 2024) and their distinct communicative features, making findings from one modality hard to generalize to the other (Huang and Sénécal, 2023).

In response, this paper seeks to advance the research streams on chatbots and voicebots by contrasting their effects on customers in a service failure setting. Following Smith *et al.* (1999) and Liu *et al.* (2023, p. 2), service failures are broadly characterized as "instances where the service provider's service fails to meet the expectations or needs of the consumer." This research focuses on service failures resulting from product performance issues, where the company is responsible for the failure (Pagiavlas, 2020; Wei and Shi, 2024). So-called functional failures, where the service failure occurs due to the use of a chatbot or voicebot interaction (i.e. where the bot malfunctions or provides incorrect information, Xing *et al.*, 2022), are not a focus for this research. Rather, the chatbot or voicebot is being used by the consumer as a means by which to rectify a perceived service failure event.

Informed by diverse theoretical approaches and methods, research on chatbots and voicebots in service failure contexts has documented both positive and negative effects. However, their comparative effects in service failure settings remain largely unexplored (see Table 1). Only a few studies have directly compared these modalities (Schindler *et al.*, 2024; Rohit *et al.*, 2024; Rzepka *et al.*, 2022; Zierau *et al.*, 2023) and, to the extent of our knowledge, none have yet examined them in the context of postpurchase service failure.

Most of these studies instead focus on earlier stages of the customer journey, prepurchase and purchase. For instance, Rohit *et al.* (2024) showed that voicebots, as a richer medium, enhance consumers' cognitive and affective engagement during prepurchase (e.g. information search) and purchase (e.g. booking) stages of the customer journey for hedonic products versus more utilitarian ones. Similarly, Schindler *et al.* (2024) found that voice (vs text) interactions promote an emotional (vs reason-based) focus, increasing satisfaction with hedonic (vs utilitarian) product choices, both at the prepurchase stage (e.g. hotel selection) and the purchase stage (e.g. food ordering). Rzepka *et al.* (2022) demonstrated that in goal-directed search tasks (e.g. finding a restaurant based on cuisine and price), consumers report higher satisfaction when using voicebots versus chatbots.

Regarding the postpurchase stage, Zierau *et al.* (2023) found that voicebots lead to more flow-like experiences than chatbots when filing insurance claims, resulting in better service evaluations, higher conversion rates and stronger retention. Collectively, these studies suggest that modality effects may vary across the customer journey (Grewal *et al.*, 2022; Lemon and Verhoef, 2016) and research directly comparing these technologies in service failure scenarios remains scarce.

Perceived relief

A critical component of service failure is the negative feelings customers experience in such situations (Gelbrich, 2010; Khamitov *et al.*, 2020). Thus, in a service failure context, a fundamental question is how to address such feelings (Balaji *et al.*, 2017; Lee, 2023; Tsarenko and Strizhakova, 2013). Even though research has emerged on bots' use of intuitive strategies to do so (e.g. chatbots' service recovery attempts that leverage apologies: Song *et al.*, 2023), there is a lack of understanding as to the differential effects of either text or voice human–robot interactions on consumer feelings derived from these strategies. Here, we contend that perceived relief, defined as the easing and smoothing of tension and negative

Table 1. Consumer research on text-based chatbots and voicebots: positive, negative and comparative effects

Study	Theoretical approach	Methodology	Key findings
Chatbots – positive effects			
Cheng *et al.* (2022)	Stimulus–organism–response (SOR)	Survey	In a service failure setting, empathy and friendliness of a chatbot have a positive effect on trust, which, in turn, has a positive effect on customer reliance. One moderator is task complexity: when a task is more complex, the positive effect of friendliness on trust is weaker
Liu *et al.* (2023)	Emotions expression	Experimental	The use of funny emojis by a chatbot has a positive effect on customer attitudes after a service failure. The effect is mediated by the perceived intelligence of the chatbot and is stronger for people with a growth mindset (incremental theory)
Agnihotri and Bhattacharya (2024)	Computers as social actors, information asymmetry	Mixed	Perceived empathy and anthropomorphism of the chatbot have a negative impact on negative word of mouth (nWOM) and a positive impact on customer forgiveness after a service failure. In contrast, privacy concerns when interacting with the chatbot have a positive impact on nWOM and a negative impact on customer forgiveness
Cai *et al.* (2024)	Mind perception	Experimental	A social-oriented (instead of task-oriented) communication style reduces negative emotions triggered by service failures, which in turn improves satisfaction with the service agent and company
Chatbots – negative effects			
Mozafari *et al.* (2020)	Attribution theory	Experimental	When customers are aware that they are interacting with a chatbot, trust decreases, particularly when the conversation revolves around a critical (vs. routine) service issue. Reduced trust leads to customer churn and negatively affects loyalty
Crolic *et al.* (2021)	Emotional appraisal	Field and experimental	Customers entering an interaction with a chatbot in an angry state experience greater dissatisfaction with the service and company if they perceive the chatbot as anthropomorphized
Huang and Dootson (2022)	Emotional coping	Experimental	After a service failure, a chatbot informing a customer of the possibility of interacting with a human at a later stage of their interaction can trigger aggressive responses from the customer. This effect is particularly stronger for those who perceive low customer participation in the interaction
Voicebots – positive effects			
Cuadra *et al.* (2021)	Conversational repair	Experimental	A voicebot capacity to self-correct (after, e.g. giving incorrect information) improves customers' assessment of the robot and interaction. This feature was more strongly valued by participants who had had recent experiences with errors
Lv *et al.* (2021)	Appraisal theory, performance expectancy	Experimental	A voicebot using a cute tone of voice leads to a higher tolerance for service failures. The effect is mediated by higher tenderness and lower performance expectancy and moderated by time pressure and failure severity

(*continued*)

Table 1. Continued

Study	Theoretical approach	Methodology	Key findings
Huang and Sénécal (2023)	Computers as social actors	Experimental	Warm verbal content used by a voicebot increases repatronage intention after a service failure. This positive effect is mediated by decreased frustration
Voicebots – negative effects			
Luo *et al.* (2019)	Negative biases towards Machines	Field	The disclosure of voicebots' identities as machines led to a drastic decrease in their effectiveness in outbound sales calls, as reflected in purchases. The effect was driven by a judgment of voicebots as less empathetic and less knowledgeable
Sun *et al.* (2022)	Cognitive load	Survey	Negative technical features (e.g. lack of perceived control resulting from cumbersome design and response delays) affect customer satisfaction and usage continuance intentions of voicebots (specifically, personal assistants) in the context of service failures. The effects are mediated by cognitive load
Li *et al.* (2023)	Self-attribution and performance expectancy	Field	A service failure by a voicebot increases the likelihood of customer complaint behavior. The effect is mediated by negative emotion
Chatbots versus voicebots – comparative effects			
Rzepka *et al.* (2022)	Technology fit theory	Experimental	Voicebots lead to higher satisfaction than chatbots in goal-directed search tasks. This effect is driven by greater perceived efficiency, lower cognitive effort and higher enjoyment
Zierau *et al.* (2023)		Experimental	Voicebots create more flow-like experiences than text-based chatbots, leading to better service evaluations and outcomes. These benefits diminish when interactions are semantically disfluent or involve too many conversational turns
Rohit *et al.* (2024)	Media richness	Experimental	In a retail setting, voicebots are more effective than chatbots at enhancing cognitive and affective engagement, especially for experiential products. The effect is moderated by localization or product customization according to location
Schindler *et al.* (2024)	Matching, congruence effects	Experimental	Text-based interactions (vs voice-based interactions) with a conversional agent encourage a reason-based (vs emotion-based) focus, leading to a greater preference for utilitarian (vs hedonic) products. This matching effect is mediated by processing fluency and is stronger for low-equity brands
This study	Emotion regulation	Experimental	In a service failure setting, chatbots lead to more positive customer reviews via increased perceived relief. The effect is stronger for low-severity failures and when outcomes are worse than expected

feelings (Kobel and Groeppel-Klein, 2021; Shurcliff, 1968), is more effectively evoked from interactions with chatbots, given the emotional regulation afforded to customers in interacting with these AI-agents. Our rationale for this prediction builds on the insights offered by emotion regulation theory, which focuses on the factors and strategies that can serve individuals control their emotional reactions (Gross, 1998). Importantly, individuals can engage in emotion regulation at an implicit or subconscious level, without making a deliberate attempt to manage their emotions (Koole *et al.*, 2015).

Specific emotion regulation mechanisms such as passive regulation through time, often referred to as "affective fading" (Skowronski *et al.*, 2014), and attentional deployment (Ferri and Hajcak, 2015; Gross, 2015) can explain why text-based (versus voice-based) interactions can be more conducive to perceived relief. Text is more asynchronous than voice, offering individuals additional time before responding: "Written conversational turns are […] more asynchronous […], so there is less of a rush to produce content" (Berger *et al.*, 2022, p. 391). Voice, by contrast, is faster and more immediate: "speech production is faster than writing because it does not involve the physical process of writing" (Rzepka *et al.*, 2022, p. 842), with "greater synchronicity due to fast-paced and more parallel processing with low revisability" (Zierau *et al.*, 2023, p. 827). Empirical evidence confirms this gap: speaking occurs at roughly 160 wpm versus 50–55 wpm for typing (Ruan *et al.*, 2018). This extra time matters because negative emotions such as anger tend to decrease almost immediately after onset and can even fade naturally within about 90 s if not retriggered (Kulper, 2016; Qiu *et al.*, 2009; Skowronski *et al.*, 2014; Taylor, 2021). Thus, the slower, more deliberate pace of text provides a natural buffer that facilitates affective fading.

In addition, writing requires more cognitive effort and deliberation than speaking: "Writing is more deliberative" (Berger *et al.*, 2022, p. 392); it is "relatively objective, structured, and detached" (Schindler *et al.*, 2024, p. 636); and "producing text involves a process of consciously applying intentionally learnt rules, while humans intuitively build and process speech" (Rzepka *et al.*, 2022, p. 843). This attentional deployment, i.e. redirecting cognitive resources to composing a message, can temporarily shift focus away from the emotional trigger, reducing arousal (Naragon-Gainey *et al.*, 2017). Consistent with these regulation mechanisms, Berger *et al.* (2022) showed that written word of mouth contains fewer emotionally charged words than spoken word of mouth, while Schindler *et al.* (2024) found that speaking about choices fosters an emotion focus, whereas writing fosters a reason focus. Across these mechanisms, the common outcome is the downregulation of negative affect and tension, which we conceptualize as perceived relief (Kobel and Groeppel-Klein, 2021).

It is further posited that these enhanced feelings of relief result in enhanced levels of customer service review, i.e. an overall assessment of the service provider that takes into account dimensions such as satisfaction and service quality (Crolic *et al.*, 2021). The ability to alleviate negative emotions plays a crucial role in shaping customer evaluations following a service failure (Nguyen and McColl-Kennedy, 2003). Research demonstrates that strategies aimed at reducing emotional distress lead to better customer outcomes, including increased brand loyalty and trust (Reeck and Onuklu, 2022). In service recovery contexts, mitigating negative emotion has been found to significantly improve customer perceptions of the service provider, as customers who experience emotional relief are less likely to hold negative attitudes toward the company (Lee, 2023). Given that perceived relief reflects the easing of negative emotions in response to a stressful event, it follows that greater perceived relief should translate into more favorable customer reviews.

Based on the above discussion, this research hypothesizes that, in a service failure setting, chatbots (versus voicebots) will be associated with increased customer review evaluations,

an effect that will be driven by increased perceived relief (i.e. the easing and smoothing of tension and negative feelings). Formally:

H1. In a service failure setting, chatbots (versus voicebots) have a positive effect on customer review.

H2. In a service failure setting, the positive effect of chatbots (versus voicebots) on customer review is mediated by increased perceived relief.

This work additionally advances the chatbot and voicebot fields by examining service failure severity as a moderator of this mediation effect.

Service failure severity

Not all service failures are perceived equally. Service failures can be differentiated along several dimensions, such as responsibility (Leo and Huh, 2020), frequency and timing (Sivakumar *et al.*, 2014) and severity (Leo and Huh, 2020; Weun *et al.*, 2004). In this work, one of the foci is the severity of the failure, defined as the perceived criticality or intensity of the service problem (Weun *et al.*, 2004), whereby more severe failures result in more intense negative feelings, which lead to decreased customer satisfaction (Smith *et al.*, 1999). Service failure severity has been frequently used as a moderator in service research (de Mesquita *et al.*, 2023; Fouroudi *et al.*, 2020) and also operationalized dichotomously, as done here (e.g. Fox *et al.*, 2018; Hess *et al.*, 2003). In the context of robots, for instance, Shams *et al.* (2024) established that humor and informal language can be successfully used by text-based chatbots in service recovery attempts as long as the severity of a failure is low, an effect that is reversed for high-severity failures. Furthermore, Lv *et al.* (2021) found that voicebots using a cute voice increase customers' tolerance of service failures, but again, only when the severity of the failure is low.

Building on the preceding section, we argue that the effect of modality (text vs voice) on perceived relief is contingent on the severity of the service failure. Emotion regulation theory (Gross, 1998, 2015) suggests that individuals deploy regulation strategies based on situational appraisals, such as urgency or threat (Matthews *et al.*, 2021). When service failures are minor, customers can retain both the motivation and cognitive capacity to regulate their emotions, as suggested before (Ferri and Hajcak, 2015; Skowronski *et al.*, 2014).

However, in high-severity situations, customers are likely to deprioritize emotional regulation in favor of swift problem resolution (Gross, 2015). When urgency is high, individuals are less inclined to invest cognitive resources in regulating emotions, especially when those emotions may functionally support goal attainment, in this case, problem resolution (Tamir and Ford, 2009; Glikson *et al.*, 2019). As cognitive demands rise as a result of a more severe failure, the regulatory benefits of chatbots, which require effortful input (Tindle and Longstaff, 2015), become less accessible. Voice-based interactions, while less conducive to relief under low-severity conditions, due to their lower cognitive demands, may become more intuitive when mental resources are depleted (Sweller *et al.*, 2011; Hawthorne *et al.*, 2019). Consequently, the relief advantage of chatbots is expected to diminish as failure severity increases.

In essence, when customers face severe service failures, emotional intensity is so high and resolution so urgent that neither the slower, self-paced nature of text nor the lower-effort fluidity of voice holds a meaningful edge; what matters is resolving the issue. Thus, we predict:

H3. The positive indirect effect of chatbot (versus voicebot) modality on customer review, via perceived relief, is only present when the service failure is not severe. When the service failure is severe, there is no difference between chatbot and

voicebot modality on perceived relief, so no modality impacts customer review differently.

Figure 1 demonstrates the conceptual model.

Overview of studies

To empirically investigate our research questions and test the proposed hypotheses, we used a multistudy research design comprising one field study and two controlled experiments. Study 1 serves as an initial exploratory field study conducted in a real-world service failure setting, aimed at establishing the baseline effect of AI bot modality on customer evaluations. This naturalistic study provides ecological validity and grounds the research in actual consumer behavior. More importantly, the initial field study was particularly important in establishing the real-world relevance of the phenomenon, namely, that differences in customer responses to AI bot modalities are not only theoretically meaningful but also practically prevalent for online service firms. As many field studies, it provides preliminary evidence for the effect of modality on important customer outcomes. Study 2 is an online experiment designed to test the underlying psychological mechanism (i.e. perceived relief) that mediates the relationship between AI bot modality and customer review. Finally, Study 3 examines the moderating role of service failure severity. Together, these studies provide a comprehensive and theory-driven examination of why and when chatbot (vs voicebot) interactions enhance customer outcomes in service failure contexts.

Study 1

To explore whether the effect of AI bot modality (chatbot vs voicebot) is observable in real-world service settings, Study 1 provides preliminary, ecologically valid evidence for *RQ1*

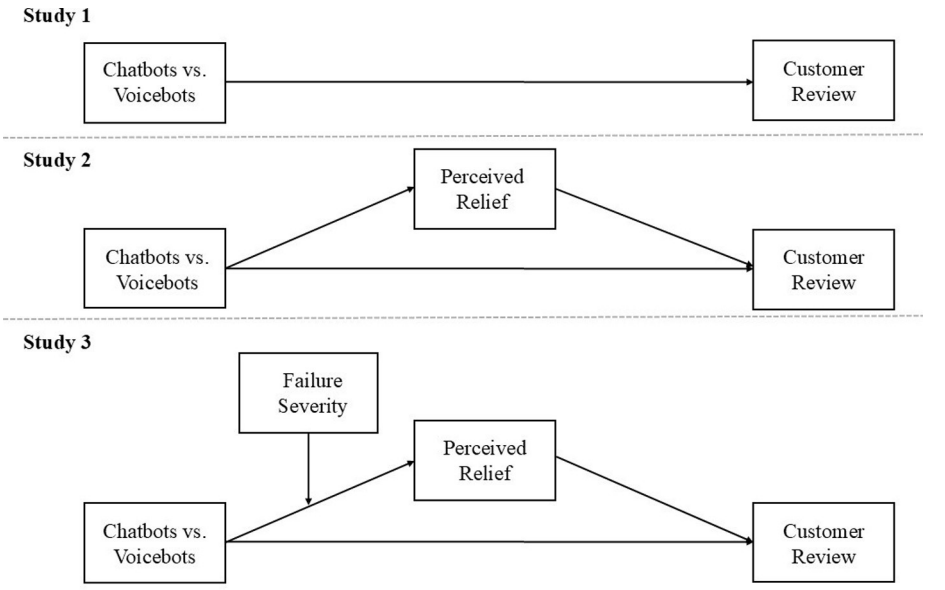

Figure 1. Conceptual model

and *H1*. Specifically, it examines whether customers interacting with a chatbot report higher satisfaction ratings than those interacting with a voicebot following a postpurchase service experience. As a naturalistic field setting, Study 1 necessarily involved channel differences (Web vs phone, as indicated below) and customer self-selection, which may confound modality effects. Rather than providing definitive causal evidence, the purpose of Study 1 is to illustrate that meaningful differences can emerge in practice, thereby motivating the controlled laboratory experiments in Studies 2 and 3.

Sample, procedure and measures

We conducted a field experiment in collaboration with a service firm within the retail IT sector. The service firm operates as a local computer dealer located in the capital city of an eastern province in China and provides both in-store and online retail services (preregistered: https://aspredicted.org/xtqt-m73h.pdf). As part of their postsales customer service infrastructure, the company uses two AI-driven service agents: a text-based chatbot embedded in their online service portal and a voicebot integrated into their telephone hotline system for any product complaint. These tools are used to assist customers with troubleshooting product issues, obtaining service information and resolving complaints outside of normal business hours. The aim of the field experiment was to examine how different AI bot modalities (i.e. chatbot vs voicebot) affect customer reviews in a service failure context. As such, this real-world setting allowed us to observe how bot modality influences customer evaluations under ecologically valid conditions. The experiment was embedded within the IT service provider's existing postsales service framework and ran for one month. Customers who encountered computer-related issues could visit the company's service website and interact with an AI-driven chatbot (text-based interaction) by clicking the "Online Customer Service" option displayed on the webpage. Alternatively, they could also call the customer service hotline (the number was provided beneath the "Online Customer Service" option) and interact with an AI-based voicebot (voice-based interaction).

Participation was naturalistic and nonintrusive such that customers were not informed that an experiment was being conducted and self-selected into one of the two AI bot modalities based on their preferred channel of service. After the interaction with the chatbot and the voicebot, customers were prompted to provide a satisfaction rating on a five-star scale (1 = very dissatisfied, 5 = very satisfied), specifically reflecting their assessment of the AI agent's ability to respond to their service request and manage the recovery interaction, by selecting the stars on the webpage or clicking on a number on phone's dial keypad (see Appendix). These customer ratings served as the primary dependent variable.

Customer responses were automatically recorded through the company's website and phone system, depending on the modality used. The field experiment lasted one month, and a total of 138 customer ratings (61 from voicebot and 77 from chatbot) were received. As this was a naturalistic field setting, no participants were excluded, and no outlier removal was conducted. The data reflects real customer behavior in an ecologically valid service context.

Results

We conducted an independent sample *t*-test to examine whether and how AI bot modality affects consumer ratings. As expected, consumers who interacted with chatbot in the Web page ($M_{Chatbot} = 4.12$, $SD_{Chatbot} = 0.628$) rated the AI services significantly higher than those who chatted with AI via the hotline ($M_{Voicebot} = 3.03$, $SD_{Voicebot} = 0.795$; $p < 0.001$, Cohen's $d = 1.54$). Thus, this finding offers preliminary field-based evidence for *H1*, demonstrating that bot modality differences manifest in real customer evaluations. Given the ecological nature of this setting and the potential for confounds (e.g. channel and severity), the result is

best interpreted as suggestive evidence that motivates further controlled studies. Importantly, this study provides a strong rationale for systematically testing the underlying mechanisms (e.g. perceived relief) and boundary conditions (e.g. failure severity and outcome expectations) in Studies 2 and 3.

Study 2

Building on the findings of Study 1, which confirmed that chatbot interactions yield more favorable customer evaluations, Study 2 addresses RQ2 and empirically tests *H2* by investigating the underlying psychological mechanism of perceived relief that may explain this effect.

Sample and procedure

We recruited 160 participants (48% females, $M_{age} = 33$, SD = 13) located in the USA from Prolific. Participants were randomly assigned to a 2× AI-driven bot (chat vs voice) modality between-subjects design.

Participants were provided with a service scenario where they were required to imagine that they had bought a camera specifically for an upcoming trip. Upon receiving the camera, they excitedly open it and snap a few pictures, only to realize they are disappointingly blurry. They suspect the lens might be cracked or misaligned. They search online for guidance on how to exchange the camera, only to find that the estimated delivery date for a replacement falls after the departure date for their trip, defeating the whole purpose of the purchase. Frustration sets in, and they contemplate reaching out to the camera company. Unfortunately, it is outside of their business hours, so they opt to explore their online customer service center instead. In the scenario with the AI chatbot, a window pops up, and there is a conversation with the chatbot about how to fix the issue. In the scenario of a voicebot, the participants discover the phone number for the company's Automated Customer Service Center, they make the call, and listen to the conversation with a voicebot. The full scenario text and AI bot interaction transcripts are provided in the Appendix.

Measures

All measures were adapted from established scales. For perceived relief (Kobel and Groeppel-Klein, 2021), participants were asked to indicate the extent to which they agree or disagree on a five-point scale with the following statements ($\alpha = 0.92$): "The bot relaxed the situation," "I perceived the bot as liberating," "The bot lightened the atmosphere," "The bot defused the situation" and "The bot enhanced the mood." Customer review was measured by asking the respondents to evaluate the service provided by the bot: overall satisfaction, customer service, helpfulness and problem resolution (Crolic *et al.*, 2021), measured on a five-point scale (1 = 1 star, 5 = 5 stars).

We conducted Confirmatory Factor Analyses (CFAs) to ensure the robustness of measurement instruments. The data were evaluated for convergent and discriminant validity (see Appendix). The values of Cronbach's α, Composite Reliability (CR) and Average Variance Extracted (AVE) were all greater than the cut-off of 0.7, 0.7 and 0.5, respectively (Hair *et al.*, 2019). Thus, we confirm that our measures met the reliability and convergent validity criteria. As depicted in the correlation matrix (see Appendix), the square root values of AVE (on the diagonals) are greater than the interconstruct correlational values (on the left of the diagonals), thereby indicating that discriminant validity was verified (Fornell and Larcker, 1981). To address potential common method bias, we compared the measurement model (i.e. perceived relief and customer review) against a single-factor model combining both perceived relief and customer review together and used the Unmeasured Latent Method

Construct (ULMC) approach by incorporating an unmeasured latent variable connected to all observed items on the basis of the measurement model (Podsakoff *et al.*, 2012). Across all studies, the fit indices for the measurement model (χ^2/df = 1.801, CFI = 0.980, TLI = 0.973, SRMR = 0.033) were significantly better than those for the single-factor model (χ^2/df = 4.743, CFI = 0.902, TLI = 0.874, SRMR = 0.054) and were not significantly different from those for the measurement model incorporating an unmeasured latent variable (χ^2/df = 1.478, CFI = 0.988, TLI = 0.984, SRMR = 0.050). These results suggest that common method bias did not pose a significant issue in any of the studies. Overall, these results suggest that we could proceed with hypotheses testing.

Pretest

We conducted a separate pretest with 101 Prolific participants based in the USA, who were randomly assigned to one of two AI bot modality conditions (chatbot vs voicebot) in a between-subjects design. Participants evaluated the scenario's believability and clarity on a five-point scale (1 = not at all, 5 = very much). They also rated the extent to which they would feel frustrated when encountering the described issue (1 = not at all, 5 = very much).

As expected, we found no significant differences between conditions; participants in both the chatbot and voicebot groups perceived the scenario as similarly believable [M_{chat} = 3.73, M_{voice} = 3.58, $t(99) = 0.75$, $p = 0.456$] and clear [M_{chat} = 4.37, M_{voice} = 4.26, $t(99) = 0.74$, $p = 0.460$]. In addition, reported frustration levels were comparable across conditions [M_{chat} = 4.29, M_{voice} = 4.40, $t(99) = 0.63$, $p = 0.530$]. These results suggest that believability, clarity and frustration were unlikely to introduce potential confounds.

Results

An independent sample *t*-test revealed that participants in the chatbot condition ($M = 3.83$) reported higher levels of customer review than those in the voicebot condition [$M = 3.53$, $t(158) = 2.15$, $p = 0.033$], further supporting *H1*.

We proposed that perceived relief mediates the impact of AI bot modality on customer review. An independent sample *t*-test revealed that participants in the chatbot condition ($M = 3.47$) reported higher levels of perceived relief than those in the voicebot condition [$M = 3.14$, $t(158) = 2.18$, $p = 0.031$]. A mediation analysis was performed (Hayes, 2017, Model 4). This analysis examined the indirect effects of AI bot modality types as a categorial variable (voicebot = 0, chatbot = 1), on customer review via perceived relief. AI bot modality types had a significant impact on perceived relief. Perceived relief was a significant predictor of customer review. The mediation effect was significant ($b = 0.24$, SE = 0.11, 95% CI: 0.03–0.45). See Figure 2 and Table 2. Thus, *H2* was supported.

Study 3

While Study 2 establishes the mediating role of perceived relief, it does so in a generalized service failure context. To further refine the conditions under which this mediation holds, Study 3 addresses RQ3 and empirically tests *H3* by examining whether the severity of the service failure moderates the indirect effect of bot modality on customer evaluations via perceived relief.

Sample and procedure

We recruited 320 participants (50% females, M_{age} = 43, SD = 14) located in the USA from Prolific. Participants were randomly assigned to a 2 AI bot modality (chat vs voicebot) × 2 service failure severity (high vs low) between-subjects design.

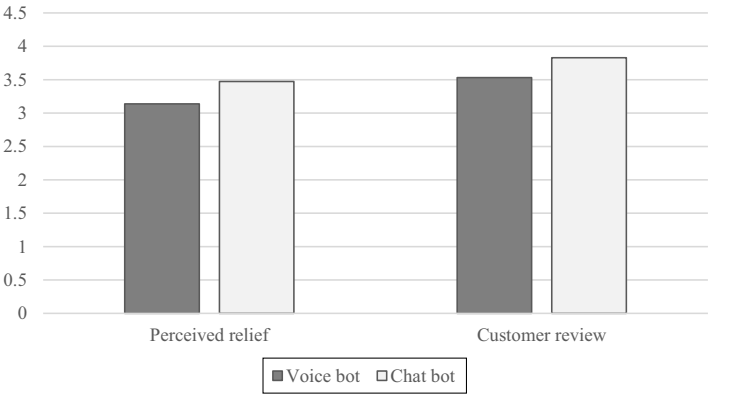

Figure 2. Perceived relief and customer review, chatbots versus voicebots

Table 2. Mediation analysis (Study 2)

	Perceived relief (*M*)				Customer review (*Y*)			
Variables	Coeff	SE	*t*	*p*	Coeff	SE	*z*	*p*
Constant	3.138	0.108	29.051	< 0.001	1.316	0.154	8.530	< 0.001
AI bot modality (*X*)	0.879	0.154	2.180	0.031	0.062	0.088	0.698	0.486
Perceived relief (*M*)	–	–	–	–	0.706	0.045	15.640	0.000

In the high-severity scenario, participants were required to imagine that they had recently purchased a brand-new refrigerator, and after just a few days of use, it had completely stopped working. Their food was at risk of spoiling, and they were facing an urgent situation. They decided to reach out to the refrigerator company for assistance. In the low-severity scenario, participants were required to imagine that they had recently bought a new refrigerator, but they have been experiencing intermittent problems with its cooling system. Sometimes, the temperature seemed fine, but on other occasions, they noticed that their food was not as cold as it should be. It was not an immediate emergency, but a concern, and they wanted to address it with the refrigerator company. Unfortunately, it was outside of their business hours, so they opted to explore their online customer service center. Participants either saw a conversation with a chatbot or listened to a conversation with a voicebot. The full scenario description and AI bot interaction transcripts are provided in the Appendix.

Measures

Perceived relief and customer review were measured with validated scales as adopted in Study 2. In a manner parallel to Study 2, we conducted the CFA tests and the results showed that the computed values for Cronbach's α, CR and AVE significantly surpassed the criteria of 0.7, 0.7 and 0.5, respectively. The discriminant validity was also achieved, as the square root values of AVE were greater than the interconstruct correlations (see Appendix). We also performed a comparison among three distinct models: the two-factor measurement model, the single-factor model and the three-factor model, introducing an unmeasured latent variable (ULMC). The fit indices of the two-factor measurement model ($\chi^2/df = 3.630$,

CFI = 0.976, TLI = 0.967, SRMR = 0.028) exhibited marked improvement when compared to the single-factor model (χ^2/df = 24.424, CFI = 0.780, TLI = 0.707, SRMR = 0.095), while the model incorporating an unmeasured latent variable (χ^2/df = 2.904, CFI = 0.984, TLI = 0.976, SRMR = 0.026) yielded similar fit indices. Thus, common method bias was not a significant concern.

Pretest

We conducted a separate pretest with 160 Prolific participants based in the USA, who were randomly assigned to a 2 AI bot modality (chat vs voicebot) × 2 service failure severity (high vs low) between-subjects design. As in Study 2, participants evaluated the scenario's believability and clarity. They also rated the extent to which they would feel frustrated when encountering the described issue, as well as the perceived severity of the issue. All items were measured on a five-point scale (1 = not at all, 5 = very much).

We conducted 2 × 2 ANOVAs and found no significant differences across conditions in how participants perceived the scenario's believability and clarity. Specifically, we found nonsignificant main effects of AI bot modality [believability: $F(1, 156) = 1.62$, $p = 0.204$; clarity: $F(1, 156) = 0.72$, $p = 0.398$], service failure severity [believability: $F(1, 156) = 0.78$, $p = 0.379$; clarity: $F(1, 156) = 0.37$, $p = 0.546$] and interaction effects [believability: $F(1, 156) = 2.16$, $p = 0.143$; clarity: $F(1, 156) = 0.72$, $p = 0.398$].

However, we found significant main effects of service failure severity on frustration and perceived severity. Specifically, participants in the high-severity condition reported higher levels of frustration [$M_{high} = 4.50$, $M_{low} = 3.52$, $F(1, 156) = 37.11$, $p < 0.001$] and perceived the issue as more severe [$M_{high} = 4.11$, $M_{low} = 3.28$, $F(1, 156) = 35.48$, $p < 0.001$] than those in the low-severity condition. However, the main effects of AI bot modality [frustration: $F(1, 156) = 0.10$, $p = 0.755$; perceived severity: $F(1, 156) = 2.28$, $p = 0.133$] as well as the interaction effects were nonsignificant [frustration: $F(1, 156) < 0.001$, $p = 1.00$; perceived severity: $F(1, 156) = 0.96$, $p = 0.330$].

Results

Customer review. Setting customer review as the dependent variable, a two-way ANOVA analysis was performed with independent variables including AI bot modality and service failure severity and their interaction. Results revealed a nonsignificant main effect of AI bot modality [$F(1, 315) = 1.32$, $p = 0.252$] and a significant main effect of service failure severity [$F(1, 315) = 51.92$, $p < 0.001$]. Moreover and as predicted, we also found a significant interaction effect [$F(1, 315) = 4.23$, $p = 0.041$]. Specifically, in the low-severity condition, participants in the chatbot condition ($M = 4.22$) reported higher levels of customer review than those in the voicebot condition [$M = 3.89$, $F(1, 315) = 5.12$, $p = 0.024$]. However, such differences were nonsignificant in the high-severity condition [$M_{chat} = 3.25$, $M_{voice} = 3.35$, $F(1, 315) = 0.42$, $p = 0.520$].

Perceived relief. Setting perceived relief as the dependent variable, a two-way ANOVA analysis was performed with independent variables including AI bot modality and service failure severity and their interaction. Results revealed a nonsignificant main effect of AI bot modality [$F(1, 315) = 2.71$, $p = 0.101$] and a significant main effect of service failure severity [$F(1, 315) = 35.48$, $p < 0.001$]. Moreover, and as predicted, we also found a significant interaction effect [$F(1, 315) = 4.85$, $p = 0.028$]. Specifically, in the low-severity condition, participants in the chatbot condition ($M = 3.49$) reported higher levels of perceived relief than those in the voicebot condition [$M = 3.07$, $F(1, 315) = 7.38$, $p = 0.007$]. However, such differences were nonsignificant in the high-severity condition [$M_{chat} = 2.58$, $M_{voice} = 2.65$, $F(1, 315) = 0.16$, $p = 0.693$].

Moderated mediation analysis. We propose that the AI bot modality types on customer reviews are mediated by perceived relief and moderated by service failure severity. A moderated mediation analysis (Hayes, 2017, Model 7) was conducted. AI bot modality types as a categorial variable (voicebot = 0, chatbot = 1) affect customer review, moderated by service failure severity (low severity = 0, high severity = 1), via perceived relief.

AI bot modality significantly predicted perceived relief, and in turn, perceived relief significantly predicted customer review. The index of moderated mediation was significant and negative ($b = -0.33$, SE = 0.15, 95% CI: −0.63 to −0.04), such that the indirect effect was significant in the low-severity condition ($b = 0.29$, SE = 0.10, 95% CI: 0.09–0.49) but not in the high-severity condition ($b = -0.04$, SE = 0.11, 95% CI: −0.27–0.18). See Table 3. *H3* was supported.

Discussion

Theoretical implications

This research extends existing theory on AI service interactions, specifically human–robot interactions, by demonstrating that chatbots and voicebots elicit different consumer responses in service failures, with chatbots (versus voicebots) significantly increasing perceived relief and, via this mechanism, having a positive influence on customer reviews. In addition, this research advances theoretical understanding by identifying a key moderator of this effect. Specifically, we show that the severity of the service failure critically determines the impact of bot modality on perceived relief and, ultimately, customer evaluations.

These findings make significant contributions to several research streams. First, we advance research on AI-driven customer service by providing one of the first empirical comparisons between bot modalities (i.e. chatbots and voicebots). While prior research has explored each modality separately (e.g. Agnihotri and Bhattacharya, 2024; Cai *et al.*, 2024; Lv *et al.*, 2021; Li *et al.*, 2023), our study is one of the first to systematically examine their differential effects in service failures. While a few studies have compared these modalities (e.g. Rohit *et al.*, 2024; Rzepka *et al.*, 2022; Schindler *et al.*, 2024), they did not examine their effects in failure scenarios, where emotional responses are heightened and customer expectations shift dramatically (Obeidat *et al.*, 2017; Valentini *et al.*, 2020).

Second, this research makes a novel contribution to the service failure literature by directly comparing the two prevailing AI bot modality interactions (i.e. chatbot and voicebot), an empirical contrast largely absent in prior service failure studies, which have typically examined these modalities in isolation (Agnihotri and Bhattacharya, 2024; Crolic *et al.*, 2021; Huang and Sénécal, 2023; Li *et al.*, 2023). In doing so, this study challenges prevailing assumptions in service failure research, which often suggest that more humanlike or richer media, such as voice-based interfaces, are inherently more effective in emotionally charged service contexts. While anthropomorphizing AI service agents has been widely

Table 3. Moderated mediation analysis (Study 3)

	Perceived relief (M)				Customer review (Y)			
Variables	Coeff	SE	t	p	Coeff	SE	t	p
Constant	3.065	0.111	27.561	< 0.001	1.706	0.126	13.548	< 0.001
AI bot modality (X)	0.429	0.158	2.717	0.007	−0.001	0.083	−0.012	0.990
Service failure sev. (W)	−0.418	0.157	−2.667	0.008	–	–	–	–
$X \times W$	−0.491	0.223	−2.203	0.028	–	–	–	–
Perceived relief (M)	–	–	–	–	0.669	0.039	17.075	< 0.001

regarded as an effective strategy (de Visser *et al.*, 2016; Huang and Sénécal, 2023; So *et al.*, 2024), our findings suggest that in certain circumstances, less humanlike interactions, such as text-based chatbots, can be more beneficial in mitigating negative emotional reactions (Crolic *et al.*, 2021; Li *et al.*, 2023). This finding contradicts the assumption that richer, more immersive media (such as voice; Rohit *et al.*, 2024) are inherently superior in emotionally charged service contexts.

Third, this study extends emotion regulation theory by identifying bot modality as a factor that influences how consumers regulate emotions in service failures. Service failures inherently elicit negative emotions (Balaji *et al.*, 2017), yet prior research has barely examined how different communication modalities facilitate or hinder emotion regulation during these events. While Berger *et al.* (2022) showed that consumers express more emotionality when speaking rather than writing in product reviews, this research extends that insight to service failure settings, demonstrating that chatbots (text) are more conducive to emotion regulation, as evidenced by increased perceived relief. Our findings are consistent with the idea that chat-based interactions facilitate mechanisms such as affective fading and attentional deployment (Ferri and Hajcak, 2015; Skowronski *et al.*, 2014).

Finally, this work adds to the literature exploring the construct of perceived consumer relief. In service failure settings, this construct has received limited attention (Kobel and Groeppel-Klein, 2021). Kobel and Groeppel-Klein (2021) investigated how humorous remarks lead to relief, demonstrating that content, i.e. *what* a service provider says, positively influences perceived relief. Here, it has been established that perceived relief is influenced not only by content but also by *form*, i.e. not *what* is said, by either the customer or service provider, but also *how* it is said, or the medium chosen for communication (text or voice).

By identifying why and when chatbots outperform voicebots in service failure settings, this research provides a novel understanding of how different AI modalities influence customer emotions and evaluations. It also lays the foundation for future work exploring the wider impact of AI communication modality across various service settings.

Managerial implications

The findings of this study provide valuable insights for managers seeking to optimize customer service strategies, particularly in the context of service failures. The findings indicate that chatbots (text-based) lead to increased perceived relief and customer review compared to voicebots, especially in instances of nonsevere service failures. Therefore, managers should consider deploying chatbots as a response mechanism in these scenarios to improve customer outcomes.

Several factors might encourage the adoption of voicebots to the detriment of chatbots. First, excitement surrounding the adoption of voicebot technology due to its complexity (e.g. its use of natural language processing and speech recognition) might favor voicebots. The extensive media coverage of ChatGPT-o's voice capabilities, which makes the LLM sound remarkably human, provides a clear example (Scharth, 2024). Second, there is a strong tendency (which, presumably, extends beyond the academic sphere) to assume that interactions with robots should be "humanized", i.e. made to approximate, to the largest extent possible, the complexities of human-to-human exchanges (Crolic *et al.*, 2021). Speech is an inherently richer form of communication than text, so voicebots once again look like a promising route to pursue in this regard (Luo *et al.*, 2019; Rohit *et al.*, 2024). The findings of this study suggest that this approach may not always yield the best results and that, indeed, in service failure settings (and, particularly, those involving low-severity failures), chatbots are the optimal strategy.

Based on the findings of this study, it would be a stretch to conclude that chatbots are invariably better. The dynamic nature of customer preferences and technological advancements necessitates continuous evaluation and adaptation of customer service strategies. Managers should regularly review the performance of both chatbots and voicebots, incorporating customer feedback and technological improvements to optimize their service delivery approaches. By understanding and applying these insights, managers can better navigate service failures, ultimately leading to improved customer outcomes.

Limitations and future research

This study has several limitations, and future research directions are suggested. First, it is important to note that while this research found that chatbots are preferable under certain circumstances, it cannot be concluded that they are invariably better. Future research should examine alternative moderators that might influence these preferences. For instance, process and outcome failures emphasize, respectively, the service delivery or its result (e.g. a product) (Fouroudi *et al.*, 2020). The scenarios chosen in this study mainly refer to outcome failures (camera and refrigerator). It would be interesting to explore if process failures, where the problem lies in the way the service is delivered rather than the end result, negate or reverse the patterns found in this research.

Second, consistent with prior research on service failures (e.g. Fox *et al.*, 2018; Shams *et al.*, 2024), we used a dichotomous manipulation of failure severity. Future research could extend this approach by using a continuous measure to capture finer variations in how bot modality interacts with failure severity.

Third, while this study explored a diverse set of products, future research could further investigate the role of product value in shaping customer expectations and responses to AI service interactions. High-value products, particularly those in luxury markets, may have customers who expect personalized, 24-h human service availability (Kapferer and Bastien, 2012), which could reduce reliance on chatbots and voicebots for service interactions. However, as luxury brands begin integrating AI-driven service tools (Chung *et al.*, 2020), understanding the boundary conditions under which such technologies influence the customer experience represents an important avenue for future research.

Fourth, one key limitation of this study is the lack of a pre- and postmeasurement of frustration, which could have provided deeper insights into how emotional responses evolve over time in reaction to service failures and AI-driven resolutions. While our findings demonstrate that frustration levels vary based on severity assessments, capturing frustration before and after the intervention would allow for a more precise assessment of AI-driven recovery strategies. Future research could adopt a longitudinal or real-time tracking approach to examine moment-to-moment changes in frustration and their subsequent impact on customer perceptions, satisfaction and loyalty.

In addition, while our randomized experimental design controls for individual differences, future studies could further explore the role of personal factors such as technology familiarity, AI aversion and service expectations, which may moderate consumers' emotional and behavioral responses to AI interactions.

Finally, based on prior work (Crolic *et al.*, 2021), this study used an overall measure of satisfaction as the dependent variable. Future research might examine similar effects with other dependent variables, such as customer loyalty, trust or perceived value. Investigating these additional dimensions could provide a more comprehensive understanding of how chatbots and voicebots influence various aspects of customer experience and behavior.

References

Agnihotri, A. and Bhattacharya, S. (2024), "Chatbots' effectiveness in service recovery", *International Journal of Information Management*, Vol. 76, p. 102679, doi: 10.1016/j.ijinfomgt.2023.102679.

Akanksha (2024), "Voicebot vs chatbot: which is better for your business? CZENTRIX", available at: www.c-zentrix.com/blog/voicebot-vs-chatbot/1000

Balaji, M.S., Roy, S.K. and Quazi, A. (2017), "Customers' emotion regulation strategies in service failure encounters", *European Journal of Marketing*, Vol. 51 Nos 5-6, pp. 960-982, doi: 10.1108/EJM-03-2015-0169.

Begum, A. (2024), "Voicebots Market - Forecast(2024 - 2030)", available at: www.linkedin.com/pulse/voicebots-market-forecast2024-2030-aamira-begum-i8qgc/

Belanche, D., Casaló, L.V., Flavián, C. and Schepers, J. (2020), "Service robot implementation: a theoretical framework and research agenda", *The Service Industries Journal*, Vol. 40 Nos 3-4, pp. 203-225, doi: 10.1080/02642069.2019.1672666.

Berger, J., Rocklage, M.D. and Packard, G. (2022), "Expression modalities: how speaking versus writing shapes word of mouth", *Journal of Consumer Research*, Vol. 49 No. 3, pp. 389-408, doi: 10.1093/jcr/ucab076.

Bezrukava, P. (2023), "Microsoft, Starbucks, Bank of America: how global companies are leveraging voice bots in customer service", available at: https://bsg.world/blog/voice-bot-real-life-examples/#:~:textinutes%20to%20read-Microsoft%2C%20Starbucks%2C%20Bank%20of%20America%3A%20How%20Global%20Companies%20Are,Voice%20Bots%20in%20Customer%20Service&text=oice%20bots%20have%20ushered%20in,reaching%20%2426.8%20billion%20by%202025

Bilquise, G., Ibrahim, S. and Shaalan, K. (2022), "Emotionally intelligent chatbots: a systematic literature review", *Human Behavior and Emerging Technologies*, Vol. 2022, pp. 1-9601630, doi: 10.1155/2022/9601630.

Cai, N., Gao, S. and Yan, J. (2024), "How the communication style of chatbots influences consumers' satisfaction, trust, and engagement in the context of service failure", *Humanities and Social Sciences Communications*, Vol. 11 No. 1, p. 687, doi: 10.1057/s41599-024-03212-0.

Cheng, X., Bao, Y., Zarifis, A., Gong, W. and Mou, J. (2022), "Exploring consumers' response to text-based chatbots in e-commerce: the moderating role of task complexity and chatbot disclosure", *Internet Research*, Vol. 32 No. 2, pp. 496-517, doi: 10.1108/INTR-08-2020-0460.

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

Crolic, C., Thomaz, F., Hadi, R. and Stephen, A.T. (2021), "Blame the bot: anthropomorphism and anger in customer–chatbot interactions", *Journal of Marketing*, Vol. 86 No. 1, pp. 132-148, doi: 10.1177/00222429211045687.

Cuadra, A., Li, S., Lee, H., Cho, J. and Ju, W. (2021), "My bad! Repairing intelligent voice assistant errors improves interaction", *Proceedings of the ACM on Human – Computer Interaction*, Vol. 5 No. CSCW1, Article, p. 27, doi: 10.1145/3449101.

de Mesquita, J.M.C., Shin, H., Urdan, A.T. and Pimenta, M.T.C. (2023), "Measuring the intention-behavior gap in service failure and recovery: the moderating roles of failure severity and service recovery satisfaction", *European Journal of Marketing*, Vol. 57 No. 7, pp. 1826-1853, doi: 10.1108/EJM-03-2022-0235.

de Visser, E.J., Monfort, S.S., McKendrick, R., Smith, M.A., McKnight, P.E., Krueger, F. and Parasuraman, R. (2016), "Almost human: anthropomorphism increases trust resilience in cognitive agents", *Journal of Experimental Psychology: Applied*, Vol. 22 No. 3, pp. 331-349, doi: 10.1037/xap0000092.

Dinh, C.-M. and Park, S. (2023), "How to increase consumer intention to use chatbots? An empirical analysis of hedonic and utilitarian motivations on social presence and the moderating effects of fear across generations", *Electronic Commerce Research*, Vol. 24 No. 4, doi: 10.1007/s10660-022-09662-5.

Ferri, J. and Hajcak, G. (2015), "Neural mechanisms associated with reappraisal and attentional deployment", *Current Opinion in Psychology*, Vol. 3, pp. 17-21, doi: 10.1016/j.copsyc.2015.01.010.

Fornell, C. and Larcker, D.F. (1981), "Evaluating structural equation models with unobservable variables and measurement error", *Journal of Marketing Research*, Vol. 18 No. 1, pp. 39-50.

Fouroudi, P., Kitchen, P.J., Marvi, R., Akarsu, T.N. and Uddin, H. (2020), "A bibliometric investigation of service failure literature and a research agenda", *European Journal of Marketing*, Vol. 54 No. 10, pp. 2575-2619, doi: 10.1108/EJM-07-2019-0588.

Fox, A.K., Deitz, G.D., Royne, M.B. and Fox, J.D. (2018), "The face of contagion: consumer response to service failure depiction in online reviews", *European Journal of Marketing*, Vol. 52 Nos 1-2, pp. 39-65, doi: 10.1108/EJM-12-2016-0887.

García-Méndez, S., Arriba-Pérez, F.D., González-Castaño, F.J., Regueiro-Janeiro, J.A. and Gil-Castiñeira, F. (2021), "Entertainment chatbot for the digital inclusion of elderly people without abstraction capabilities", *IEEE Access*, Vol. 9, pp. 75878-75891, doi: 10.1109/ACCESS.2021.3080837.

Gelbrich, K. (2010), "Anger, frustration, and helplessness after service failure: coping strategies and effective informational support", *Journal of the Academy of Marketing Science*, Vol. 38 No. 5, pp. 567-585, doi: 10.1007/s11747-009-0169-6.

Glikson, E., Rees, L., Wirtz, J., Kopelman, S. and Rafaeli, A. (2019), "When and why a squeakier wheel gets more grease: the influence of cultural values and anger intensity on customer compensation", *Journal of Service Research*, Vol. 22 No. 3, pp. 223-240, doi: 10.1177/1094670519838623.

Grewal, D., Guha, A., Schweiger, E., Ludwig, S. and Wetzels, M. (2022), "How communications by AI-enabled voice assistants impact the customer journey", *Journal of Service Management*, Vol. 33 Nos 4-5, pp. 705-720, doi: 10.1108/JOSM-11-2021-0452.

Gross, J.J. (1998), "The emerging field of emotion regulation: an integrative review", *Review of General Psychology*, Vol. 2 No. 3, pp. 271-299, doi: 10.1037/1089-2680.2.3.271.

Gross, J.J. (2015), "Emotion regulation: current status and future prospects", *Psychological Inquiry*, Vol. 26 No. 1, pp. 1-26, doi: 10.1080/1047840X.2014.940781.

Hair, J.F., Risher, J.J., Sarstedt, M. and Ringle, C.M. (2019), "When to use and how to report the results of PLS-SEM", *European Business Review*, Vol. 31 No. 1, pp. 2-24.

Harrison-Walker, L. (2012), "The role of cause and affect in service failure", *Journal of Services Marketing*, Vol. 26 No. 2, pp. 115-123, doi: 10.1108/08876041211215275.

Harrison-Walker, L.J. (2019), "The effect of consumer emotions on outcome behaviors following service failure", *Journal of Services Marketing*, Vol. 33 No. 3, pp. 285-302, doi: 10.1108/JSM-04-2018-0124.

Hawthorne, B.S., Vella-Brodrick, D.A. and Hattie, J. (2019), "Well-being as a cognitive load reducing agent: a review of the literature", *Frontiers in Education*, Vol. 4, doi: 10.3389/feduc.2019.00121.

Hernandez-Ortega, B. and Ferreira, I. (2021), "How smart experiences build service loyalty: the importance of consumer love for smart voice assistants", *Psychology and Marketing*, Vol. 38 No. 7, pp. 1122-1139, doi: 10.1002/mar.21497.

Hess, R.L., Jr, Ganesan, S. and Klein, N.M. (2003), "Service failure and recovery: the impact of relationship factors on customer satisfaction", *Journal of the Academy of Marketing Science*, Vol. 31 No. 2, pp. 127-145, doi: 10.1177/0092070302250898.

Holland, J., Kingston, L., McCarthy, C., Armstrong, E., O'Dwyer, P., Merz, F. and McConnell, M. (2021), "Service robots in the healthcare sector", *Robotics*, Vol. 10 No. 1, p. 47, available at: www.mdpi.com/2218-6581/10/1/47

Hoy, M.B. (2018), "Alexa, Siri, Cortana, and more: an introduction to voice assistants", *Medical Reference Services Quarterly*, Vol. 37 No. 1, pp. 81-88, doi: 10.1080/02763869.2018.1404391.

Huang, B. and Sénécal, S. (2023), "How should voice assistants be heard? The mitigating effect of verbal and vocal warmth in voice assistant service failure", *The Service Industries Journal*, Vol. 43 Nos 11-12, pp. 806-826, doi: 10.1080/02642069.2023.2208522.

Huang, Y.-S. and Dootson, P. (2022), "Chatbots and service failure: when does it lead to customer aggression", *Journal of Retailing and Consumer Services*, Vol. 68, p. 103044, doi: 10.1016/j.jretconser.2022.103044.

Jenneboer, L., Herrando, C. and Constantinides, E. (2022), "The impact of chatbots on customer loyalty: a systematic literature review", *Journal of Theoretical and Applied Electronic Commerce Research*, Vol. 17 No. 1, pp. 212-229.

Jiang, K., Qin, M. and Li, S. (2022), "Chatbots in retail: how do they affect the continued use and purchase intentions of Chinese consumers? ", *Journal of Consumer Behaviour*, Vol. 21 No. 4, pp. 756-772, doi: 10.1002/cb.2034.

Kapferer, J.N., and Bastien, V. (2012), *The Luxury Strategy: Break the Rules of Marketing to Build Luxury Brands*, Kogan Page Publishers.

Kaur, R., Sandhu, R.S., Gera, A., Kaur, T., and Gera, P. (2020), "Intelligent voice bots for digital banking", *Smart Systems and IoT: Innovations in Computing*, Springer; Singapore.

Khamitov, M., Grégoire, Y. and Suri, A. (2020), "A systematic review of brand transgression, service failure recovery and product-harm crisis: integration and guiding insights", *Journal of the Academy of Marketing Science*, Vol. 48 No. 3, pp. 519-542, doi: 10.1007/s11747-019-00679-1.

Klaus, P. and Zaichkowsky, J. (2020), "AI voice bots: a services marketing research agenda", *Journal of Services Marketing*, Vol. 34 No. 3, pp. 389-398, doi: 10.1108/JSM-01-2019-0043.

Kobel, S. and Groeppel-Klein, A. (2021), "No laughing matter, or a secret weapon? Exploring the effect of humor in service failure situations", *Journal of Business Research*, Vol. 132, pp. 260-269, doi: 10.1016/j.jbusres.2021.04.034.

Koole, S.L., Webb, T.L. and Sheeran, P.L. (2015), "Implicit emotion regulation: feeling better without knowing why", *Current Opinion in Psychology*, Vol. 3, pp. 6-10, doi: 10.1016/j.copsyc.2014.12.027.

Kulper, D.A. (2016), "The time course of anger: an experimental investigation ", (Doctoral dissertation, Temple University). Temple University ScholarShare, available at: https://scholarshare.temple.edu/items/d72fd8bc-b74f-4e20-8290-8de4d96c5a4e

Lee, J.E. (2023), "Frustration and anger mitigation after service failure: the effect of other consumer's explanation and employee's explanation on frustration and anger after the service failure", *Psychology Research and Behavior Management*, Vol. 16, pp. 1763-1785, doi: 10.2147/PRBM.S406753.

Lemon, K.N. and Verhoef, P.C. (2016), "Understanding customer experience throughout the customer journey", *Journal of Marketing*, Vol. 80 No. 6, pp. 69-96, doi: 10.1509/jm.15.0420.

Leo, X. and Huh, Y.E. (2020), "Who gets the blame for service failures? Attribution of responsibility toward robot versus human service providers and service firms", *Computers in Human Behavior*, Vol. 113, p. 106520, doi: 10.1016/j.chb.2020.106520.

Li, B., Liu, L., Mao, W., Qu, Y. and Chen, Y. (2023), "Voice artificial intelligence service failure and customer complaint behavior: the mediation effect of customer emotion", *Electronic Commerce Research and Applications*, Vol. 59, p. 101261, doi: 10.1016/j.elerap.2023.101261.

Liu, D., Lv, Y. and Huang, W. (2023), "How do consumers react to chatbots' humorous emojis in service failures", *Technology in Society*, Vol. 73, p. 102244, doi: 10.1016/j.techsoc.2023.102244.

Luo, X., Tong, S., Fang, Z. and Qu, Z. (2019), "Frontiers: machines vs. humans: the impact of artificial intelligence chatbot disclosure on customer purchases", *Marketing Science*, Vol. 38 No. 6, pp. 937-947, doi: 10.1287/mksc.2019.1192.

Lv, X., Liu, Y., Luo, J., Liu, Y. and Li, C. (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*, Vol. 87, p. 103114, doi: 10.1016/j.annals.2020.103114.

McKinsey and Company (2024), "The state of AI in early 2024", available at: www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai

McLean, G. and Osei-Frimpong, K. (2019), "Hey Alexa … examine the variables influencing the use of artificial intelligent in-home voice assistants", *Computers in Human Behavior*, Vol. 99, pp. 28-37, doi: 10.1016/j.chb.2019.05.009.

Matthews, M., Webb, T.L., Shafir, R., Snow, M. and Sheppes, G. (2021), "Identifying the determinants of emotion regulation choice: a systematic review with meta-analysis", *Cognition and Emotion*, Vol. 35 No. 6, pp. 1056-1084, doi: 10.1080/02699931.2021.1945538.

Mishra, A., Shukla, A. and Sharma, S.K. (2022), "Psychological determinants of users' adoption and word-of-mouth recommendations of smart voice assistants", *International Journal of Information Management*, Vol. 67, p. 102413, doi: 10.1016/j.ijinfomgt.2021.102413.

Mohamad Suhaili, S., Salim, N. and Jambli, M.N. (2021), "Service chatbots: a systematic review", *Expert Systems with Applications*, Vol. 184, p. 115461, doi: 10.1016/j.eswa.2021.115461.

Mozafari, N., Weiger, W. and Hammerschmidt, M. (2020), "The chatbot disclosure dilemma: desirable and undesirable effects of disclosing the non-human identity of chatbots", *Icis 2020*.

Naragon-Gainey, K., McMahon, T. and Chacko, T. (2017), "The structure of common emotion regulation strategies: a meta-analytic examination", *Psychological Bulletin*, Vol. 143 No. 4, pp. 384-427, doi: 10.1037/bul0000093.

Nguyen, D.T. and McColl-Kennedy, J.R. (2003), "Diffusing customer anger in service recovery: a conceptual framework", *Australasian Marketing Journal*, Vol. 11 No. 2, pp. 46-55, doi: 10.1016/S1441-3582(03)70128-1.

Obeidat, Z.M.I., Xiao, S.H., Iyer, G.R. and Nicholson, M. (2017), "Consumer revenge using the internet and social media: an examination of the role of service failure types and cognitive appraisal processes", *Psychology and Marketing*, Vol. 34 No. 4, pp. 496-515, doi: 10.1002/mar.21002.

Pagiavlas, S. (2020), "Two essays examining consumer responses to product and service failures", (Doctoral dissertation, University of South Carolina), Scholar Commons, available at: https://scholarcommons.sc.edu/etd/5670/

Pérez, J.Q., Daradoumis, T. and Puig, J.M.M. (2020), "Rediscovering the use of chatbots in education: a systematic literature review", *Computer Applications in Engineering Education*, Vol. 28 No. 6, pp. 1549-1565, doi: 10.1002/cae.22326.

Podsakoff, P.M., MacKenzie, S.B. and Podsakoff, N.P. (2012), "Sources of method bias in social science research and recommendations on how to control it", *Annual Review of Psychology*, Vol. 63 No. 1, pp. 539-569.

Qiu, P., Yang, R. and Potegal, M. (2009), "Statistical modeling of the time course of Tantrum Anger", *The Annals of Applied Statistics*, Vol. 3 No. 3, pp. 1013-1034, doi: 10.1214/09-AOAS242.

Rapp, A., Curti, L. and Boldi, A. (2021), "The human side of human – chatbot interaction: a systematic literature review of ten years of research on text-based chatbots", *International Journal of Human – Computer Studies*, Vol. 151, p. 102630, doi: 10.1016/j.ijhcs.2021.102630.

Reeck, C. and Onuklu, N.N.Y. (2022), "Interpersonal emotion regulation: consequences for brands in customer service interactions", *Frontiers in Psychology*, Vol. 13, p. 872670, doi: 10.3389/fpsyg.2022.872670.

Rohit, K., Shankar, A., Katiyar, G., Mehrotra, A. and Alzeiby, E.A. (2024), "Consumer engagement in chatbots and voicebots. A multiple-experiment approach in online retailing context", *Journal of Retailing and Consumer Services*, Vol. 78, p. 103728, doi: 10.1016/j.jretconser.2024.103728.

Ruan, S., Wobbrock, J.O., Liou, K., Ng, A. and Landay, J.A. (2018), "Comparing speech and keyboard text entry for short messages in two languages on touchscreen phones", *Proceedings of ACM on*

Interactive, Mobile, Wearable and Ubiquitous Technologies, Vol. 1 No. 4, pp. 1-23, doi: 10.1145/3161187.

Rzepka, C., Berger, B. and Hess, T. (2022), "Voice assistant vs. chatbot – examining the fit between conversational agents' interaction modalities and information search tasks", *Information Systems Frontiers*, Vol. 24 No. 3, pp. 839-856, doi: 10.1007/s10796-021-10226-5.

Salesforce (2024), "The state of service report", 6th edition, available at: www.salesforce.com/au/resources/research-reports/state-of-service/

Scharth, M. (2024), "ChatGPT is now better than ever at faking human emotion and behaviour", The Conversation, available at: https://theconversation.com/chatgpt-is-now-better-than-ever-at-faking-human-emotion-and-behaviour-230254

Schindler, D., Maiberger, T., Koschate-Fischer, N. and Hoyer, W.D. (2024), "How speaking versus writing to conversational agents shapes consumers' choice and choice satisfaction", *Journal of the Academy of Marketing Science*, Vol. 52 No. 3, pp. 634-652, doi: 10.1007/s11747-023-00987-7.

Schöbel, S., Schmitt, A., Benner, D., Saqr, M., Janson, A. and Leimeister, J.M. (2024), "Charting the evolution and future of conversational agents: a research agenda along five waves and new frontiers", *Information Systems Frontiers*, Vol. 26 No. 2, pp. 729-754, doi: 10.1007/s10796-023-10375-9.

Sezgin, E., Huang, Y., Ramtekkar, U. and Lin, S. (2020), "Readiness for voice assistants to support healthcare delivery during a health crisis and pandemic", *Npj Digital Medicine*, Vol. 3 No. 1, pp. 122-122, doi: 10.1038/s41746-020-00332-0.

Shahzad, M.F., Xu, S., An, X. and Javed, I. (2024), "Assessing the impact of AI-chatbot service quality on user e-brand loyalty through chatbot user trust, experience and electronic word of mouth", *Journal of Retailing and Consumer Services*, Vol. 79, p. 103867, doi: 10.1016/j.jretconser.2024.103867.

Shams, G., Kim, K.K. and Kim, K. (2024), "Enhancing service recovery satisfaction with chatbots: the role of humor and informal language", *International Journal of Hospitality Management*, Vol. 120, p. 103782, doi: 10.1016/j.ijhm.2024.103782.

Sharma, S. and Singh, G. (2021), "Comparison of voice based virtual assistants fostering indian higher education – a technical perspective", *2021 International Conference on Technological Advancements and Innovations (ICTAI)*.

Shurcliff, A. (1968), "Judged humor, arousal, and the relief theory", *Journal of Personality and Social Psychology*, Vol. 8 No. 4, pp. 360-363, doi: 10.1037/h0025493.

Singh, M., Mittal, M., Dewan, P., Kaur, A., Kaur, G. and Gupta, A. (2024, 28 Feb.-1 March 2024), "From text to treatment: an overview of artificial intelligence chatbots in healthcare", *2024 11th International Conference on Computing for Sustainable Global Development (INDIACom)*.

Sivakumar, K., Li, M. and Dong, B. (2014), "Service quality: the impact of frequency, timing, proximity, and sequence of failures and delights", *Journal of Marketing*, Vol. 78 No. 1, pp. 41-58, doi: 10.1509/jm.12.0527.

Skowronski, J.J., Walker, W.R., Henderson, D.X., and Bond, G.D. (2014), "The fading affect bias: its history, its implications, and its future", In J. M. Olson and M. P. Zanna (Eds), *Advances in Experimental Social Psychology*, Elsevier Academic Press, Vol. 49, pp. 163-218.

Smith, A.K., Bolton, R.N. and Wagner, J. (1999), "A model of customer satisfaction with service encounters involving failure and recovery", *Journal of Marketing Research*, Vol. 36 No. 3, pp. 356-372, doi: 10.1177/002224379903600305.

So, K.K.F., Kim, H., Liu, S.Q., Fang, X. and Wirtz, J. (2024), "Service robots: the dynamic effects of anthropomorphism and functional perceptions on consumers' responses", *European Journal of Marketing*, Vol. 58 No. 1, pp. 1-32, doi: 10.1108/EJM-03-2022-0176.

Song, M., Zhang, H., Xing, X. and Duan, Y. (2023), "Appreciation vs. apology: research on the influence mechanism of chatbot service recovery based on politeness theory", *Journal of Retailing and Consumer Services*, Vol. 73, p. 103323, doi: 10.1016/j.jretconser.2023.103323.

Sun, Y., Li, S. and Yu, L. (2022), "The dark sides of AI personal assistant: effects of service failure on user continuance intention", *Electronic Markets*, Vol. 32 No. 1, pp. 17-39, doi: 10.1007/s12525-021-00483-2.

Sweller, J., Ayres, P., and Kalyuga, S. (2011), *Cognitive Load Theory*, Springer, New York, NY.

Tamir, M. and Ford, B.Q. (2009), "Choosing to be afraid: preferences for fear as a function of goal pursuit", *Emotion*, Vol. 9 No. 4, pp. 488-497, doi: 10.1037/a0015882.

Tärning, B. and Silvervarg, A. (2019), "I didn't understand, I'm really not very smart"–how design of a digital tutee's self-efficacy affects conversation and student behavior in a digital math game", *Education Sciences*, Vol. 9 No. 3, doi: 10.3390/educsci9030197.

Taylor, J.B. (2021), *Whole Brain Living: The Anatomy of Choice and the Four Characters That Drive Our Life*, Hay House, Inc.

Tindle, R. and Longstaff, M.G. (2015), "Writing, reading, and listening differentially overload working memory performance across the serial position curve", *Advances in Cognitive Psychology*, Vol. 11 No. 4, pp. 147-155, doi: 10.5709/acp-0179-6.

Tsarenko, Y. and Strizhakova, Y. (2013), "Coping with service failures", *European Journal of Marketing*, Vol. 47 Nos 1-2, pp. 71-92, doi: 10.1108/03090561311285466.

Valentini, S., Orsingher, C. and Polyakova, A. (2020), "Customers' emotions in service failure and recovery: a meta-analysis", *Marketing Letters*, Vol. 31 Nos 2-3, pp. 199-216, doi: 10.1007/s11002-020-09517-9.

Valenzuela, A., and Hadi, R. (2018), "Implications of product anthropomorphism through design", in Solomon, Tina, M.L. (Ed.), *The Routledge Companion to Consumer Behavior*, Taylor and Francis.

Valtolina, S., Barricelli, B.R. and Di Gaetano, S. (2020), "Communicability of traditional interfaces VS chatbots in healthcare and smart home domains", *Behaviour and Information Technology*, Vol. 39 No. 1, pp. 108-132, doi: 10.1080/0144929X.2019.1637025.

Wei, T. and Shi, Y. (2024), "A service failure assessment model for smart product consumption experience based on customer perception", *Scientific Reports*, Vol. 14 No. 1, p. 23097, doi: 10.1038/s41598-024-75283-7.

Weun, S., Beatty, S.E. and Jones, M.A. (2004), "The impact of service failure severity on service recovery evaluations and post-recovery relationships", *The Journal of Services Marketing*, Vol. 18 No. 2, pp. 133-146, doi: 10.1108/08876040410528737.

Xing, X., Song, M., Duan, Y. and Mou, J. (2022), "Effects of different service failure types and recovery strategies on the consumer response mechanism of chatbots", *Technology in Society*, Vol. 70, p. 102049, doi: 10.1016/j.techsoc.2022.102049.

Yun, J. and Park, J. (2022), "The effects of chatbot service recovery with emotion words on customer satisfaction, repurchase intention, and positive Word-Of-Mouth [original research]", *Frontiers in Psychology*, Vol. 13, doi: 10.3389/fpsyg.2022.922503.

Zierau, N., Hildebrand, C., Bergner, A., Busquet, F., Schmitt, A. and Marco Leimeister, J. (2023), "Voice bots on the frontline: voice-based interfaces enhance flow-like consumer experiences and boost service outcomes", *Journal of the Academy of Marketing Science*, Vol. 51 No. 4, pp. 823-842, doi: 10.1007/s11747-022-00868-5.

Further reading

Bendapudi, N. and Leone, R.P. (2003), "Psychological implications of customer participation in co-production", *Journal of Marketing*, Vol. 67 No. 1, pp. 14-28, doi: 10.1509/jmkg.67.1.14.18592.

Choi, S. and Mattila, A.S. (2008), "Perceived controllability and service expectations: influences on customer reactions following service failure", *Journal of Business Research*, Vol. 61 No. 1, pp. 24-30, doi: 10.1016/j.jbusres.2006.05.006.

Folkman, S., Lazarus, R.S., Dunkel-Schetter, C., DeLongis, A. and Gruen, R.J. (1986), "Dynamics of a stressful encounter: cognitive appraisal, coping, and encounter outcomes", *Journal of Personality and Social Psychology*, Vol. 50 No. 5, pp. 992-1003, doi: 10.1037/0022-3514.50.5.992.

Appendix

Study 1

Field experiment setup

(a) The local dealer website and online chatbot

(b) Voicebot greeting transcript

Chinese (origin): 您好, 我是智能语音助手小智, 欢迎致电 XX电脑售后服务中心。我们致力于为您提供高效便捷的服务支持。在接下来的通话中, 我会根据您的语音描述, 识别您遇到的问题, 并为您提供相应的解决方案。请尽量用简洁明确的语言描述您的问题, 比如“电脑开不了机”、“屏幕黑屏”、“键盘失灵”等。感谢您的耐心, 现在请您开始描述您遇到的问题。

English (Translation): Hello, this is Smart Voice Assistant Xiao Zhi. Thank you for calling the XX Computer After-Sales Service Center. We're here to provide you with fast and convenient support.

During this call, I will listen to your voice input, identify the issue you're experiencing and offer the most appropriate solution. Please try to describe your problem clearly and briefly, such as “the computer won't turn on,” “the screen is black” or “the keyboard is not working.” Thank you for your patience. Now, please go ahead and describe the issue you're experiencing.

Study 2

Broken camera scenario

Imagine you have bought a camera specifically for an upcoming trip. Upon receiving the camera, you excitedly open it and snap a few pictures, only to realize they are disappointingly blurry. You suspect the lens might be cracked or misaligned. You search online for guidance on how to exchange the camera, only to find that the estimated delivery date for a replacement falls after your departure date – defeating the whole purpose of your purchase. Frustration sets in, and you contemplate reaching out to the camera company. Unfortunately, it is outside of their business hours, so you opt to explore their online customer service center instead.

A chatbot window pops up, and the conversation between you and the chatbot is as follows. You discovered the phone number for the company's Automated Customer Service Center. You made the call, and the conversation between you and a voicebot is below.

Conversation with chatbot/voicebot

Jamie: Hey there, I'm Jamie, your Customer Service Assistant. Thanks for getting in touch with our Automated Customer Service Center. What can I help you with today?

Your response: Hi Jamie, I'm honestly so frustrated right now. I bought a camera from your company last week, and the pictures I took are all unbelievably blurry. I've got a trip coming up, and I really need this sorted out.

Jamie: I'm really sorry to hear that you're going through this. Did you know about our Advanced Video Stabilization software? It's designed to fix that exact issue caused by camera movement. You can download it from our website.

Your response: No, I had no clue about that software. This whole situation is just messing with my plans for the trip, and it's driving me crazy.

Jamie: I totally understand your frustration, and I'm sorry for the inconvenience. Downloading the software might make a significant difference. Hope it helps salvage your photos before your trip.

Your response: Thanks for the info. I'll give it a shot, though I'm seriously hoping it works. This trip means a lot to me. Thanks for your help. I appreciate it.

Jamie: No problem at all. Wishing you the best with your photos and your trip. If anything else comes up, don't hesitate to reach out. Take care!

Reliability and validity tests

Table A1. Factor loadings, composite reliability (CR) and average variance extracted (AVE)

			Study 2	
Construct	Item	Loading	CR	AVE
Perceived relief (α = 0.921)	PR_1	0.835	0.922	0.703
	PR_2	0.822	–	–
	PR_3	0.869	–	–
	PR_4	0.788	–	–
	PR_5	0.875	–	–
Customer review (α = 0.919)	CR_1	0.905	0.930	0.769
	CR_2	0.873	–	–
	CR_3	0.880	–	–
	CR_4	0.848	–	–

Table A2 Correlation matrix

Constructs	PR	CR
PR	*0.838*	
CR	0.787	*0.877*

Note(s): PR = perceived relief; CR = customer review. Diagonal values in bold indicate the square root of average variances extracted (AVE). Scores in the lower diagonal indicate interconstruct correlations

Study 3

High-severity scenario (refrigerator completely stopped working)

Imagine you recently purchased a brand-new refrigerator, and after just a few days of use, it has completely stopped working. Your food is at risk of spoiling, and you are facing an urgent situation. You decide to reach out to the refrigerator company for assistance.

Low-severity scenario (intermittent cooling problems)

Imagine you have recently bought a new refrigerator, but you have been experiencing intermittent problems with its cooling system. Sometimes, the temperature seems fine, but on other occasions, you notice that your food is not as cold as it should be. It is not an immediate emergency, but it is definitely a concern, and you want to address it with the refrigerator company.

Bot modality

Unfortunately, it is outside of their business hours, so you opt to explore their online customer service center instead.

A chatbot window pops up, and the conversation between you and the chatbot is as follows.

You discovered the phone number for the company's Automated Customer Service Center. You made the call, and the conversation between you and a voicebot is below.

High-severity scenario (refrigerator completely stopped working)

Jenny: Hey there, I'm Jenny, your Customer Service Assistant. Thanks for getting in touch with our Automated Customer Service Center. What can I help you with today?

Your response: Hi, I recently purchased one of your refrigerators, and it's completely stopped working after just a few days. I'm really worried because all my food is inside, and it's at risk of spoiling. Can you please help me with this urgent issue?

Jenny: I'm sorry to hear about the issue with your refrigerator. We understand the urgency. Let's try to troubleshoot this together. Have you checked if it's plugged in properly, and is the power outlet working?

Your response: Yes, I've checked that, and everything seems fine on that end. It's just not cooling at all.

Jenny: Thank you for confirming that. It might be a more serious issue. Can you provide your purchase reference number, so we can verify the warranty status and will schedule a technician to come and inspect the refrigerator as soon as possible?

Your response: Sure, my reference number is 1368. Please send someone as soon as you can; I'm really concerned about my food.

Jenny: Thank you for the information. We'll prioritize your case and get a technician out to you as soon as possible. We apologize for any inconvenience this has caused and appreciate your patience.

Low-severity scenario (intermittent cooling problems)

Jenny: Hey there, I'm Jenny, your Customer Service Assistant. Thanks for getting in touch with our Automated Customer Service Center. What can I help you with today?

Your response: Hello, I recently bought one of your refrigerators, and I've been having some issues with the cooling system. It's not consistently keeping my food as cold as it should be. It's not an emergency, but I'd like to address this issue.

Jenny: Thank you for reaching out. I'm sorry to hear about the cooling problem you're experiencing. Let's try to resolve this. Have you checked if the temperature settings on the refrigerator are correctly adjusted?

Your response: Yes, I've double-checked the settings, and they seem to be okay. It just seems to work fine sometimes and not so well at other times.

Jenny: I see. It could be a minor issue, but we want to make sure it's resolved for you. Can you provide your purchase reference number, so we can verify the warranty status and schedule a technician to come and inspect the refrigerator?

Your response: Sure, my reference number is 1368. Please let me know when someone can come take a look.

Jenny: Thank you for providing the information. We'll schedule a technician to investigate the issue and make the necessary repairs. We appreciate your patience while we work to resolve this for you.

Reliability and validity tests

Table A3. Factor loadings, composite reliability (CR) and average variance extracted (AVE)

			Study 3	
Construct	Item	Loading	CR	AVE
Perceived relief (α = 0.943)	PR_1	0.878	0.943	0.768
	PR_2	0.856	–	–
	PR_3	0.879	–	–
	PR_4	0.883	–	–
	PR_5	0.885	–	–
Customer review (α = 0.941)	CR_1	0.909	0.943	0.806
	CR_2	0.916	–	–
	CR_3	0.910	–	–
	CR_4	0.854	–	–

Table A4. Correlation matrix

Constructs	PR	CR
PR	*0.877*	
CR	0.694	*0.898*

Note(s): PR = perceived relief; CR = customer review. Diagonal values in bold indicate the square root of average variances extracted (AVE). Scores in the lower diagonal indicate interconstruct correlations

Corresponding authors
Mai Nguyen can be contacted at: mai.journalarticles@gmail.com and Yunen Zhang can be contacted at: ianzhang508@gmail.com