Trust me, I'm a bot – repercussions of chatbot disclosure in different service frontline settings

Nika Mozafari
Faculty of Business and Economics, University of Goettingen, Goettingen, Germany
Welf H. Weiger
College of Business, Alfaisal University, Riyadh, Saudi Arabia and Faculty of Business and Economics, University of Goettingen, Goettingen, Germany, and
Maik Hammerschmidt
Faculty of Business and Economics, University of Goettingen, Goettingen, Germany

Received 31 October 2020
Revised 6 April 2021
Accepted 22 May 2021

Abstract

Purpose – Chatbots are increasingly prevalent in the service frontline. Due to advancements in artificial intelligence, chatbots are often indistinguishable from humans. Regarding the question whether firms should disclose their chatbots' nonhuman identity or not, previous studies find negative consumer reactions to chatbot disclosure. By considering the role of trust and service-related context factors, this study explores how negative effects of chatbot disclosure for customer retention can be prevented.

Design/methodology/approach – This paper presents two experimental studies that examine the effect of disclosing the nonhuman identity of chatbots on customer retention. While the first study examines the effect of chatbot disclosure for different levels of service criticality, the second study considers different service outcomes. The authors employ analysis of covariance and mediation analysis to test their hypotheses.

Findings – Chatbot disclosure has a negative indirect effect on customer retention through mitigated trust for services with high criticality. In cases where a chatbot fails to handle the customer's service issue, disclosing the chatbot identity not only lacks negative impact but even elicits a positive effect on retention.

Originality/value – The authors provide evidence that customers will react differently to chatbot disclosure depending on the service frontline setting. They show that chatbot disclosure does not only have undesirable consequences as previous studies suspect but can lead to positive reactions as well. By doing so, the authors draw a more balanced picture on the consequences of chatbot disclosure.

Keywords Chatbots, Chatbot identity disclosure, Service criticality, Chatbot failure, Trust, Customer retention

Paper type Research paper

Introduction

Recent advancements in artificial intelligence encourage more and more firms to use chatbots for service delivery and incorporate them into the frontline (van Doorn *et al.*, 2017). Chatbots are text-based virtual robots that emulate human-to-human conversation through natural language processing (Schuetzler *et al.*, 2018; Wirtz *et al.*, 2018). They offer the chance to provide efficient customer service around the clock, therefore serving as a crucial strategic asset for firms (Thomaz *et al.*, 2020). For instance, a recent industry report forecasts that by

This is a submission for the JOSM special section on "Living and Working with (Ro)bots – The Impact of (Ro)bots on the Service Frontline".

The authors thank Thorsten Hennig-Thurau, Michael Paul, Tillmann Wagner, Gianfranco Walsh, and all participants of the 2019 and 2020 Research Bootcamp on Marketing for their valuable feedback on earlier drafts of the manuscript.

Journal of Service Management

1757-5818
DOI 10.1108/JOSM-10-2020-0380

2025, 95% of all consumer interactions with a firm will be powered – that is, augmented or replaced – by chatbots (Servion, 2020).

In contrast to traditional self-service technologies with a merely functional character, chatbots are equipped with additional social–emotional and relational elements (Wirtz *et al.*, 2018). Not only does the natural language interface remind of human conversation (Tuzovic and Paluch, 2018), chatbots also take on roles that were so far fulfilled by human frontline employees and provide personalized responses based on sophisticated speech recognition tools that create an anthropomorphic conversation (Nass and Moon, 2000; Wilson *et al.*, 2017; Wuenderlich and Paluch, 2017). However, this rapid advancement of chatbot technology comes with a dark side: As chatbots become increasingly anthropomorphic, consumers find it increasingly challenging to correctly distinguish between human or artificial conversational partners (Candello *et al.*, 2017).

As this challenge gains traction, firms are confronted with the question whether or not to disclose information on the nonhuman identity of chatbots. Previous studies that attempt to address this question consistently find negative consumer reactions (i.e. perceiving it as less empathetic or knowledgeable; Luo *et al.*, 2019) to disclosed vs undisclosed chatbots and therefore stress the detrimental effects of chatbot disclosure. As these negative reactions in service interactions jeopardize customer retention by alienating customers (Puntoni *et al.*, 2021), this study aims to answer the following research question:

RQ1. How does disclosing chatbot identity influence customer retention?

In order to find ways for mitigating or at best reversing potential negative retention effects of chatbot disclosure, firms need to understand the mechanism that fuels the relationship between chatbot disclosure and retention. Approaches that try to explain negative reactions highlight consumers' aversion toward algorithms that is rooted in the lack of trust in their performance in service delivery (Dietvorst *et al.*, 2015). The importance of trust as an explaining mechanism is stressed through human's tendency to react to computers as they would to humans in social contexts (Nass and Moon, 2000; Nass *et al.*, 1994) particularly as interfaces become more and more anthropomorphic (Holtgraves *et al.*, 2007). Not only is trust in an exchange partner a key mediator between service attributes and customer retention (Morgan and Hunt, 1994), it is also a focal construct in a variety of chatbot studies (e.g. de Visser *et al.*, 2016). However, the role of trust has not yet been empirically investigated in the context of chatbot disclosure despite current calls to examine what drives consumers' (mis)trust in chatbots (De Keyser *et al.*, 2019; Wirtz *et al.*, 2018). Therefore, to provide insights into the underlying mechanism responsible for negative consumer reactions, the following question is posed:

RQ2. Does trust in the conversational partner mediate the effect of chatbot disclosure on customer retention?

Finally, this research seeks to find contextual factors that influence customers' trust responses to chatbots in order to reveal settings in which negative retention effects might be mitigated, eliminated or even reversed. In doing so, this research answers recent calls to include service characteristics as potential moderators of the customer–chatbot interaction (Wirtz *et al.*, 2018). This paper draws on service characteristics that can shape the effectiveness of chatbot communication for retention (Webster and Sundaram, 1998, 2009): service criticality and service outcome. First, as chatbot technology becomes more sophisticated, it shifts from merely being used to answering simple FAQ-style questions (Luger and Sellen, 2016) to handling more complex and critical service requests (Luo *et al.*, 2019). So far, no study examines different levels of criticality of services delivered by chatbots. As trust in disclosed versus undisclosed chatbots should vary with different levels of service criticality, this research considers service criticality as a moderator in the

relationship between chatbot disclosure and trust. Second, service outcome – whether a chatbot fails to handle service requests or not – remains the most prominent influence on trust (Nordheim *et al.*, 2019). Prior studies, however, did not examine variations in service outcome, in that whether chatbots were able to handle service requests as good as undisclosed chatbots or human counterparts would. Related studies have shown that trust will be affected differently for failures induced by more or less anthropomorphic conversational partners (de Visser *et al.*, 2016), and outcome attributions differ depending on whether a service was delivered by a human or a robot (Belanche *et al.*, 2020). Recent studies call to examine customer reactions to chatbot failures, specifically in scenarios where they may not be aware of the nonhuman identity of the conversational partner (Sheehan *et al.*, 2020). This research therefore considers service outcome as a further moderator. Hence, the third research question is as follows:

RQ3. Do service-related context factors (i.e. service criticality and service outcome) moderate the effect of chatbot disclosure on trust?

In tackling these questions, this study contributes to research on the impact of chatbots on the service frontline. First, this research is the first to show that, for certain service contexts, positive outcomes of chatbot disclosure prevail. By doing so, it advances the literature that so far has taken a pessimistic view on disclosing the nonhuman nature of conversational agents and advices against disclosure. By taking a more nuanced view on the typical contextual settings in which conversational agents operate, this paper provides a more balanced picture on the consequences of disclosing the machine nature of the agent. Specifically, results empirically show that response to chatbot disclosure is detrimental for highly critical service issues while a disclosure is beneficial once a conversational agent cannot resolve the service issue. Second, the role of trust in the conversational partner is highlighted as a mediator between chatbot disclosure and customer retention. Thus, this research shows the relevance of this consumer response in chatbot-mediated interactions as it determines whether behavioral outcomes of chatbot disclosure are desirable or undesirable for firms. Together, these insights guide firm's design of chatbot systems in terms of whether and under which circumstances to disclose chatbot identity.

The rest of this research article is structured as follows: After presenting the conceptual framework, related research on chatbot disclosure and the roles of trust and service-related factors in the context of technology-mediated interactions are discussed. Mechanisms from attribution theory are introduced as the underpinning for the hypotheses regarding the relationships between the focal variables. Next, this paper presents two experiments simulating interactions with a chatbot to identify whether chatbot disclosure will yield positive or negative effects on customer retention in different service frontline settings. Finally, findings are summarized and implications are outlined.

Conceptual background

Research framework

Figure 1 illustrates the research framework. To evaluate whether and under what circumstances chatbot disclosure produces favorable outcomes for firms, this research considers the effect of chatbot disclosure for different levels of service criticality (i.e. high vs low service criticality) and different service outcomes (i.e. chatbot failure vs no chatbot failure) on customer retention through trust. Customer retention is chosen as a metric to capture the outcome of chatbot disclosure as it is key to company profitability (McCollough *et al.*, 2000). Further, this research focuses on retention instead of purchase behavior, as most chatbots today are deployed in postpurchase customer service settings (Shevat, 2017).

Literature review on chatbot disclosure

As chatbot technology becomes increasingly sophisticated and chatbots are increasingly able to pose as humans, the more relevant it becomes for firms to understand the repercussions of disclosing or not disclosing chatbot identity (Skjuve *et al.*, 2019). The discussion was already sparked in 2018 by Google Duplex. The intelligent phone assistant employed a variety of anthropomorphic cues that were characteristic to human conversations, for example, the incorporation of speech disfluencies, creating an uncannily realistic experience. In 2020, Google presented their chatbot Meena, whose conversational quality level was rated nearly as high as that of human conversations (Adiwardana *et al.*, 2020). Scholars argue that in these environments, transparency on the identity of the conversational agent is essential for consumers to evaluate an interaction and form trust (Donath, 1999; Wang and Benbasat, 2008).

However, existing empirical research on the effect of chatbot disclosure has thus far found largely negative reactions to disclosed (vs undisclosed) chatbots, despite identical performance, suggesting that transparency about identity comes at a cost. In a behavioral experiment on the prisoner's dilemma, Ishowo-Oloko *et al.* (2019) demonstrate that bots are better at generating cooperative behavior; however, this efficiency is negated once the bot identity is disclosed. The study's participants also do not recover from their negatively biased assessment of the bot. If an interaction partner is perceived as a bot, they receive more negative user ratings (Murgia *et al.*, 2016) and are evaluated as less persuasive (Shi *et al.*, 2020), less socially present and less human (Hendriks *et al.*, 2020). Luo *et al.* (2019) further demonstrate that chatbot disclosure drastically decreases interaction length and subsequent purchase behavior.

Overall, these studies dominantly take on the algorithm aversion perspective (Dietvorst *et al.*, 2015; Jussupow *et al.*, 2020), in that negative biases toward chatbot are caused by lack of trust in their performance. Only one existing study argues adversatively, saying that undisclosed bots will negatively affect user experience due to feelings of uncertainty, however, find no significant evidence that a chatbot that is believed to be human is perceived as less pleasant than a chatbot whose identity is disclosed (Skjuve *et al.*, 2019). The consistent absence of positive effects of chatbot disclosure in any of the mentioned study is startling: Not only has research observed negative biases toward disclosed bots, although performance levels in service delivery were held constant across disclosed and undisclosed bots, but also did some studies provide evidence on superior performance of bots over humans (e.g. Ishowo-Oloko *et al.*, 2019), speaking for positive responses toward disclosed bots.

Trust in human–chatbot interactions

Some of the existing studies on chatbot disclosure attempt to provide explanations for negative biases. A common argument is the lack of trust in algorithms, that is mentioned in all, but not tested in any of the studies (Murgia *et al.*, 2016; Ishowo-Oloko *et al.*, 2019; Luo *et al.*, 2019; Skjuve *et al.*, 2019; Hendriks *et al.*, 2020; Shi *et al.*, 2020). This section highlights

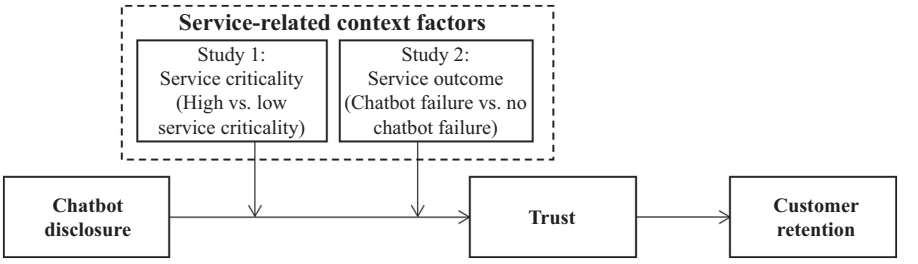

Figure 1.
Research framework

arguments for considering the role of trust for human–chatbot interactions as an explanatory mechanism for behavioral responses to chatbot disclosure.

Trust is defined as the willingness to rely on an exchange partner, more specifically the willingness to rely on the trustee to be able to fulfill their obligations (i.e. competence), to tell the truth (i.e. integrity) and to act in the trustor's interest (i.e. benevolence) (Komiak and Benbasat, 2004; Moorman *et al.*, 1993). Trust is central for environments that produce high levels of uncertainty, for example, online settings (Riedl *et al.*, 2011). According to commitment–trust theory, trust in an exchange partner is a key mediating variable between services and relational outcomes (Hart and Johnson, 1999; Morgan and Hunt, 1994). A lack of trust is described as a "stumbling block" (Sherman, 1992, p. 78) for successful relationships, as especially for services, trust is the basis for loyalty (Berry and Parasuraman, 1991). In exchanges between buyers and sellers, trust is a central element to constructive dialogue (Schurr and Ozanne, 1985). If trust is established in a relationship, the trustor will commit themselves to that relationship (Hrebiniak, 1974).

Importantly, the key role of trust is also enforced in agent-mediated interactions (Komiak and Benbasat, 2004). Consumers generalize social concepts such as trust to computers, even if they know they are not interacting with a living being (Nass and Moon, 2000). Neurological research confirms that trust-building processes within human–computer interactions can in fact be compared to that of human–human interactions (Riedl *et al.*, 2011).

In a chatbot context, a variety of studies have focused on the examination of trust as a reaction to chatbot design (e.g. de Visser *et al.*, 2016; Nunamaker *et al.*, 2011; Sameh *et al.*, 2010). However, in the context of chatbot disclosure, the difference in trust between disclosed and undisclosed chatbots remains largely unexplored, although it is likely to determine whether behavioral outcomes (e.g. retention) of chatbot disclosure are favorable or unfavorable for the chatbot-employing firm. Further, this research aims to create a comprehensive framework by including not only trust but also subsequent customer retention.

Service-related context factors of human–chatbot interactions

As all the studies discussed in the literature review have only examined main effects of disclosure implying universal consequences of disclosure across service situations, this research suggests that these insights should be enriched by testing whether the effects of chatbot disclosure on business-relevant outcomes vary across different service contexts as these contexts arguably shape consumers' trust responses to revealed chatbots.

The role of service criticality. Existing studies fail to address that consumers are likely to react differently to chatbot disclosure in interactions with different levels of service criticality, which refers to a customer's perceived importance of whether a service is successfully delivered (Ostrom and Iacobucci, 1995; Webster and Sundaram, 1998). Retrospectively, many consumers mainly use chatbots to handle simple rule-based tasks, such as finding an answer to an FAQ-style question or other menial requests (Huang and Rust, 2018; Zamora, 2017). However, as technology evolves, so does the scope of services chatbots are asked to deliver. With more complex, conversational-like tasks, consumers are likely to have increased situational involvement with the service, which should alter consumer reactions to such services (Webster and Sundaram, 1998).

To explain how and why chatbot disclosure affects consumer trust and thus retention differently depending on service criticality, this research draws upon attribution theory. People are inherently driven to assign causes to other's behavior and events in order to better understand their environment. Attribution theory investigates this formation of causal judgment, which is based on situational factors, such as external circumstances, or dispositional factors, such as beliefs about the ability or motivation of others (Heider, 1958). These causal attributions subsequently will affect consumer responses such as trust and retention (van Vaerenbergh *et al.*, 2014).

A core tenet of attribution theory suggests that the process of inferring a cause for behavior or events is prone to the attribution bias (Forsyth, 1987). The attribution bias describes the tendency of humans to overly rely on dispositions relative to situational influences, that is, hastily forming judgments based on personal beliefs, overlooking the actual situational behavior of an exchange partner (Ross, 1977). That means, when reacting to an event, people tend to ascribe the outcome of a situation to the perceived characteristics of involved parties instead of the actual situational environment. This cognitive bias occurs as a result of a spontaneous, premature attribution.

Research shows that humans are skeptical toward algorithms (Dawes, 1979), tend to have less confidence in their performance (Dietvorst *et al.*, 2015) and perceive chatbots as less knowledgeable and empathetic, especially with regard to high criticality services (Luo *et al.*, 2019). Following this line of reasoning, consumers have the belief that chatbots are not capable of handling critical service issues. The notion that humans prefer interacting with a human(like) counterpart for high criticality services is supported by a recent meta-analysis, which finds that anthropomorphic robot design has a stronger positive effect on usage intentions for critical than for noncritical services (Blut *et al.*, 2021). Based on the attribution bias, for high criticality services, consumers rely on their negative disposition toward chatbots when learning about the chatbot identity of the conversational partner and hence form reduced trust. Therefore:

H1. If service criticality is high, disclosing (vs not disclosing) chatbot identity reduces trust in the conversational partner.

As stated earlier, service research has frequently found a positive impact of trust on customer retention (e.g. Morgan and Hunt, 1994; Silitonga *et al.*, 2020). As the direct relationship between trust and retention is well-established, this research does not formulate hypotheses for this relationship. Under the notion that trust positively affects customer retention, if chatbot disclosure in a setting with high service criticality reduces trust, retention will indirectly be affected negatively. In this case, trust takes in a mediating role between the service interaction and the behavioral outcome (Morgan and Hunt, 1994). Hence:

H2. If service criticality is high, chatbot disclosure has an indirect negative effect on retention, which is mediated by trust in the conversational partner.

The role of service outcome. In all the studies discussed in the literature review, bot performance is at a high level, so that reactions to chatbot disclosure in failure settings remain yet to be investigated. However, research shows that consumers react differently to robot errors than to human errors (Belanche *et al.*, 2020). Particularly, as chatbot design influences error tolerance and trust resilience (de Visser *et al.*, 2016), consumers should react differently to errors from disclosed versus undisclosed chatbots. Some studies have approached examining the impact of chatbot failure on consumer reactions (e.g. Sheehan *et al.*, 2020), showing that chatbot failure significantly decreases adoption intent. However, regarding the impact of chatbot disclosure on trust and retention, the consideration of different service outcomes remains unexamined.

As stated earlier, attributions made by humans are oftentimes biased as they occur spontaneously. However, attribution becomes less spontaneous and more elaborated if the valence of an outcome is negative (Kanazawa, 1992). That is, if a negative outcome (e.g. failure) occurs, individuals feel the need to comprehend, control and predict their environment in order to effectively cope with the situation (Weiner, 2000). In this case, individuals invest higher effort to more deeply understand the cause for a negative outcome (van Vaerenbergh *et al.*, 2014; Weiner, 1985).

In search for a cause of the negative outcome, chatbot disclosure represents a concrete cue that stimulates attributional activity and allows a better understanding of the reasons of the

failure (Weiner, 1985). This in turn serves as a coping mechanism to help deal with frustration and anger caused by the failure (Gelbrich, 2010). If chatbot identity is not disclosed, information on the cause of the negative service outcome remains abstract and the customer is not able to identify a specific entity the failure can be attributed to. Through chatbot disclosure, customers should therefore be able to better cope with the situation. Empirical evidence from research on decision support systems further demonstrates that providing explanation for outcomes nurtures trust (Wang and Benbasat, 2008).

Taken together, being able to locate the cause for failure should enhance trust compared to not locating it.

H3. If a chatbot failure occurs, disclosing (vs not disclosing) chatbot identity enhances trust in the conversational partner.

Under the notion that trust positively affects retention, if chatbot disclosure (vs no disclosure) in a chatbot failure setting enhances trust, retention will be affected positively. Again, trust takes in a mediating role between the service interaction and the behavioral outcome (Morgan and Hunt, 1994). Hence:

H4. If a chatbot failure occurs, chatbot disclosure has an indirect positive effect on retention, which is mediated by trust in the conversational partner.

Empirical examination

Study 1: Chatbot disclosure and service criticality

Study design. The goal of study 1 was to examine how disclosing the nonhuman chatbot identity impacts consumer trust and in turn retention for different levels of service criticality (but holding service outcome constant in terms of considering successful service delivery). To examine this, the study applies a 2 (chatbot disclosure vs no disclosure) $\times$ 2 (high vs low service criticality) between-subject experiment. The study was conducted as a scenario-based experiment, to be able to control for confounding influences and ensure high internal validity. This enabled the creation of a human–chatbot interaction, from which participants could not infer the identity of the conversational partner without explicit disclosure.

Participants of the online experiment were recruited through a European online panel provider (i.e. Clickworker) with monetary compensation and were randomly assigned to one of the two service criticality scenarios. In both scenarios, participants were instructed to imagine that they were moving into a new apartment and had to contact their current energy provider to inform about the address change in order to register their current electricity contract under the new address. In addition to this information, the participants in the high criticality scenario were informed that if they did not succeed in reregistering their contract, they would automatically receive their electricity from the public utility provider, which would result in a significantly higher monthly rate. Afterward, participants of both scenarios were exposed to identical service interactions up to the disclosure manipulation. The manipulation for service criticality is common in service research (Ostrom and Iacobucci, 1995; Webster and Sundaram, 1998), as the criticality refers to the subjective importance of the service being delivered. The energy sector was chosen for the context of the study, as it represents an industry in which the usage of AI is common (Tata Consultancy Services, 2020). Furthermore, as the service represents a commodity, there should be no brand preference, which would distort the results. In the online chat, to initially conceal the identity of the chatbot, the conversational partner did not present itself as a bot, but simply introduced himself as "Leon." Pieces of the conversation were presented to the participants in sequence. During the conversation, the customer's issue was resolved in the chat, in that the electricity contract was successfully reregistered with the new meter at the new address. At the end of the conversation, it was revealed to half of the participants that the service agent of the

presented chat dialogue was in fact not a human person, but a chatbot. The other half of the participants did not receive this information, but instead read the text that they may now close the chat window. Apart from criticality and disclosure manipulations, the course of the chatbot interactions was of identical length and depth. For screenshots of the disclosure scenario, see Figure 2. For a full description of all scenarios, see Appendix 1.

Measures, manipulation checks and validity. After going through these scenarios of service encounters, measures for trust and customer retention were taken (see Table 1 for items). The study further contained manipulation and attention checks. Multi-item constructs were measured by taking the mean of participants' statements on seven-point Likert scales, anchored by 1 = strongly disagree and 7 = strongly agree. The initial sample consisted of 249 participants. Those who failed to answer attention checks correctly ("Please tick the scale point (5) here to show that you have read the questionnaire carefully" and "What was the service request you approached the company with?") and those who did not fill out the survey conscientiously ("I have answered the questionnaire conscientiously") or in an unrealistic completion time were discarded from further analyses (Haenel *et al.*, 2019). The effective sample thus consisted of 201 participants (45% female, M_{age} = 38 years). The scenarios were perceived as realistic, with a mean of 6.24 and an SD of 0.97 on a 7-point scale ("The scenario is realistic" anchored by (1) strongly disagree and (7) strongly agree) (Bagozzi *et al.*, 2016).

The manipulation check for perceived identity ("Please rate whether you think you talked to an automated chatbot or a human service employee" anchored by automated chatbot (1) and human service employee (7)) (Go and Sundar, 2019) is significant at $p < 0.001$, with respondents in the chatbot disclosure scenario perceiving their conversational partner significantly more as a chatbot than in the no chatbot disclosure scenario ($M_{disclosed}$ = 1.83,

Figure 2.
Exemplary scenario

Construct	Measurement	Study 1 Item loadings	Study 1 α	Study 1 AVE	Study 1 CR	Study 2 Item loadings	Study 2 α	Study 2 AVE	Study 2 CR
Trust in conversational partner (Bhattacherjee 2002)	The conversational partner has the necessary skills to deliver the service[1]	0.84	0.90	0.63	0.92	0.84	0.94	0.72	0.95
	The conversational partner has access to the information needed to handle my service request adequately[1]	0.83				0.85			
	The conversational partner is fair in its conduct of my service request[2]	0.85				0.88			
	The conversational partner has high integrity[2]	0.76				0.82			
	The conversational partner is receptive to my service request[3]	0.80				0.85			
	The conversational partner makes efforts to address my service request[3]	0.76				0.88			
	Overall, the conversational partner is trustworthy[4]	0.71				0.81			
Customer retention (Wallenburg 2009; Bhattacherjee *et al.*, 2012)	I would continue being a customer of the energy provider	0.80	0.91	0.69	0.93	0.91	0.95	0.81	0.96
	I would extend my existing contract with the energy provider when it expires	0.84				0.92			
	If I had to decide, I would again select this energy provider	0.78				0.87			
	I would terminate my existing contract with the energy provider (*R*)	0.88				0.89			
	I would intend to switch my energy provider (*R*)	0.85				0.91			
	I would plan to abandon the energy provider (*R*)	0.85				0.91			

Note(s): *R* = reverse scaled items, α = Cronbachs alpha, AVE = average variance extracted, CR = construct reliability, all item loadings are significant at $p < 0.001$; Trust Dimensions: [1]competence, [2]integrity, [3]benevolence, [4]overall trust

Table 1.
Measures of multi-item constructs, indicator and construct reliability

$SD = 1.3$; $M_{undisclosed} = 3.02$, $SD = 1.8$; $t = 5.36$). The manipulation check for perceived service criticality ("The resolution of my service request is. . . uncritical (1) / critical (7) to me") (Webster and Sundaram, 1998) is significant at $p < 0.01$, with respondents in the high criticality scenario perceiving it more critical than in the low criticality scenario ($M_{highcriticality} = 5.38$, $SD = 1.8$; $M_{lowcriticality} = 4.71$, $SD = 1.9$; $t = -2.56$).

Construct reliability and validity of the two multi-item constructs (trust in conversational partner and customer retention) were examined by employing different methods. First, in both studies, all Cronbach's alpha and composite reliability measures are above the cutoff value of 0.7, indicating construct-level reliability (see Table 1) (Hulland *et al.*, 2018). Second, the Fornell and Larcker (1981) test indicates initial evidence for convergent validity as the average variance extracted (AVE) for each multiple-item construct exceeds 0.50 and is larger than their shared variance (Hulland *et al.*, 2018). Third, as suggested by prior research, this research additionally relies on the heterotrait–monotrait (HTMT) method to further demonstrate discriminant validity (Henseler *et al.*, 2015; Krämer *et al.*, 2020). Estimating the HTMT ratios of the two multi-item constructs in study 1 (study 2) yields a value of 0.62 (0.63), which is well below the conservative cutoff value of 0.85. The upper limit of the 97.5% bias-corrected confidence interval in study 1 (study 2) is 0.74 (0.73), which strengthens the confidence in the discriminant validity exhibited by the focal constructs.

Method and results. To first test the effect of the manipulations on trust, analysis of variance (ANOVA) is used. Chatbot disclosure, service criticality and the interaction of disclosure and service criticality were used as independent variables and trust in the conversational partner as the dependent variable.

The results show a significant negative main effect of chatbot disclosure on trust ($M_{disclosed} = 6.18$, $SE = 0.08$; $M_{undisclosed} = 6.45$, $SE = 0.08$; $F = 5.86$, $p = 0.016$), which is in line with extant studies that suggest negative responses. The main effect of service criticality is not significant ($M_{highcriticality} = 6.33$, $SE = 0.08$; $M_{lowcriticality} = 6.30$, $SE = 0.08$; $F = 0.11$, $p = 0.738$). The interaction effect of chatbot disclosure and service criticality is marginally significant ($F = 3.00$, $p = 0.085$). This suggests that at least two treatment conditions do not yield significant differences. Thus, to identify nuances in the effects of the four scenarios on trust, we used planned contrasts. There was no effect of chatbot disclosure on trust if service criticality is low ($M_{undiscl \times lowcrit} = 6.33$, $M_{discl \times lowcrit} = 6.26$, $t = -0.47$, $p = 0.638$). However, results show that there is a negative effect of chatbot disclosure if the service criticality is high ($M_{undiscl \times highcrit} = 6.56$, $M_{discl \times highcrit} = 6.10$, $t = -3.04$, $p = 0.003$). Despite seemingly high mean values in trust in all scenarios, the effect size measure (Cohen's *d*) for this effect amounts to 0.6. Taken together, these results provide support for hypothesis 1, which stated that disclosing chatbot identity reduces trust more if service criticality is high. Figure 3 illustrates

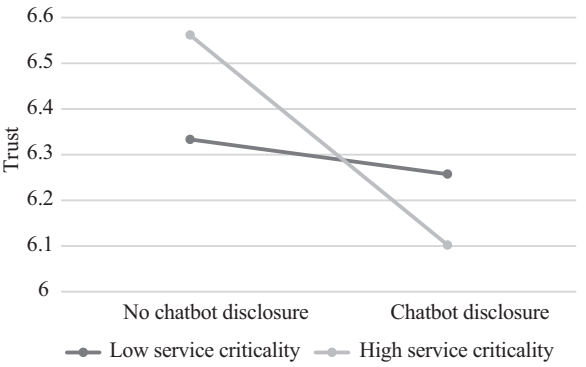

Figure 3.
Study 1: Interaction of chatbot disclosure × service criticality on trust

the interaction effect of chatbot disclosure and service issue on trust, predictive margins can be found in Table 2.

To test hypothesis 2, a mediation analysis was further conducted using the products of coefficient method to estimate the indirect effects and bias-corrected bootstrapped confidence intervals (Zhao *et al.*, 2010). Results refer to the effect of chatbot disclosure in high criticality service issues and are shown in Table 3. In line with the expectations, results show that the interaction of chatbot disclosure and service criticality has a significant negative indirect effect on retention through trust ($\beta = -0.265$, lower-level confidence interval [LLCI] = −0.526; upper-level confidence interval [ULCI] = −0.065) because the 95% confidence intervals do not include zero, supporting hypothesis 2. There were no significant direct effects of the interaction term of chatbot disclosure and service criticality on retention, suggesting full (or indirect-only) mediation (Zhao *et al.*, 2010).

Post hoc analysis. In order to gain a deeper understanding of the effect of chatbot disclosure for different levels of service criticality, a post hoc analysis was conducted where trust in the conversational partner was disentangled in its three subdimensions: competence, integrity and benevolence. To estimate the effects, three ANOVAs were conducted, followed by pairwise comparisons of predictive margins. In line with the main analysis, chatbot disclosure had no effect on trust dimension when service criticality was low ($\Delta_{Comp} = -0.02$, $SE = 0.18$, $t = -0.09$, $p = 0.929$; $\Delta_{Int} = -0.14$, $SE = 0.19$, $t = -0.72$, $p = 0.469$; $\Delta_{Bene} = -0.08$, $SE = 0.17$, $t = -0.47$, $p = 0.638$). However, for high criticality service issues, results show that there is a significant decrease in perceptions of competence and benevolence ($\Delta_{Comp} = -0.58$, $SE = 0.16$, $t = -3.55$, $p < 0.001$; $\Delta_{Bene} = -0.33$, $SE = 0.16$, $t = -2.09$, $p = 0.038$), when chatbot identity is disclosed. The effect of chatbot disclosure on integrity perceptions was insignificant for high criticality services ($\Delta_{Int} = -0.29$, $SE = 0.18$, $t = -1.59$, $p = 0.114$). An overview of the results can be found in Table 4 and Figure 4.

Study 2: Chatbot disclosure and service outcome

Study design. The goal of study 2 was to examine the effect of disclosing the nonhuman chatbot identity on consumer trust and retention for different service outcomes. Therefore, another 2 (chatbot disclosure vs no disclosure) × 2 (chatbot failure vs no chatbot failure)

	Study 1		Study 2	
	Low service criticality	High service criticality	No chatbot failure	Chatbot failure
No chatbot disclosure	$M = 6.33$ $SE = 0.12$	$M = 6.56$ $SE = 0.10$	$M = 6.28$ $SE = 0.17$	$M = 3.55$ $SE = 0.16$
Chatbot disclosure	$M = 6.26$ $SE = 0.10$	$M = 6.10$ $SE = 0.11$	$M = 6.27$ $SE = 0.16$	$M = 4.06$ $SE = 0.16$

Note(s): $N_{Study1} = 201$; $N_{Study2} = 197$; M = mean; SE = standard error

Table 2. Predictive margins for trust

Study	Path	*Coeff.*	*SE*	*LLCI*	*ULCI*	Mediation
Study 1	Chatbot disclosure × High service criticality → Trust → Retention	−0.265	0.118	−0.526	−0.065	✓
Study 2	Chatbot disclosure × Chatbot failure → Trust → Retention	0.109	0.083	0.005	0.348	✓

Note(s): Number of bootstrap samples = 5,000; *Coeff.* = coefficient; *SE* = standard error; *LLCI* = 95 % lower level confidence interval; *ULCI* = 95 % upper level confidence interval

Table 3. Mediation testing

	Study 1		Study 2	
	Competence		*Competence*	
	Low service criticality	High service criticality	No chatbot failure	Chatbot failure
No chatbot disclosure	$M = 6.40$, $SE = 0.13$	$M = 6.75$, $SE = 0.11$	$M = 6.41$, $SE = 0.21$	$M = 3.13$, $SE = 0.19$
Chatbot disclosure	$M = 6.38$, $SE = 0.11$	$M = 6.16$, $SE = 0.12$	$M = 6.39$, $SE = 0.19$	$M = 3.12$, $SE = 0.20$
	Integrity		*Integrity*	
	Low service criticality	High service criticality	No chatbot failure	Chatbot failure
No chatbot disclosure	$M = 6.36$, $SE = 0.15$	$M = 6.42$, $SE = 0.12$	$M = 6.22$, $SE = 0.20$	$M = 3.77$, $SE = 0.18$
Chatbot disclosure	$M = 6.22$, $SE = 0.13$	$M = 6.13$, $SE = 0.14$	$M = 6.21$, $SE = 0.18$	$M = 4.58$, $SE = 0.19$
	Benevolence		*Benevolence*	
	Low service criticality	High service criticality	No chatbot failure	Chatbot failure
No chatbot disclosure	$M = 6.46$, $SE = 0.13$	$M = 6.69$, $SE = 0.10$	$M = 6.52$, $SE = 0.21$	$M = 3.58$, $SE = 0.19$
Chatbot disclosure	$M = 6.38$, $SE = 0.11$	$M = 6.36$, $SE = 0.12$	$M = 6.53$, $SE = 0.19$	$M = 4.39$, $SE = 0.19$

Table 4. Predictive margins for trust dimensions

between-subject experiment was conducted. Participants were gathered through the same panel provider as in study 1, while making sure there was no overlap between the samples. For chatbot design, the same materials as in study 1 were used. Service criticality was held constant across scenarios. This means the scenarios were identical to study 1, except that there was no manipulation of high service criticality. Participants had to imagine that they were customers of an energy provider and about to use the online chat to contact the energy provider to reregister their contract to the new address. Participants were randomly assigned to one of the service outcome conditions. In the chatbot failure condition, the conversational partner was not able to handle the customer's inquiry and thus could not resolve the customer issue, whereas in the other condition, the customer's issue was handled successfully. Identically to study 1, at the end of the conversation, half of the participants received the information that the conversational partner was a chatbot. For a full transcript of the scenarios, see Appendix 2.

Measures, manipulation checks and validity. After going through the scenarios, measures on trust, retention, demographics and manipulation checks were collected. All items and scales used were identical to study 1, except for the manipulation check used for service outcome. Furthermore, study 2 includes a control variable on attribution of responsibility, as in failure settings, consumers will attribute the outcome of the service interaction differently than in settings where no failure occurs ("Which entity was responsible for the service outcome?" anchored by "I myself" (1) and "my conversational partner" (7)) (Belanche *et al.*, 2020; Russel, 1982).

The initial sample consisted of 254 participants. Again, those who failed to attention checks and did not fill out the survey conscientiously were discarded from further analyses. The effective sample consisted of 197 participants (50% female, $M_{age} = 38$ years).

The scenarios were perceived as realistic, with a mean of 6.13 and an *SD* of 1.18 on the 7-point perceived realism scale. The manipulation check for perceived identity is significant at $p < 0.01$, with respondents in the chatbot disclosure scenario perceiving their conversational partner significantly more as a chatbot than in the no chatbot disclosure scenario ($M_{disclosed} = 2.14$, $SD = 1.5$; $M_{undisclosed} = 2.68$, $SD = 1.7$; $t = 2.41$). The manipulation check for service outcome ("How would you describe the outcome of the service interaction?" anchored by success (1) and failure (7)) (Belanche *et al.*, 2020) is significant at $p < 0.001$, with

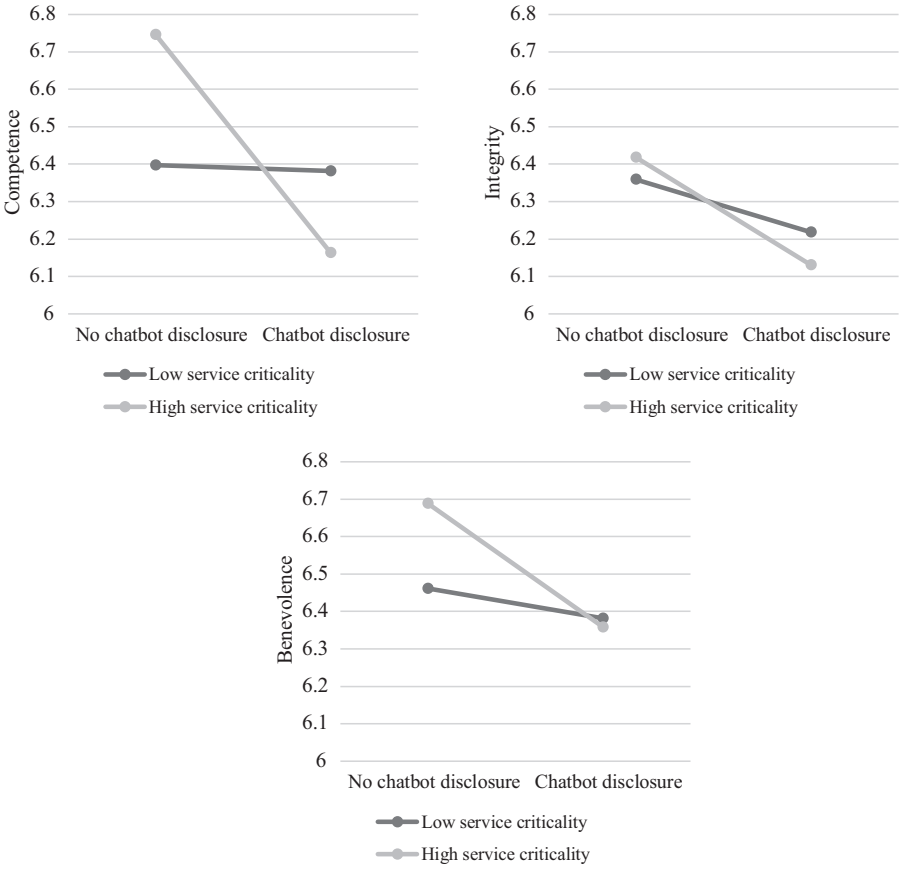

Figure 4.
Study 1: Predictive margins for trust dimensions

respondents in the chatbot failure scenario perceiving the service outcome significantly more as a failure than participants in the no chatbot failure scenario ($M_{\text{failure}} = 1.32$, $SD = 0.9$; $M_{\text{nofailure}} = 6.71$, $SD = 0.7$; $t = 45.97$).

Validity measures are reported in the description of study 1 as well as in Table 1.

Method and results. To test the effect of chatbot disclosure on trust for different service outcomes, an ANCOVA was conducted. Chatbot disclosure, service outcome and the interaction of disclosure and service outcome were used as independent variables, responsibility attribution as a covariate and trust in the conversational partner as the dependent variable. Responsibility attribution did not yield any significant effects on the dependent variable in any of the analyses and excluding the variable from the model did not change the results.

There was no significant main effect of chatbot disclosure on trust ($M_{\text{disclosed}} = 5.17$, $SE = 0.11$; $M_{\text{undisclosed}} = 4.91$, $SE = 0.12$; $F = 2.40$, $p = 0.123$). Not surprisingly, the main effect of service outcome on trust is negative ($M_{\text{failure}} = 3.80$, $SE = 0.11$; $M_{\text{nofailure}} = 6.28$, $SE = 0.12$; $F = 222.42$, $p < 0.001$). The interaction of chatbot disclosure and service outcome does not yield an effect on trust ($F = 2.66$, $p = 0.104$). To identify whether there are nuances in the effects of the four scenarios on trust, planned contrasts were used. Chatbot disclosure had no significant

effect on trust when no chatbot failure occurred ($M_{undiscl \times nofailure} = 6.28$, $M_{discl \times nofailure} = 6.27$, $t = -0.06$, $p = 0.953$). However and interestingly, the effect of chatbot disclosure on trust is positive in case of a chatbot failure ($M_{undiscl \times failure} = 3.55$, $M_{discl \times failure} = 4.06$, $t = 2.30$, $p = 0.023$), supporting hypothesis 3, which stated that disclosing chatbot identity enhances trust if the service outcome is a failure. The effect size measure for this effect amounts to $d = -0.36$. For an illustration of the interaction, see Figure 5.

To further test hypothesis 4, mediation analysis was conducted. Results refer to the effect of chatbot disclosure when the chatbot failure occurred and are shown in Table 3. As expected, results show a significant positive indirect effect of chatbot disclosure on retention, when a chatbot failure occurred ($\beta = 0.109$, LLCI = 0.005; ULCI = 0.348), supporting H4. There were no significant direct effects of chatbot disclosure or service outcome on retention, suggesting full (or indirect-only) mediation (Zhao *et al.*, 2010).

Post hoc analysis. Mirroring the analysis procedure of study 1, a post hoc analysis was conducted in study 2 to deepen the understanding of the trust mechanism. The ANCOVA results and pairwise comparisons of predictive margins show that when considering the separate trust dimensions, chatbot disclosure has no significant effects on any of the trust dimensions when no chatbot failure occurred, which is in line with the results of the main analysis ($\Delta_{Comp} = -0.02$, $SE = 0.28$, $t = -0.07$, $p = 0.947$; $\Delta_{Int} = -0.01$, $SE = 0.27$, $t = -0.05$, $p = 0.958$; $\Delta_{Bene} = 0.01$, $SE = 0.28$, $t = 0.01$, $p = 0.988$). However, in the case of chatbot failure, chatbot disclosure positively affected perceptions of integrity ($\Delta_{Int} = 0.82$, $SE = 0.26$, $t = 3.13$, $p = 0.002$) as well as benevolence ($\Delta_{Bene} = 0.81$, $SE = 0.27$, $t = 3.02$, $p = 0.003$). There was no significant effect of chatbot disclosure on competence perceptions, when a chatbot failure occurred ($\Delta_{Comp} = -0.01$, $SE = 0.27$, $t = -0.04$, $p = 0.967$). An overview of the predictive margins and visualization of results can be found in Table 4 and Figure 6.

An additional analysis on differences in responsibility attribution shows that the results are subject to the self-serving bias, that is, attributing responsibility for positive outcomes to oneself, while blaming external factors for negative outcomes (Miller and Ross, 1975) ($M_{failure} = 5.64$, $SD = 1.5$; $M_{nofailure} = 4.97$, $SD = 1.5$; $t = -3.09$; $p = 0.001$). Interestingly, this bias emerges independently from whether chatbot disclosure occurred.

General discussion

The goal of this article was to provide empirical insight on the effects of chatbot disclosure on trust and customer retention. In a similar vein as prior research, study 1 finds that chatbot disclosure will negatively impact retention through trust if the conversation proceeds flawlessly, that is, if the chatbot delivers the expected service. More specifically, for highly critical services, if the chatbot is able to solve the customer's issue, disclosure will negatively

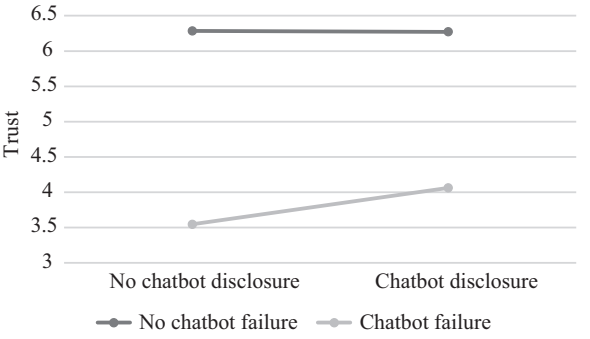

Figure 5.
Study 2: Interaction of chatbot disclosure × service outcome on trust

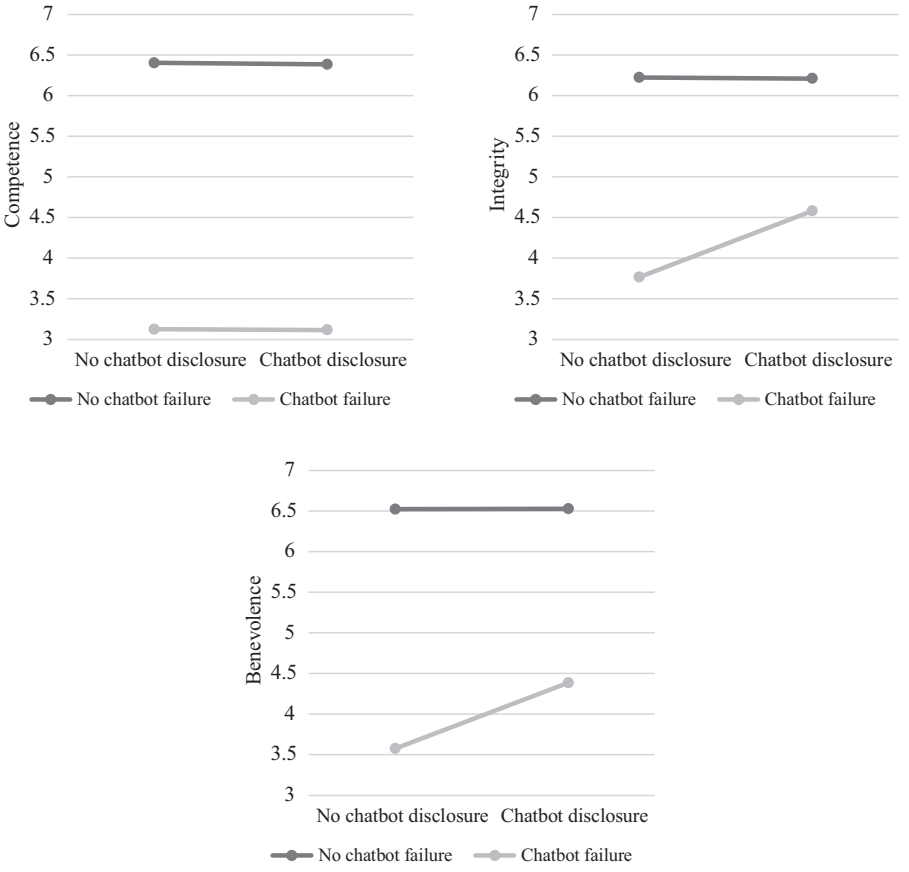

Figure 6.
Study 2: Predictive margins for trust dimensions

impact customer trust and thus hamper retention. In contrast, for less critical services, chatbot disclosure does not impact trust at all.

Interestingly, examining the three trust dimensions individually suggests that the loss of trust stems from lower perceptions of conversational partner's competence and benevolence, but not lower perceptions of integrity. The decrease in competence perceptions suggests that despite identical service interaction and outcome, conversational partners that are disclosed as chatbots are perceived as less knowledgeable. Furthermore, the loss in benevolence belief can be explained by lower perceived empathy (Bhattacherjee, 2002). Both results are in line with results of Luo *et al.* (2019), who did not examine trust, but perceived chatbot knowledge and perceived chatbot empathy. The fact that integrity perception remains unaffected by chatbot disclosure could be explained by the fact that integrity is perceived as a constituent characteristic of any successful service delivery (Mayer *et al.*, 1995).

Study 2 further showed that, as expected, the main driver of trust in the conversational partner was whether or not the customer's issue could be solved (de Matos *et al.*, 2007; Kelley *et al.*, 1993). This result is well established in service research and should therefore not be the focus of this discussion. However, chatbot disclosure still plays a significant role: the interaction effect shows that a significant increase in trust can be observed if chatbot identity

is disclosed in a chatbot failure setting. This suggests that when the customer's issue cannot be resolved in the online chat, chatbot disclosure helps to mitigate the negative failure effect. Therefore, disclosing a chatbot's nonhuman identity enhances overall trust and thus retention (instead of mitigating it) in cases where the chatbot fails to deliver the expected service. This work theorized that this happens because the disclosure offers a type of explanation for the negative outcome being perceived like an "apology." The latter represents a well-established recovery instrument, which apparently is effective and "accepted" by customers even in situations where a machine delivers the apology and not a human. Moreover, this tactic even works when not provided explicitly but implicitly. While trust in the failure setting is significantly lower than in the no failure setting, study 2 was able to demonstrate a positive reaction to chatbot disclosure, compared to no disclosure.

Taking a closer look at the data reveals that the positive reaction to chatbot disclosure was not a result of a higher belief in the chatbot's competence, but in its integrity and benevolence. Thus, while chatbot disclosure does not result in more favorable perceptions of the chatbot's abilities, it does indeed ameliorate perceptions of integrity and benevolence in cases where the chatbot is unable to handle the customer's service issue. This is in line with findings from service robot research, which show that after process-type failures, machine-like robots are evaluated as more empathetic and warm than human-like robots, while there is no significant difference in competence perceptions of the two in failure situations (Choi *et al.*, 2020).

The effects could be a result of differences in trust resilience between undisclosed and disclosed chatbots (de Visser *et al.*, 2016). When experiencing a failure, disclosure relieves the customer of the attributional search and creates certainty. In case identity remains undisclosed following a failure, the conversational partner is perceived as abstract and less tangible. Trust resilience research shows that feedback, which reduces uncertainty, will increase trust. This would imply that when experiencing a failure, high uncertainty perceptions of interactions due to an undisclosed agent result in customers considering their conversational partner to be less empathetic and less of integrity than disclosed chatbots.

All in all, the results of the studies confirm the theoretical mechanisms of attribution theory in the context of chatbot disclosure (Davison and Martinsons, 2016). That is, the results demonstrate a spontaneous and biased attribution for high criticality services despite being addressed successfully (confirming the negative outcomes suggested in extant studies) as well as an elaborated attribution following a negative service outcome (pointing to positive outcomes that diverge from the current view).

The results challenge and advance the current thinking on chatbot-based service delivery. They challenge the skeptical view on disclosing the nonhuman nature of service agents prevalent in extant studies by empirically demonstrating that chatbot disclosure can lead to positive outcomes in certain service frontline settings. In providing several insights that explain this diverging observation, this research advances current chatbot literature. First, this research offers insights on how customers respond to chatbot disclosure by showing that trust mediates the relationship between chatbot disclosure and customer retention. Also, the studies demonstrate that the service context shapes the trust response to chatbot disclosure in that this response can be positive and negative depending on the context and yielding corresponding retention outcomes.

Implications

Implications for consumers

This research complements existing literature on consumer reactions to service delivery with chatbots in that it highlights that consumers today show a general skepticism toward services delivered by bots instead of humans. Interestingly, the results show that consumers

do trust chatbots to deliver services with low criticality. However, when it comes to service inquiries with higher criticality, consumers do not feel secure to rely on bots, which will hamper their service experience and relationship with the service provider in turn. The results further suggest that consumers may find comfort in transparency, if the transparently communicated information offers an explanation for the service outcome that matches their disposition.

All in all, chatbot technology offers a variety of opportunities for consumers. For instance, chatbots enable consumers to engage in service interactions around the clock and enable them to resolve their service requests more efficiently than when interacting with potentially stressed and unfriendly human employees (Luo *et al.*, 2019). However, apparently consumers fail to acknowledge these advantages, which might rob them of benefits of higher efficiency provided by chatbots, as they remain skeptical toward the technology.

Notably, from a consumer's perspective, understanding the repercussions of chatbot disclosure (or broader, artificial intelligence disclosure) gains relevance due to current technological advancements, such as Google Duplex. In situations like this, the undisclosed, highly anthropomorphic conversational agent does not represent the service agent, but the customer and is able to, for example, make reservations or appointments with firms in the place of the customer. That means customers are replaced by artificial intelligence, which then interacts with human or nonhuman service frontline employees. This "inverted" perspective is currently radically transforming service encounter frameworks in service research (Robinson *et al.*, 2020). Thus far, how reactions on disclosure versus nondisclosure will differ for this perspective has not yet been investigated.

Implications for service providers and chatbot designers

As a core innovation in customer service, proper implementations of chatbots in the service frontline are crucial for customer retention and hence firm profitability. The study results suggest that firms' decision whether to disclose chatbot identity should be informed by and aligned with contextual service factors. If chatbot identity is disclosed in high criticality service settings, stereotypes regarding chatbots dominate customers' thinking and customer trust may be reduced, which in turn is detrimental for the customer relationship.

However, when chatbot identity is disclosed in response to a chatbot failure, it offers an explanation and relieves the customer of the search for the cause of failure. Therefore, "merely" disclosing chatbot identity serves as an actionable and inexpensive instrument for failure recovery. This shows that there are cases in which chatbot disclosure does have positive effects on business-relevant consumer behavior. Chatbot disclosure is thus a viable lever service providers can use for damage control in case of chatbot failure. These findings have also direct implication for chatbot designers responsible for engineering the algorithms behind the bot: Once a chatbot cannot solve the service issue, an automated message should be triggered informing the customer about the artificial intelligence nature of the conversational agent serving as an "apology" for the failure and making the customer forgive.

Of course, service providers should continue striving for error-free service delivery with chatbots. However, when recognizing that a complete elimination of service failures is "an insurmountable task" (Webster and Sundaram, 1998, p. 153), finding instruments of mitigating negative effects of service failures only gains more relevance. Overall, firms should focus on creating trust through and in chatbot communication, so that in the long term, consumers' beliefs about chatbot abilities evolve corresponding to actual chatbot abilities and the current aversion toward bots evolves to appreciation. In doing so, the currently prevailing view of a negative relationship between chatbot disclosure and customer retention might be mitigated or even reversed.

Implications for society

Both positive and negative effects found in this research originate from consumers' initial mistrust in chatbots or more generally, algorithmic entities. Apart from situations where chatbot disclosure can mitigate negative effects from chatbot failure, the results of this paper imply that companies should prevent disclosing the algorithmic identity to their customers in online chats in order to not hamper customer retention. However, it is questionable if withholding chatbot identity is tenable ethically and legally in the long term. As chatbots become increasingly anthropomorphic, they should be implemented using "ethical anthropomorphism" (Kaminski *et al.*, 2017; Thomaz *et al.*, 2020). Specifically, this means that anthropomorphism should not be used to intentionally mislead or even manipulate consumers (Leong and Selinger, 2019). For instance, the state of California has already passed a bill to prevent companies from doing so for political and commercial bots (California Legislative Information, 2018). If this development gained traction worldwide and disclosure became legally inevitable, based on the largely negative effects shown in prior studies, firms would have to scrutinize whether they deploy chatbots at all. While this research finds negative reactions to chatbot disclosure too, the results also prove that disclosure can in fact produce positive reactions. The way forward should thus not be to question deployment of chatbots, but to develop a disclosure strategy that consistently produces positive outcomes and is ethically tenable.

Research implications and limitations

In order to come up with a successful disclosure strategy, further research should shift the discussion from whether and under what circumstances to disclose chatbot identity to how to disclose chatbot identity to create trust and customer retention. First studies are tackling this question by adding further explanations to chatbot disclosure that frame the chatbot perception in a desired manner and find that it might be possible to create trust levels as high as that of undisclosed bots (Mozafari *et al.*, 2021). A theoretically grounded approach to examine different framing strategies for chatbot disclosure should be provided with the goal to minimize negative reactions and even produce positive reactions beyond those that were found in the results of this paper.

Finally, the present study is not free of limitations. In this study, screenshots of a conversation were shown to the study participants to ensure high internal validity. Future studies should let study participants interact with a chatbot themselves in order to increase external validity. Additionally, the results of study 1 showed that in service interactions that proceed flawlessly, trust in the conversational partner is generally quite high. While the analyses showed significant differences in trust for disclosed and undisclosed bots, the differences in trust could be more prominent if trust was measured after disclosure, but prior to service delivery. In this way, further studies would be able to assess consumer attitudes and beliefs toward chatbots even better. In contrast, study 2 showed a wide gap in trust caused by the failure manipulation. Future studies should consider examining repercussions of chatbot disclosure in settings with different failure types or failure severities. It can be assumed that if a failure is less severe, the difference in trust will not be as large, and as a consequence, the effect size of the chatbot disclosure should be stronger.

Also, the results of study 2 suggest the existence of a self-serving bias in a chatbot context, along the lines of "claim success, but blame the bot." This introduces a new perspective on research on outcome attributions in service encounters, which has been the subject of some existing studies on human–robot interactions (Belanche *et al.*, 2020) and could dive into even deeper in future research. Whether the emergence of this bias is a result of the "Computers are Social Actors" paradigm (Nass *et al.*, 1994) or has other explanations remains unanswered in this research. Further research could analyze whether the consequences of the self-serving bias in chatbot interactions are comparable to face-to-face human encounters.

References

Adiwardana, D., Luong, M.-T., So, D.R., Hall, J., Fiedel, N., Thoppilan, R., Yang, Z., Kulshreshtha, A., Nemade, G., Lu, Y. and Le, V.Q. (2020), "Towards a human-like open-domain chatbot", available at: http://arxiv.org/pdf/2001.09977v2.

Bagozzi, R.P., Belanche, D., Casaló, L.V. and Flavián, C. (2016), "The role of anticipated emotions in purchase intentions", *Psychology and Marketing*, Vol. 33 No. 8, pp. 629-645.

Belanche, D., Casaló, L.V., Flavián, C. and Schepers, J. (2020), "Robots or frontline employees? Exploring customers' attributions of responsibility and stability after service failure or success", *Journal of Service Management*, Vol. 31 No. 2, pp. 267-289.

Berry, L.L. and Parasuraman, A. (1991), *Marketing Service*, The Free Press, New York.

Bhattacherjee, A. (2002), "Individual trust in online firms: scale development and initial test", *Journal of Management Information Systems*, Vol. 19 No. 1, pp. 211-241.

Bhattacherjee, A., Limayem, M. and Cheung, C.M. (2012), "User switching of information technology: a theoretical synthesis and empirical test", *Information & Management*, Vol. 49 No. 7-8, pp. 327-333, doi: 10.1016/j.im.2012.06.002.

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

California Legislative Information (2018), "SB-1001 bots: disclosure", available at: https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180SB1001 (accessed 21 April 2020).

Candello, H., Pinhanez, C. and Figueiredo, F. (2017), "Typefaces and the perception of humanness in natural language chatbots", *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ACM Press, Denver, CO, pp. 3476-3487.

Choi, S., Mattila, A.S. and Bolton, L.E. (2020), "To err is human(-oid): how do consumers react to robot service failure and recovery?", *Journal of Service Research*, pp. 1-18, available at: https://journals.sagepub.com/doi/10.1177/1094670520978798.

Davison, R.M. and Martinsons, M.G. (2016), "Context is king! Considering particularism in research design and reporting", *Journal of Information Technology*, Vol. 31 No. 3, pp. 241-249.

Dawes, R.M. (1979), "The robust beauty of improper linear models in decision making", *American Psychologist*, Vol. 34 No. 7, pp. 571-582.

De Keyser, A., Köcher, S., Alkire, L., Verbeeck, C. and Kandampully, J. (2019), "Frontline Service Technology infusion: conceptual archetypes and future research directions", *Journal of Service Management*, Vol. 30 No. 1, pp. 156-183.

de Matos, C.A., Henrique, J.L. and Alberto Vargas Rossi, C. (2007), "Service recovery paradox: a meta-analysis", *Journal of Service Research*, Vol. 10 No. 1, pp. 60-77.

de Visser, E.J., Monfort, S.S., McKendrick, R., Smith, M.A.B., 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.

Dietvorst, B.J., Simmons, J.P. and Massey, C. (2015), "Algorithm aversion: people erroneously avoid algorithms after seeing them err", *Journal of Experimental Psychology: General*, Vol. 144 No. 1, pp. 114-126.

Donath, J. (1999), "Identity and deception in the virtual community", in Kollock, P. and Smith, M. (Eds), *Communities in Cyberspace*, Routledge, London.

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.

Forsyth, D.R. (1987), *Social Psychology*, Brooks/Cole Publ, Pacific Grove, CA.

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.

Go, E. and Sundar, S.S. (2019), "Humanizing chatbots: the effects of visual, identity and conversational cues on humanness perceptions", *Computers in Human Behavior*, Vol. 97, pp. 304-316, doi: 10.1016/j.chb.2019.01.020.

Haenel, C.M., Wetzel, H.A. and Hammerschmidt, M. (2019), "The perils of service contract divestment: when and why customers seek revenge and how it can Be attenuated", *Journal of Service Research*, Vol. 22 No. 3, pp. 301-322.

Hart, C.W.L. and Johnson, M.D. (1999), "Growing the trust relationship", *Marketing Management*, Vol. 8 No. 1, pp. 9-19.

Heider, F. (1958), *The Psychology of Interpersonal Relations*, Lawrence Erlbaum Associates, Hillsdale, NJ.

Hendriks, F., Ou, C., Amiri, A.K. and Bockting, S. (2020), "The power of computer-mediated communication theories in explaining the effect of chatbot introduction on user experience", *Proceedings of the 53rd Hawaii International Conference on System Sciences*.

Henseler, J., Ringle, C.M. and Sarstedt, M. (2015), "A new criterion for assessing discriminant validity in variance-based structural equation modeling", *Journal of the Academy of Marketing Science*, Vol. 43 No. 1, pp. 115-135.

Holtgraves, T.M., Ross, S.J., Weywadt, C.R. and Han, T.L. (2007), "Perceiving artificial social agents", *Computers in Human Behavior*, Vol. 23 No. 5, pp. 2163-2174.

Hrebiniak, L.G. (1974), "Effects of job level and participation on employee attitudes and perceptions of influence", *Academy of Management Journal*, Vol. 17 No. 4, pp. 649-662.

Huang, M.-H. and Rust, R.T. (2018), "Artificial intelligence in service", *Journal of Service Research*, Vol. 21 No. 2, pp. 155-172.

Hulland, J., Baumgartner, H. and Smith, K.M. (2018), "Marketing survey research best practices: evidence and recommendations from a review of JAMS articles", *Journal of the Academy of Marketing Science*, Vol. 46 No. 1, pp. 92-108.

Ishowo-Oloko, F., Bonnefon, J.-F., Soroye, Z., Crandall, J., Rahwan, I. and Rahwan, T. (2019), "Behavioural evidence for a transparency–efficiency tradeoff in human–machine cooperation", *Nature Machine Intelligence*, Vol. 1 No. 11, pp. 517-521.

Jussupow, E., Benbasat, I. and Heinzl, A. (2020), *Why Are We Averse towards Algorithms? A Comprehensive Literature Review on Algorithm Aversion*.

Kaminski, M.E., Rueben, M., Smart, W.D. and Grimm, C.M. (2017), "Averting robot eyes", *Maryland Law Review*, Vol. 76 No. 4, pp. 983-1024.

Kanazawa, S. (1992), "Outcome or expectancy? Antecedent of spontaneous causal attribution", *Personality and Social Psychology Bulletin*, Vol. 18 No. 6, pp. 659-668.

Kelley, S.W., Hoffman, K.D. and Davis, M.A. (1993), "A typology of retail failures and recoveries", *Journal of Retailing*, Vol. 69 No. 4, pp. 429-452.

Komiak, S.X. and Benbasat, I. (2004), "Understanding customer trust in agent-mediated electronic commerce, web-mediated electronic commerce, and traditional commerce", *Information Technology and Management*, Vol. 5 No. ½, pp. 181-207.

Krämer, T., Weiger, W.H., Gouthier, M.H.J. and Hammerschmidt, M. (2020), "Toward a theory of spirals: the dynamic relationship between organizational pride and customer-oriented behavior", *Journal of the Academy of Marketing Science*, Vol. 22 No. 1, p. 60.

Leong, B. and Selinger, E. (2019), "Robot eyes wide shut", *Proceedings of the Conference on Fairness, Accountability, and Transparency - FAT* '19*, Atlanta, GA, 29.01.2019 - 31.01.2019, ACM Press, New York, NY, pp. 299-308.

Luger, E. and Sellen, A. (2016), *"Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents*.

Luo, X., Tong, S., Fang, Z. and Qu, Z. (2019), "Machines versus humans: the impact of AI chatbot disclosure on customer purchases", *Marketing Science*, Vol. 38 No. 6, pp. 937-947.

Mayer, R.C., Davis, J.H. and Schoorman, F.D. (1995), "An integrative model of organizational trust", *Academy of Management Review*, Vol. 20 No. 3, pp. 709-734.

McCollough, M.A., Berry, L.L. and Yadav, M.S. (2000), "An empirical investigation of customer satisfaction after service failure and recovery", *Journal of Service Research*, Vol. 3 No. 2, pp. 121-137.

Miller, D.T. and Ross, M. (1975), "Self-serving biases in the attribution of causality: fact or ficiton?", *Psychological Bulletin*, Vol. 82 No. 2, pp. 213-225.

Moorman, C., Deshpandé, R. and Zaltman, G. (1993), "Factors affecting trust in market research relationships", *Journal of Marketing*, Vol. 57 No. 1, pp. 81-101.

Morgan, R.M. and Hunt, S.D. (1994), "The commitment-trust theory of relationship marketing", *Journal of Marketing*, Vol. 58 No. 3, pp. 20-38.

Mozafari, N., Weiger, W.H. and Hammerschmidt, M. (2021), "Resolving the chatbot disclosure dilemma: leveraging selective self-presentation to mitigate the negative effect of chatbot disclosure", *Hawaii International Conference on System Sciences*, pp. 2916-2923.

Murgia, A., Janssens, D., Demeyer, S. and Vasilescu, B. (2016), "Among the Machines: Human-Bot Interaction on Social Q&A Websites", *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, San Jose, CA, pp. 1272-1279.

Nass, C. and Moon, Y. (2000), "Machines and mindlessness: social responses to computers", *Journal of Social Issues*, Vol. 56 No. 1, pp. 81-103.

Nass, C., Steuer, J. and Tauber, E.R. (1994), *Computers Are Social Actors*, Human Factors in Computing Systems, Boston, MA.

Nordheim, C.B., Følstad, A. and Bjørkli, C.A. (2019), "An initial model of trust in chatbots for customer service—findings from a questionnaire study", *Interacting with Computers*, Vol. 31 No. 3, pp. 317-335.

Nunamaker, J.F., Derrick, D.C., Elkins, A.C., Burgoon, J.K. and Patton, M.W. (2011), "Embodied conversational agent-based Kiosk for automated interviewing", *Journal of Management Information Systems*, Vol. 28 No. 1, pp. 17-48.

Ostrom, A.L. and Iacobucci, D. (1995), "Consumer trade-offs and the evaluation of services", *Journal of Marketing*, Vol. 59, pp. 17-28.

Puntoni, S., Reczek, R.W., Giesler, M. and Botti, S. (2021), "Consumers and artificial intelligence: an experiential perspective", *Journal of Marketing*, Vol. 85 No. 1, pp. 131-151.

Riedl, R., Mohr, P., Kenning, P., Davis, F. and Heekeren, H. (2011), *Trusting Humans and Avatars: Behavioral and Neutral Evidence*, Shanghai.

Robinson, S., Orsingher, C., Alkire, L., de Keyser, A., Giebelhausen, M., Papamichail, K.N., Shams, P. and Temerak, M.S. (2020), "Frontline encounters of the AI kind: an evolved service encounter framework", *Journal of Business Research*, Vol. 116, pp. 366-376.

Ross, L. (1977), "The intuitive psychologist and his shortcomings: distortions in the attribution process", in Berkowitz, L. (Ed.), *Advances in Experimental Social Psychology, Advances in Experimental Social Psychology*, Academic Press, New York, Vol. 10, pp. 173-220.

Russel, D. (1982), "The causal dimension scale: a measure of how individuals perceive causes", *Journal of Personality and Social Psychology*, Vol. 42 No. 6, pp. 1137-1145.

Sameh, A.-N., Benbasat, I. and Cenfetelli, R. (2010), "Trustworthy virtual advisors and enjoyable interactions: designing for expressiveness and transparency", *Proceedings of the European Conference on Information Systems*, Regensburg, Germany, available at: https://aisel.aisnet.org/ecis2010/116.

Schuetzler, R.M., Giboney, J.S., Grimes, G.M. and Nunamaker, J.F. Jr (2018), "The Influence of Conversational Agents on Socially Desirable Responding", *Hawaii International Conference on System Sciences*, Hawaii, USA, pp. 283-292.

Schurr, P.H. and Ozanne, J.L. (1985), "Influences on exchange processes: buyers' preconceptions of a seller's trustworthiness and bargaining toughness", *Journal of Consumer Research*, Vol. 11, pp. 939-953.

Servion (2020), "AI will power 95% of customer interactions by 2025", available at: https://www.financedigest.com/ai-will-power-95-of-customer-interactions-by-2025.html.

Sheehan, B., Jin, H.S. and Gottlieb, U. (2020), "Customer service chatbots: anthropomorphism and adoption", *Journal of Business Research*, Vol. 115, pp. 14-24.

Sherman, S. (1992), "Are strategic alliances working?", *Fortune*, pp. 77-78, September 21.

Shevat, A. (2017), *Designing Bots: Creating Conversational Experiences*, O'Reilly Media, Beijing.

Shi, W., Wang, X., Oh, Y.J., Zhang, J., Sahay, S. and Yu, Z. (2020), "Effects of persuasive dialogues: testing bot identities and inquiry strategies", *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Hawaii, USA, available at: http://arxiv.org/pdf/2001.04564v2.

Silitonga, K.A.A., Ikhsan, R.B. and Fakhrorazi, A. (2020), "Drivers of buyer retention in e-commerce: the role of transaction characteristics and trust", *Management Science Letters*, Vol. 10 No. 15, pp. 3485-3494, doi: 10.5267/j.msl.2020.6.046.

Skjuve, M., Haugstveit, I.M., Følstad, A. and Brandtzaeg, P.B. (2019), "Help! Is my chatbot falling into the uncanny valley? An empirical study of user experience in human-chatbot interaction", *Human Technology*, Vol. 15 No. 1, pp. 30-54.

Tata Consultancy Services (2020), "Getting smarter by the sector: how 13 global industries use artificial intelligence", available at: https://www.tcs.com/content/dam/tcs/pdf/Industries/global-trend-studies/ai/TCS-GTS-how-13-global-industries-use-artificial-intelligence.pdf.

Thomaz, F., Salge, C., Karahanna, E. and Hulland, J. (2020), "Learning from the Dark Web: leveraging conversational agents in the era of hyper-privacy to enhance marketing", *Journal of the Academy of Marketing Science*, Vol. 48 No. 1, pp. 43-63.

Tuzovic, S. and Paluch, S. (2018), "Conversational commerce – a new era for service business development?", in Bruhn, M. and Hadwich, K. (Eds), *Service Business Development*, Springer Fachmedien Wiesbaden, Wiesbaden, Vol. 31, pp. 81-100.

van Doorn, J., Mende, M., Noble, S.M., Hulland, J., Ostrom, A.L., Grewal, D. and Petersen, J.A. (2017), "Domo arigato Mr. Roboto", *Journal of Service Research*, Vol. 20 No. 1, pp. 43-58.

van Vaerenbergh, Y., Orsingher, C., Vermeir, I. and Larivière, B. (2014), "A meta-analysis of relationships linking service failure attributions to customer outcomes", *Journal of Service Research*, Vol. 17 No. 4, pp. 381-398.

Wallenburg, C.M. (2009), "Innovation in logistics outsourcing relationships: proactive improvement by logistics service providers as a driver of customer loyality", *Journal of Supply Chain Management*, Vol. 45 No. 2, pp. 75-93.

Wang, W. and Benbasat, I. (2008), "Attributions of trust in decision support technologies: a study of recommendation agents for E-commerce", *Journal of Management Information Systems*, Vol. 24 No. 4, pp. 249-273.

Webster, C. and Sundaram, D.S. (1998), "Service consumption criticality in failure recovery", *Journal of Business Research*, Vol. 41 No. 2, pp. 153-159.

Webster, C. and Sundaram, D.S. (2009), "Effect of service provider's communication style on customer satisfaction in professional services setting: the moderating role of criticality and service nature", *Journal of Services Marketing*, Vol. 23 No. 2, pp. 104-114.

Weiner, B. (1985), "'Spontaneous' causal thinking", *Psychological Bulletin*, Vol. 97 No. 1, pp. 74-84.

Weiner, B. (2000), "Attributional thoughts about consumer behavior", *Journal of Consumer Research*, Vol. 27 No. 3, pp. 382-387.

Wilson, H.J., Daugherty, P.R. and Morini-Bianzino, N. (2017), "The jobs that artificial intelligence will create", *MIT Sloan Management Review*, Vol. 58 No. 4, pp. 13-17.

Wirtz, J., Patterson, P.G., Kunz, W.H., Gruber, T., Lu, V.N., Paluch, S. and Martins, A. (2018), "Brave new world: service robots in the frontline", *Journal of Service Management*, Vol. 29 No. 5, pp. 907-931.

Wuenderlich, N. and Paluch, S. (2017), "A nice and friendly chat with a bot: user perceptions of AI-based service agents", *Proceedings of the International Conference on Information Systems*, pp. 1-11.

Zamora, J. (2017), *I'm Sorry, Dave, I'm Afraid I Can't Do That*, International Conference on Human Agent Interaction, Bielefeld, Germany, pp. 253-260.

Zhao, X., Lynch, J.G. and Chen, Q. (2010), "Reconsidering Baron and Kenny: Myths and truths about mediation analysis", *Journal of Consumer Research*, Vol. 37 No. 2, pp. 197-206.

Appendix 1

Study 1	Low service criticality	High service criticality
Scenario description	Please imagine that you are a customer of the fictitious energy provider NEO. Your contract number is 100218 Because you are moving to a new apartment, you want to reregister your electricity contract to your new address. Your energy provider offers you to do so via online customer service. You would like to use this service offer Assume that the conversation between you and your energy provider went according to the course of the conversation shown below	Please imagine that you are a customer of the fictitious energy provider NEO. Your contract number is 100218 Because you are moving to a new apartment, you want to reregister your electricity contract to your new address. Your energy provider offers you to do so via online customer service. You would like to use this service offer Your landlord has informed you that – if you fail to reregister your current electricity contract – you will automatically receive your electricity from the public utility company. However, the rate with the public provider is significantly higher than your previous rate with NEO. Since you want to avoid the risk of paying a higher price, it is very important to you to properly reregister your electricity contract Assume that the conversation between you and your energy provider went according to the course of the conversation shown below
Service interaction (Identical in both service criticality manipulations)	Agent [A]: Hi, my name is Leon A: How can I help you? Customer [B]: I'm moving and would like to change my electricity contract to the new apartment A: Sure thing A: Please let me know your contract number B: 100218 A: Perfect, I found your account A: What is your new address? B: Hauptstrasse 12, 31535 Neustadt A: What is the number of your new electricity meter? B: 78822 A: All right, I found the meter with the number 78822 in the system A: On what date are you moving? B: March 15, 2021 A: All right, I've recorded your entry A: Your contract has been re-registered to the meter number 78822	
Information displayed in chatbot disclosure manipulation	Your conversational partner was not a human person, but a chatbot	

Table A1. Full description of scenarios in study 1

Appendix 2

Study 2	No chatbot failure	Chatbot failure
Scenario description (Identical in both service outcome manipulations)	Please imagine that you are a customer of the fictitious energy provider NEO. Your contract number is 100218 Because you are moving to a new apartment, you want to reregister your electricity contract to your new address. Your energy provider offers you to do so via online customer service. You would like to use this service offer Assume that the conversation between you and your energy provider went according to the course of the conversation shown below	
Service interaction	Agent [A]: Hi, my name is Leon A: How can I help you? Customer [B]: I'm moving and would like to change my electricity contract to the new apartment A: Sure thing A: Please let me know your contract number B: 100218 A: Perfect, I found your account A: What is your new address? B: Hauptstrasse 12, 31535 Neustadt A: What is the number of your new electricity meter? B: 78822 A: All right, I found the meter with the number 78822 in the system A: On what date are you moving? B: March 15, 2021 A: All right, I've recorded your entry A: Your contract has been reregistered to the meter number 78822	A: Hi, my name is Leon A: How can I help you? B: I'm moving and would like to change my electricity contract to the new apartment A: Sure thing A: Please let me know your contract number B: 100218 A: Perfect, I found your account A: What is your new address? B: Hauptstrasse 12, 31535 Neustadt A: What is the number of your new electricity meter? B: 78822 A: Unfortunately, I cannot find the meter number 78822 in the system A: Please check again what the correct meter number is B: I looked it up, the number is 78822 A: Under the number 78822, I unfortunately cannot find an electricity meter A: Sorry, I cannot help you with this
Information displayed in chatbot disclosure manipulation	Your conversational partner was not a human person, but a chatbot	

Table A2. Full description of scenarios in study 2

Corresponding author
Maik Hammerschmidt can be contacted at: maik.hammerschmidt@wiwi.uni-goettingen.de