“Your conversational partner is a chatbot” — An Experimental Study on the Influence of Chatbot Disclosure and Service Outcome on Trust and Customer Retention in the Fashion Industry

Noemi Funke*, Katja Stadler*, Heidi Vakkuri*, Anna Wagner*, Marc Lunkenheimer*, Alexander H. Kracklauer*

ABSTRACT

Should companies disclose their chatbots’ nonhuman identity or not? Previous studies have found both negative and positive consumer reactions to chatbot disclosure. This experimental study explores how trust and customer retention change when the nonhuman identity of the chatbot is revealed and when different service outcomes apply in the context of the German fashion industry. The results of this experiment provide evidence that disclosing chatbot identity influences neither trust nor customer retention, but service outcome has an effect on both.

Companies should therefore focus on developing a functional customer service as chatbot failure has tremendous consequences for the volume of reliable customers and profits. The main limitation of this study is that the respondents were only shown screenshots, leaving the impact of a real interaction with chatbots undiscovered.

KEYWORDS

Chatbot, customer retention, identity disclosure, service outcome, trust

1. Introduction

“Hello, how can I help you today?” This is a chatbot ready to help customers and facilitate their experience on a website. Chatbots are sometimes difficult to distinguish from human conversational partners [1]. This presents companies with the challenge of considering whether the identity of their chatbots should be revealed to users or not. Studies have been conducted on this topic, with varying

* University of Applied Sciences Neu-Ulm, Wileystrasse 1, 89231 Neu-Ulm, Germany
https://doi.org/10.25929/jair.v1i1.113

results. Some previous studies have been pessimistic about the disclosure of chatbots in general. For example, Luo et al. [2] suggest that the disclosure of artificial agents has negative effects; their data reveal that exposing the identity of a chatbot before a conversation reduces purchase rates by more than 79.7 %. On the other hand, Mozafari et al. [3] present a more positive perspective: When the outcome is not satisfactory and the identity of the chatbot is revealed, the effect of disclosure is positive.

This inconsistency in results merits further investigation. This study aims to replicate the concept of recent publications with a varied research design. The study focuses on the fashion industry, as there is a particular need for automated order processes due to the vast number of order processes in e-commerce and the high returns rate. Chatbots can take over simple service tasks such as changing order details. For this reason, the focus here is not on purchase abandonment rates due to a sales call, as in Luo et. al [2], but on easier-to-complete, more frequent service processes in the context of e-commerce and the associated longer-term customer loyalty. According to Mozafari et al. [3] 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 elicits a positive effect on retention.

Part of the purpose of this study is to answer the following question:

Research Question (RQ) 1: How does disclosing chatbot identity influence customer retention? [3]

Customers nowadays expect a positive service outcome. Belanche et al. [4] point out that customers react differently to chatbot and human service failures. According to Chizhik and Zherebtsova [5] building a full-fledged chatbot that interacts with humans in a human-like manner is a very challenging and complex task. Therefore, Adam et al. [6] report that many users are still not content to interact with chatbots (e.g., because of their high failure rate), which may lead to skepticism and resistance to the technology. This in turn may discourage users from following the recommendations or requests made by a chatbot.

Previous research shows that users' behaviors differ according to whether they think they are interacting with a chatbot or a human being [7]. According to Følstad et al. [8], trust in a chatbot is due not only to the perceived characteristics of the chatbot but also and more particularly to the service context in which the chatbot is situated. It is noteworthy that the brand that provides the chatbot is crucial for trust. In other words, users are more likely to trust chatbots provided by trusted brands. From these observations, the following question emerges:

RQ2: Does service outcome moderate the effect of chatbot disclosure on trust or customer retention? [3]

Customer trust has a major influence on customer retention and the intention to buy [9]. For instance, if a customer feels angry about the service, this is likely to have a negative impact on customer retention (i.e., the customer feels the need to show aversion towards the company by no longer buying the product). This negative impact may extend to other products of the company [10]. The following research question derives from these points:

RQ3: Can the impact of trust on customer retention also be observed in this study?

This article is structured as follows: After a summary of the existing literature in the fields of chatbot disclosure, trust, service outcome, and customer retention, the research framework is laid out, along with the hypotheses. Next, this paper presents an experimental simulation of four kinds of interactions with a chatbot and investigates whether disclosure of the chatbot's identity has a positive or negative impact on customer retention in different frontline service situations. Finally, the findings are summarized, and implications are identified.

2. Theoretical background

Technological progress allows consumers to be better informed, receive targeted offers, and gain faster access to services [11]. New technologies continue to shape the ways shoppers choose channels, products, and services and how they make purchases [11].

With the current technological standard, today's consumers are demanding 24/7 service for assistance. Artificial intelligence (AI) has become a huge part of addressing this need. As a result, companies are rapidly looking to develop text-based automated conversational agents, i.e., chatbots and virtual assistants, to answer questions customers may have at any time of the day [12]. Chatbots are required not only to provide users with necessary consultancy and guidance but also to communicate in a friendly and social way with the aid of the ability to interpret natural (not pre-defined) language used by humans [7, 19, 32]. According to Chizhik and Zherebtsova [5], building a human-like chatbot is a very challenging and complex task. Despite chatbots' performance not always being error-free, chatbots are replacing human chat service agents due to the time and cost savings they offer [1].

During the COVID-19 pandemic, there has been a radical change in retail: The trend toward online shopping has drastically increased [13]. McKinsey & Company [14] report that in 2020, within eight months, e-commerce's share of fashion sales rose from 16 % to 29 % globally, which equals six regular years of growth. According to Luce [12] and Silvestri [13], specialized chatbot services are well used and will become the norm for fashion brands looking for AI-assisted product discovery, product care, and customer service. Today, customers use their smartphones to navigate on e-commerce websites, and chatting has become the most popular method of communication, especially for younger generations.

Chatbot disclosure

Machine learning endows chatbots with interaction skills for impersonating human behavior in order to meet high expectations for good customer service [15]. Sometimes this makes it challenging for users to determine whether they are interacting with a machine or a human when this information is not explicitly provided [16].

Skjuve et al. [17] find that even when the true nature of the conversational agent is unclear to the user, communication is still feasible when the transaction is easy and when expectations about the capabilities of the agent do not lead to inadequate or frustrating service outcomes. In the context of customer service, not revealing the identity of the chatbot can be useful in some situations, especially in situations where human and automated agents overlap seamlessly. Furthermore, Corti and Gillespie [18] point out that finding common ground seems to be more important in interactions with a non-disclosed chatbot with human-like behavior than in interactions with a chatbot whose behavior is not human-like.

Mozafari et al. [3] find that transparently communicating chatbot identity generates positive user reactions. When interactions with chatbots are successful, there is no significant effect on trust. However, in cases of chatbot failure, chatbot disclosure has a significant positive effect on trust. Luo et al. [2], however, suggest the opposite: the disclosure of artificial agents has negative effects due to people's subjective perceptions of machines. Mozafari et al. [3] add that the mere knowledge that users are interacting with a chatbot rather than a human causes a biased reaction. However, if the disclosure of the chatbot identity is combined with selectively presented information about the chatbot, the disclosure dilemma can be solved appropriately. Even with responses that are thoughtful, responsive, and polite, Hendriks et al. [19] find users' perceptions and evaluations of the whole process to be completely changed by chatbot impersonation. If these selected variables, which include humanity and satisfaction, are considered in terms of the overall user experience, their study shows that users still prefer to talk to a real person instead of a chatbot.

Trust in chatbots

Trust is one of the most important factors in building customer loyalty and in strengthening and retaining a relationship [20]. The previously reviewed research shows that users behave differently according to whether they think they are interacting with a chatbot or a human being. Despite promising forecasts, many chatbots built for commercial use underperformed in practice and had to be shut down, which according to industry reports was mainly necessitated by a general lack of trust in chatbots. This lack of trust also prevents widespread adoption of chatbots. This is particularly evident in human–computer interaction [7]. While chatbots mimic human behavior and even replace humans in their tasks, building trust with humans works differently from building trust with chatbots [21]. According to Komiak and Benbasat [22] 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) [3].

The concept of trust is more complex in the context of AI-enabled customer service, where trust is limited not only by technology and brand but also by the purpose and process of using AI [23]. Følstad et al. [8] add that trust in a chatbot is due not only to its perceived characteristics but also, and more notably, to the service context. Building trust takes time; trust can be broken within seconds and needs a long time to be repaired. This is a dynamic process, moving from initial trust to the ongoing development of trust. To achieve the latter, a chatbot should be reliable and easy to use. The fear that AI will replace and displace jobs, along with the potential threat of AI to the existence of humanity, creates distrust and hinders the continued development of trust [23]. The findings of Ameen et al. [24] show, on the one hand, that consumer trust is an important factor to consider in AI adoption and AI experiences. On the other hand, achieving trust presents the greatest of challenges given the complexity and ambiguity of AI technology. Results also confirm that there is a positive relationship between trust and service quality. Thus, concentrating on trust, providers could offer (among other things) better service quality in terms of security, interface design, and reliability.

In cases of negative service outcomes, identifying the conversational partner as a chatbot increases trust and retention [3]. In customer service situations, a chatbot that is human-like enough to recognize a potential misunderstanding seems adequate. The ability to resolve miscommunication appears to be as effective as avoiding it.

Service outcome

When buying a product or service, consumers experience a purchase outcome that leaves them either satisfied or dissatisfied. They then try to determine what led to this result, which influences their future buying behavior. Customers are more likely to search for a responsible entity when the outcome was a failure [10]. The customer attributes responsibility for the negative service outcome to either the product, the employee, or the firm rather than to him- or herself, being convinced that it could have been avoided [10]. As a covariate, however, responsibility attribution has no effect on the relationship between chatbot disclosure, service outcome, and their interaction with trust, according to Mozafari et al. [3]. When experiencing a positive service outcome, customers tend to be less satisfied with the service provider when the conversational partner is a chatbot. A reason for this could be that the customers attribute the positive outcome to themselves and are therefore not surprised [25].

The customer's expectations play a large role in the causal attribution of responsibility for the service outcome. Unexpected outcomes are accompanied by significantly more spontaneous attributions [26]. Attributions to the chatbot of both controllability and service stability are high when negative emotions emerge about the service outcome, meaning that service failure increases customers' reason to complain and, therefore, the spread of negative word of mouth [27]. Dealing with negative service outcomes is linked to anger and helplessness [28]; subsequent explanation of why an error occurred mitigates that anger and helplessness [28].

Belanche et al. [4] show that customers react differently to chatbot and human service failures. The attribution of responsibility to the agent is greater when the agent is a human employee. The agent's responsibility is assigned differently depending on the service outcome. For example, in case of a service failure, little responsibility is attributed to a chatbot, while a human agent is seen as responsible in a similar manner as for a service success. Customers may assume that chatbots are less capable of solving service issues than humans [4]. According to Mozafari et al. [3], this mistrust can be overcome by revealing the identity of the chatbot. This leads, in cases of chatbot failure, to higher levels of trust and retention. Regarding these findings, Blut et al. [1] discuss the merits of anthropomorphism: whether a chatbot should be designed to imitate human behavior or if it should be obvious to the customer when the agent is a chatbot. They conclude that anthropomorphism creates a positive effect because the human-like appearance allows the customer to act as in a human-to-human encounter, using the same social rules.

Customer retention

Even though companies are always striving to improve customer retention and satisfaction, it is ultimately always the customer who makes the decision, for various reasons, whether to stay with or leave the company [29]. For instance, if a customer feels anger with the service, this is likely to have a negative impact on customer retention. The customer feels the need to punish the company by not buying the product anymore, a decision that may extend to other products of the company [10]. Confessing the mistake helps the customer to understand that it was not due to a malicious act on the company's part but only an error, encouraging the customer to believe that the mistake will not be repeated [10].

Silitonga et al. [30] find that on e-commerce websites, the factors with the greatest effect on buyer retention are customization, contact interactivity, care, character, and trust [30]. To gain trust, companies should design their websites in a way that customers would perceive as easy to use and functional. When websites are not designed like this, trust decreases, and customers reduce their participation in online commerce with the companies. Customer trust therefore has a notable influence on customer retention and intention to buy [9]. In a mediation analysis, Mozafari et al. [3] demonstrate that a significant relationship between chatbot disclosure, service outcome, and retention exists only through the mediator of trust. Their research also shows chatbot disclosure to have a positive effect on customer retention when the service outcome is negative and when it is influenced by trust. The effect of chatbot disclosure or service outcome on retention shows no significance.

3. Research framework

For scientific progress, it is important to review research findings by replicating results from a third party. This aims to make the findings more reliable. According to Pesaran [7], there are two types of replication studies, replication in a narrow sense and replication in a wide sense. Replication in a narrow sense means checking the data of the original study for consistency and accuracy. The aim of such a study is to verify the accuracy of the data analysis performed as described in the original study. Replication in a wide sense aims to verify the research method as such. For this purpose, the original data can be used, or new data can be generated and evaluated under changed conditions (e.g., with a change in the date of the data generation, the gender of the participants, or the industries under scrutiny). The aim of such a study is to verify the findings and check whether they apply in general or just under certain conditions. This study belongs to the latter category. It replicates the study by Mozafari et al. [3] in order to verify whether their findings can be transferred to another, sufficiently different industry. If so, these results can be seen as a first hint whether a chatbot's identity disclosure has similar effects in other industries as well.

As the current state of research shows, there is still no definitive answer to the question of whether chatbot identity should be disclosed as studies have observed contrasting results. This leaves the question whether the identity should be communicated honestly still open for discussion [17]. Accordingly, this paper focuses on the effects of service outcome and chatbot disclosure on trust and customer retention in the context of the fashion industry. Trust is one of the most important factors for customer loyalty [20]. Particularly in the case of a negative service outcome, the customer reacts with frustration [28]. Yet, if the service outcome is a failure and chatbot identity is revealed, Belanche et al. [4] say that customers attribute little responsibility to the chatbot, although they would give the chatbot more credit if it were a human being. However, successful service outcomes have no significant effect on trust according to Mozafari et al. [3], whereas when service outcomes are negative, chatbot disclosure has a positive effect on trust. To test the findings of Mozafari et al. [3] for another sector, the first hypothesis transfers this context to the online fashion industry:

Hypothesis (H) 1: If chatbot failure occurs, disclosing (vs. not disclosing) chatbot identity enhances trust in the conversational partner.

As the customer always has the last decision, either staying with or leaving the company, customer retention is an important aspect to consider [29]. A service outcome that triggers the feeling of anger can reduce retention, causing other services to be avoided in the future. For this reason, the second hypothesis addresses whether chatbot identity disclosure increases customer retention in the event of a chatbot failure in the online fashion industry, as follows:

H2: If chatbot failure occurs, chatbot disclosure (vs. non-disclosure) enhances customer retention.

Additionally, this paper considers the relation between trust and customer retention separately. Trust has a major influence on retention and the intention to buy [30]. Since research shows that trust and customer retention are contiguous, this paper tests the effect of trust on retention for the German online fashion industry. Consequently, the third and last hypothesis is formulated as follows:

H3: Trust has a positive effect on customer retention.

4. Research design

A cross-sectional online survey was conducted using a convenience sample, distributed via WhatsApp messages and e-mail as the most efficient ways to reach people. A screening question at the beginning of the survey made sure that the sample included only respondents who had interacted with a customer service feature of a homepage via online chat in the past year. Mozafari et al. [3] did not include such a screening question. However, since it would bias the results if people took part in the survey who had never experienced such customer service, it was necessary to insert a screening question at the beginning of the survey. The initial sample size was n = 191; after the screening question, the sample decreased to n = 148. Attention checks were included in the survey, as they were by Mozafari et al. [3], to make sure that people did not answer the questions randomly. Those who did not pass the attention checks ("Please tick the scale point (5) if you have read the questionnaire carefully" and "What was the service request you approached the company with?") or did not fill out the questionnaire consciously ("Have you answered the questionnaire consciously?") were excluded from further analysis, so the total final sample size was n = 128, with n = 64 male and n = 64 female respondents.

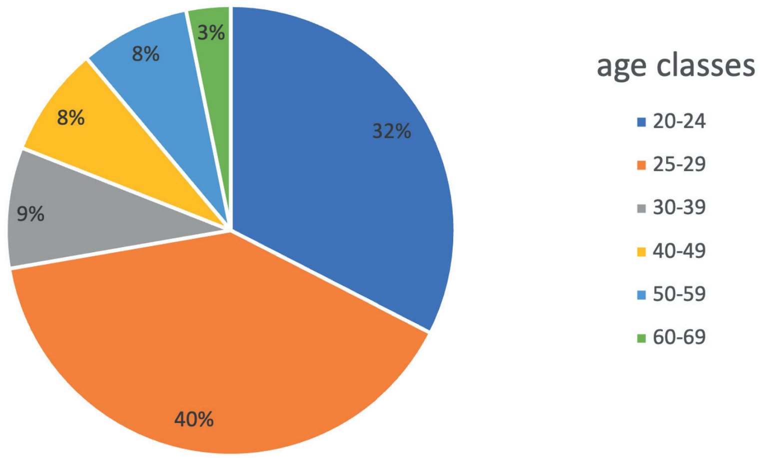

Figure 1: Age distribution in years (n = 128).

Figure 1 illustrates the age distribution of the participants. The average respondent age was 31 years. The procedure including a screening question at the beginning of the survey resulted in a relatively young sample of participants. This phenomenon can be explained by the fact that customers who have experience in dealing with chatbots are predominantly young people. A study conducted in Switzerland found that 53 % of 18−30 year-olds have already had contact with a chatbot. In the 31−50 age group, it is 42 % and of those over 50, it is only 27 % [31]. Thus, the overrepresentation of the younger age group can be explained by this. With regard to gender, equal distribution was taken into account in order to achieve isomorphism here as well.

In contrast to the study by Mozafari et al. [3], in which participants had to imagine that they were customers of an energy provider, the participants in this study had to imagine being customers of a fashion label. The scenario was the following: They, as customers, had just ordered a pair of trousers online but had used the wrong shipping address. Now, they would like to change the address for the package delivery via the customer service's online chat. "Faulty" addresses are a well-known phenomenon in the e-commerce sector - ambiguity, spelling errors, invalid or incomplete addresses are examples for that. Each of the four groups received a different outcome for this scenario. The participants were sent four such screenshots in a row, showing the chronological development of the chat. This ensured that the participants did not already know at the beginning how the chat would proceed (see Figure 2 for a sample scenario). In the first subsample, which included n = 31 participants, the respondents experienced a successful change of the delivery address. The experience of the n = 32 members of the second subsample was a successful change of shipping address and the disclosure that the interaction was with a chatbot and not a human service employee. In the third subsample, the n = 34 respondents tried to change shipping address, which proved unsuccessful. Finally, the experience of the n = 31 members of the fourth subsample was a failed attempt to change the address and chatbot disclosure. The answers were measured with a Likert scale from 1 = strongly disagree to 7 = strongly agree. The conversations were presented in German to the German-speaking participants and translated for the purpose of this publication afterward.

	Chatbot success	**Chatbot failure**
Chatbot non-disclosure	n = 31	n = 34
Chatbot disclosure	n = 32	n = 31

Table 1: The distribution of respondents among groups.

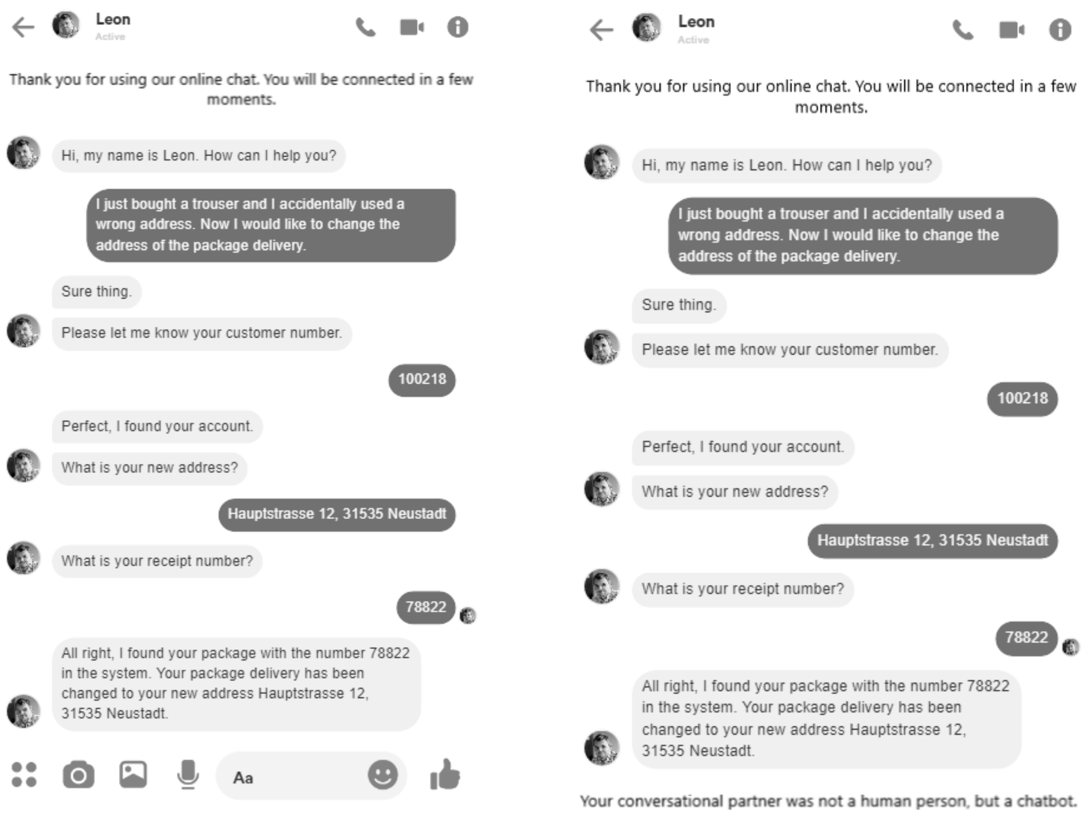

Figure 2: Sample scenario: success with non-/disclosure condition.

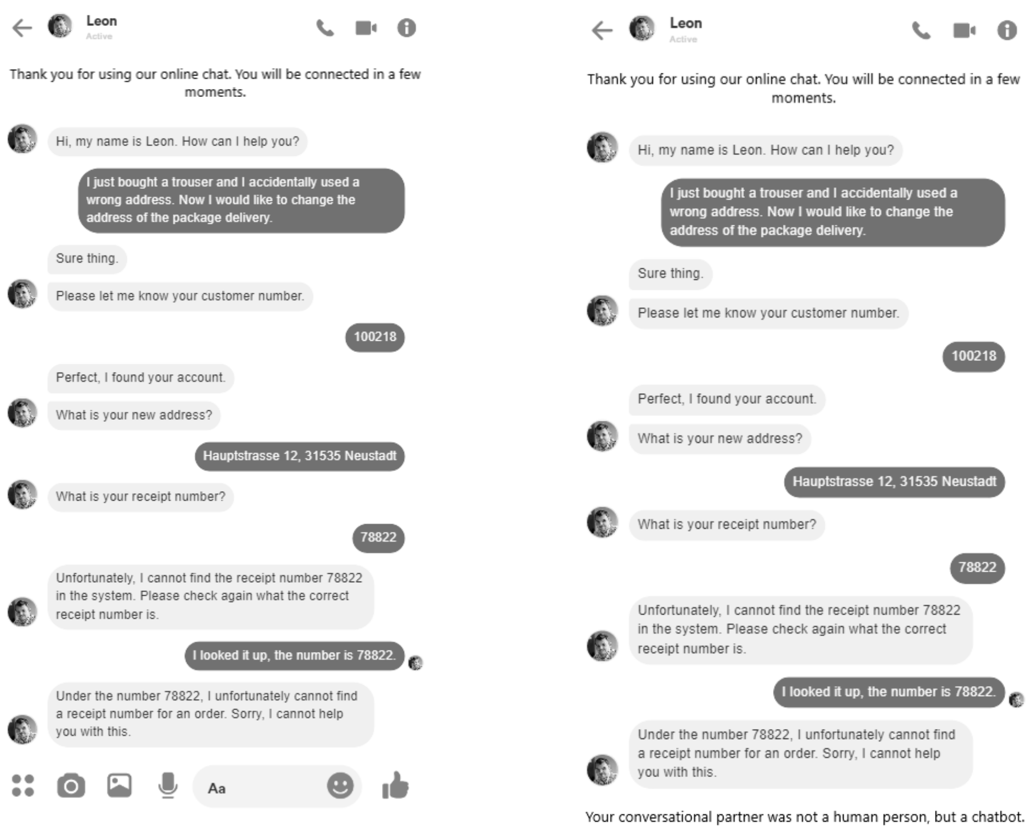

Figure 3: Sample scenario: failure with non-/disclosure condition.

5. Findings

The findings of this study provide new insights into the target phenomenon and contribute to the overall research on trust in relation to service performance and the revelation or non-revelation of the identity of a chatbot. In particular, the results show that regardless of the disclosure or non-disclosure of a chatbot's identity, the success of a chatbot in achieving the desired service outcome leads to higher trust than the chatbot's failure. Furthermore, analysis of the effects of service outcome and chatbot disclosure on customer loyalty shows that only service outcome has an impact. Together, these results can influence the design of corporate chatbot systems in terms of whether and under what circumstances the identity of a chatbot should be revealed.

To test the data against manipulation and for validity, this paper follows the approach of Mozafari et al. [3]. The manipulation check ("Do you think you talked to an automated chatbot or a human service employee?") is statistically significant (see Table 5 in the Appendix). All statistically significant relationships in this paper are significant at the 95 % level of confidence. Compared to respondents who did not know the identity of their conversational partner, respondents experiencing chatbot disclosure were significantly more likely to perceive their conversational partner as a chatbot than as a human service employee ($M_{disclosed}$ = 1.13, SD = 0.34, $M_{undisclosed}$ = 1.45, SD = 0.50, t = 4.22). The validity check ("Which entity was responsible for the service outcome? Me, the customer, or Leon, the employee?") shows whether respondents really perceived customer service failure differently from success. As this relationship is also statistically significant (see Table 6 in the Appendix), it indicates that respondents experiencing service failure attribute the responsibility for the service outcome significantly more to the employee, Leon, than the group that experienced service success ($M_{failure}$ = 6.25, SD = 1.12, $M_{success}$ = 5.33, SD = 1.66, t = −3.67).

To test reliability, a Cronbach's alpha test was conducted, testing the internal consistency of the scales measuring the concepts of trust and customer retention. As Table 2 shows, the Cronbach's alpha values for both concepts are above 0.7, indicating reliability [3]. However, the Cronbach's alpha for the construct trust is above 0.9, which would hint at possible redundant values, meaning that one dimension in the concept is very similar to another dimension. A high correlation is good, but too high a correlation needs to be avoided as it would display multicollinearity. Yet, this is a matter of interpretation, as a broader concept is better for inferences but decreases reliability. Consequently, in this paper this slightly higher value of 0.91 for the construct trust is accepted. The component analysis shows the contribution of each item to the variance in the construct. The higher the value, the more valuable the information contributed by a certain item in explaining the variance within a construct. The question "Would you give the trousers back if you received them?" shows a notable low item loading with a value of 0.46 for the construct customer retention. However, leaving the variable out of the construct increases Cronbach's alpha to 0.93, which is higher than the current value of 0.89. Consequently, the variable is not excluded from the construct.

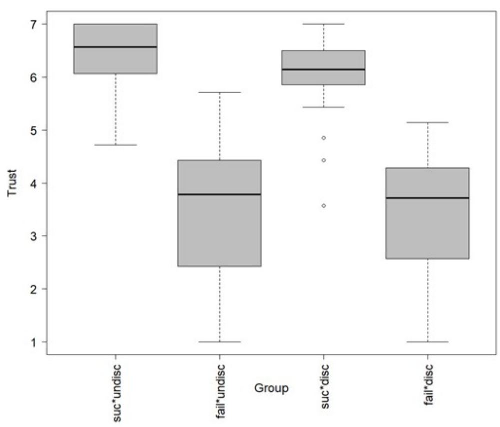

Figure 4: SEQ Figure * ARABIC 4: Boxplot of trust for each group.

Construct	*Dimension*	*Measurement*	*Component analysis*	*α*
Trust in the conversational partner [32]	Competence	Does the conversational partner have the necessary skills to deliver the service?	0.87	0.91
	Competence	Does the conversational partner have access to the information needed to handle my service request adequately?	0.74	
	Integrity	Is the conversational partner's conduct in response to my service request fair?	0.84	
	Integrity	Does the conversational partner have high integrity?	0.84	
	Benevolence	Is the conversational partner receptive to my service request?	0.81	
	Benevolence	Does the conversational partner make efforts to address my service request?	0.82	
	Overall trust	Is the conversational partner trustworthy overall?	0.79	
Customer retention [32]		Would you continue being a customer of this fashion label?	0.88	0.89
		Would you buy additional products beyond this pair of trousers from this fashion label in the future?	0.90	
		If you had to decide, would you select this fashion label again?	0.88	
		Would you return the trousers if you received them? (R)	0.46	
		Would you intend to switch to another fashion label? (R)	0.81	
		Would you plan to abandon this fashion label? (R)	0.88	

Table 2: Measures of multi-item constructs and of dimension and construct reliability.
Notes: R = reverse scaled items; α = Cronbach's alpha. Concepts based on Mozafari et al. [3].

Figure 4 provides a descriptive view of how trust is distributed across the four groups: chatbot success with non-disclosure of chatbot identity, chatbot failure with non-disclosure, chatbot success with disclosure, and chatbot failure with disclosure. One can see that the groups that experienced chatbot success had a higher level of trust than the groups with chatbot failure, consonant with the findings of Mozafari et al. [3]. Mozafari et al. [3] did not find an effect deriving from the covariate "responsibility service outcome" while conducting an ANCOVA analysis. Consequently, this paper follows their approach and conducts an ANOVA instead [3]. The ANOVA analysis of chatbot disclosure, service outcome, and their interactions with trust indicates that only service outcome has a strong positive significant effect, with a p-value of <0.001 (see Table 7 in the Appendix). While this, however, only explains the variance, meaning that the groups differ significantly, a post-hoc test is conducted to further analyze which groups display significant differences. The t-test shows whether the difference among groups is significant by comparing means. It reveals that chatbot disclosure, compared to non-disclosure, had no significant impact on trust when the service outcome was a successful change of address ($M_{success*non-disclosure} = 6.35$, $M_{success*disclosure} = 6.05$, $p = 0.45$). Surprisingly, chatbot disclosure also had no significant effect on trust when the service outcome was a failure ($M_{failure*non-disclosure} = 3.51$, $M_{failure*disclosure} = 3.41$, $p = 0.69$). This result stands in contrast to Mozafari et al. [3]. Consequently, the significance in the ANOVA can only derive from the difference between success and failure as service outcomes. Analyzing these means shows that success had a positive and significant impact on trust compared to failure when the identity of the chatbot was not revealed ($M_{non-disclosure*success} = 6.35$, $M_{non-disclosure*failure} = 3.51$, $p < 0.001$). The same applies to the case of chatbot identity disclosure ($M_{disclosure*success} = 6.05$, $M_{disclosure*failure} = 3.41$, $p < 0.001$). The insignificance of the effect of chatbot disclosure or non-disclosure on trust is shown by comparing the groups experiencing chatbot success and chatbot failure. Independent of chatbot disclosure or non-disclosure, chatbot success as a service outcome meant higher levels of trust compared to chatbot failure (see Table 3; also see Tables 9, 10, and 11 in the Appendix for the individual trust dimensions). Therefore, the first hypothesis needs to be rejected.

	Chatbot success	**Chatbot failure**
Chatbot non-disclosure	M = 6.35, SD = 0.65	M = 3.51, SD = 1.24
Chatbot disclosure	M = 6.05, SD = 0.77	M = 3.41, SD = 1.13

Table 3: Mean and standard deviation for the concept trust.

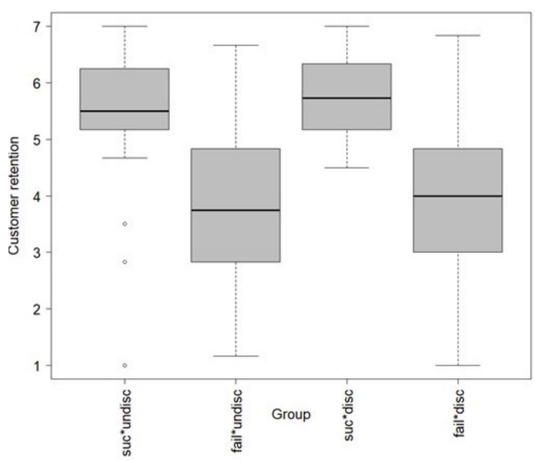

Figure 5: SEQ Figure * ARABIC 5: Boxplot of customer retention for each group.

Figure 4 provides a descriptive view of how customer retention is distributed across the four groups. One can see that the groups experiencing chatbot success had higher levels of customer retention than the groups with chatbot failure. The ANOVA analysis of chatbot disclosure, service outcome, and their interactions with customer retention indicates that only service outcome had a strong positive significant effect, with a p-value of <0.001 (see Table 8 in the Appendix). The post-hoc test shows that chatbot disclosure, compared to non-disclosure, had no significant impact on customer retention when the service outcome was a successful change of address ($M_{success*non\text{-}disclosure} = 5.46$, $M_{success*disclosure} = 5.77$, $p = 0.63$). Chatbot disclosure also had no significant effect on retention when the service outcome was failure ($M_{failure*non\text{-}disclosure} = 3.80$, $M_{failure*disclosure} = 3.88$, $p = 0.78$). Consequently, the significance in the ANOVA can only derive from the difference between success and failure as service outcomes. Analyzing these means shows that success had a positive and significant impact on customer retention compared to failure when the identity of the chatbot was not revealed ($M_{non\text{-}disclosure*success} = 5.46$, $M_{non\text{-}disclosure*failure} = 3.80$, $p < 0.001$). The same applies to the case of chatbot identity disclosure ($M_{disclosure*success} = 5.77$, $M_{disclosure*failure} = 5.88$, $p = <0.001$). The insignificance of the effect of chatbot disclosure or non-disclosure on customer retention is shown by comparing the groups that experienced chatbot success with the groups that experienced chatbot failure (see Table 4). Independent of chatbot disclosure or non-disclosure, chatbot success as a service outcome was shown to mean higher levels of customer retention compared to chatbot failure. Therefore, the second hypothesis also needs to be rejected.

	Chatbot success	**Chatbot failure**
Chatbot non-disclosure	M = 5.46, SD = 1.24	M = 3.80, SD = 1.30
Chatbot disclosure	M = 5.77, SD = 0.74	M = 3.88, SD = 1.37

Table 4: Mean and standard deviation for the concept customer retention.

To analyze Hypothesis 3, a regression analysis was conducted (see Table 5). Sorting out the missing data points left 128 observations. The R2 shows that 51.6 % of the variation in the data can be explained with this model, which is a mediocre prediction, leading to the conclusion that other factors also influence the relationship. The F statistic of 138.408 shows that the overall model is significant as the independent variable has a highly significant positive effect on the dependent variable, with a t-value of 11.593. Hence, trust influences customer retention positively. Increasing trust by 1 increases customer retention by 0.63. The relationship is statistically significant. Consequently, as this supports the assumption, the third hypothesis cannot be rejected.

	Dependent variable:
	retention
trust	0.630
	$t = 11.593^{***}$
Constant	1.682
	$t = 6.071^{***}$
Observations	128
R^2	0.516
Adjusted R^2	0.512
Residual Std. Error	1.033 (df = 126)
F Statistic	134.408^{***} (df = 1; 126)
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Figure 6: SEQ Figure * ARABIC 6: Regression table of the impact of trust on customer retention (n = 128).

Figure 6 displays the descriptive relationship between trust and customer retention. It shows the observation points and the regression as well as the confidence intervals. It clearly reveals that trust and customer retention are highly correlated with each other.

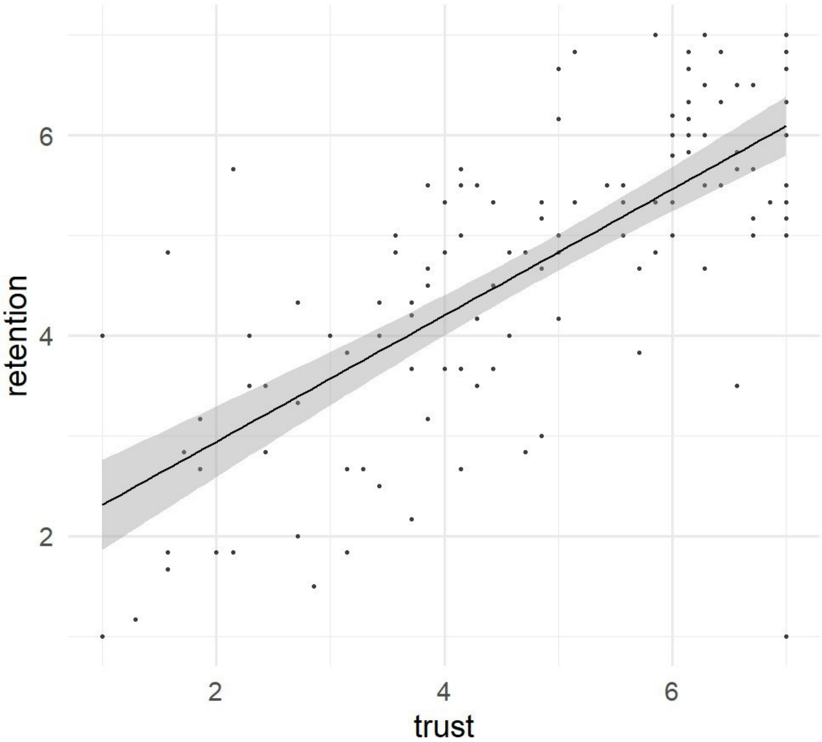

Figure 7: SEQ Figure * ARABIC 7: Regression of trust on customer retention with confidence intervals (n = 128).

6. Discussion

In summary, this paper supports the potential cross-sectoral validity of some of the findings of Mozafari et al. [3] but cannot confirm others for the online fashion industry.

Like Mozafari et al. [3], this paper shows that experiencing a positive service outcome leads to higher levels of trust compared to a negative service outcome. In general, chatbot disclosure does not have an impact on trust, neither in this paper nor in the study by Mozafari et al. [3]. Differentiating between the two service outcomes also leads to no effect of chatbot disclosure on trust when the service outcome is a success. This applies for Mozafari et al. [3] as well. However, in the present study, chatbot disclosure in cases of failure as a service outcome still does not influence trust. This result stands in contrast to Mozafari et al. [3]. To sum up, while chatbot disclosure becomes important in the case of chatbot failure in the European energy sector, for online fashion industry, only service outcome has an impact on trust, leaving disclosure and non-disclosure out of the equation.

In analyzing whether customer retention is influenced by service outcome or chatbot disclosure, an effect is found only for service outcome. Looking at the groups experiencing chatbot success and failure separately shows no impact of chatbot disclosure on customer retention in conjunction with either of the service outcomes. This leads to the conclusion that service outcome is the important factor and not chatbot disclosure, which is generally in line with current research [10, 3]. Nonetheless, this paper also stands in contrast to new research findings as it finds, unlike Mozafari et al. [3], that disclosure is also not important in cases of chatbot failure. While looking at the relationship between trust and customer retention, which was already discovered in previous research [9], this paper reports evidence that trust does also influence customer retention in the online fashion industry.

Although the results of this paper contribute to the current state of literature by challenging a few recent findings, there are some limitations of this study that need to be addressed. As the respondents were quite young, findings might differ for an older population less familiar with digital technology; in that case, chatbot disclosure in combination with service outcome might have a greater effect. Additionally, a comparison of male and female respondents might be illuminating. It should be noted that as with Mozafari et al. [3], respondents were shown only screenshots, leaving the impact of a real interaction with chatbots undiscovered. In addition, testing other constructs of trust and customer retention would be conceivable. Beyond this, it would be interesting to see whether the same results apply in another industry or country, since the scope of this study is on the German sector only. It would be also interesting to replicate the study of Luo et al. [2] concerning negative effects of chatbot disclosure and purchase rates in the fashion sector; however, this would need another research design. From a quantitative point of view, looking at the regression model again shows that it can only predict about 50 % of the variance, leading to the conclusion that other factors influence the relationship as well. Future research should address these issues.

7. Conclusion and implications

Finally, to answer the original research questions, this paper shows that disclosing chatbot identity influences neither trust nor customer retention. It reveals that service outcome not only moderates the effect of chatbot disclosure on trust and customer retention — it is the sole factor having an impact. Lastly, this paper also extends the widely discussed influence of trust on customer retention to the online fashion industry.

Contributing to existing research are some implications deriving from this research. Consumers are generally quite critical when it comes to new technology. However, this study shows that service outcome is more important for trust in a brand and customer retention than is revealing chatbot identity. Brands offering services via chatbots should keep the critical attitude of consumers in mind. However, they should focus on developing functional customer service as chatbot failure has tremendous consequences for levels of trust and customer retention, which impact the brand directly through decreased volumes of reliable customers and profits.

Conflict of interest statement

The authors declare that there is no conflict of interest in connection with the present work.

References and notes

[1] Blut, M., Wang, C., Wünderlich, N. V.; Brock, C. (2021): Understanding anthropomor-phism in service provision: a meta-analysis of physical robots, chatbots, and other AI. In: Journal of the Academy of Marketing Science, pp. 632–658.

[2] Luo, X., Tong, S., Fang, Z.; Qu, Z. (2019): Frontiers: Machines vs. Humans: The Impact of Artificial Intelligence Chatbot Disclosure on Customer Purchases. In: Marketing Science, vol. 38 (6), pp. 937-947.

[3] Mozafari, N., Weiger, W. H.; Hammerschmidt, M. (2021b): Trust me, I'm a bot – repercussions of chatbot disclosure in different service frontline settings. In: Journal of Service Management, vol. 33 (2), pp. 221-245.

[4] Belanche, D., Casaló, L. V., Flavián, C.; Schepers, J. (2020): Robots or frontline em-ployees? Exploring customers' attributions of responsibility and stability after service failure or success. In: Journal of Service Management, pp. 267–289.

[5] Chizhik, A.; Zherebtsova, Y. (2020): Challenges of Building an Intelligent Chatbot, Proceedings of the International Conference" Internet and Modern Society"(IMS-2020).

[6] Adam, M., Wessel, M.;Benlian, A. (2021): AI-based chatbots in customer service and their effects on user compliance, in: Electronic Markets, vol. 31, (2), pp. 427–445.

[7] Pesaran, H. (2003): Introducing a replication section. In: Journal of Applied Econometrics (1), pp. 111.

[8] Følstad, A., Nordheim, C. B.; Bjørkli, C. A. (2019): What Makes Users Trust a Chatbot for Customer Service? An Exploratory Interview Study. In: Internet Science: 5th International Conference, vol. 2019, pp. 194–208.

[9] Ranaweera, C.; Prabhu, J. (2003): On the relative importance of customer satisfaction and trust as determinants of customer retention and positive word of mouth. In: Journal of Targeting, Measurement and Analysis for Marketing, vol. 12, pp. 82-90.

[10] Weiner, B. (2000): Attributional Thoughts about Customer Behaviour. In: Journal of Consumer Research, vol. 27 (3), pp. 382-387.

[11] Grewal, D., Roggeveen, A. L.; Nordfält, J. (2017): The Future of Retailing. In: Journal of Retailing (1), p. 1–6.

[12] Luce, L. (2019): Artificial Intelligence for Fashion. How AI is Revolutionizing the Fashion Industry, New York: Apress.

[13] Silvestri, B. (2020): The Future of Fashion: How the Quest for Digitization and the Use of Artificial Intelligence and Extended Reality Will Reshape the Fashion Industry After COVID-19. In: ZoneModa Journal, 10 (2), pp. 61-73.

[14] McKinsey & Company (2020): The State of Fashion 2021. Avialable at: https://www.mckinsey.com/~/media/mckinsey/industries/retail/our%20insights/state%20of%20fashion/2021/the-state-of-fashion-2021-vf.pdf [Accessed: 23 August 2022].

[15] Hyken, S. (2017): AI And Chatbots Are Transforming The Customer Experience [Online]. Available at https://www.forbes.com/sites/shephyken/2017/07/15/ai-and-chatbots-are-transforming-the-customer-experience/?sh=71bbe5ad41f7 [Accessed:19 December 2021].

[16] Candello, H., Pinhanez, C.; Figueiredo, F. (2017): Typefaces and the Perception of Hu-manness in Natural Language Chatbots, CHI '17: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 3476–3487.

[17] Skjuve, M., Haugstveit, I. M., Følstad, A.; Brandtzaeg, P. B. (2019): Help! Is my chatbot falling into the uncanny valley? An empirical study of user experience in human-chatbot interaction. In: Human Technology, pp. 30–54.

[18] Corti, K.; Gillespie, A. (2016): Co-constructing intersubjectivity with artificial conversational agents: People are more likely to initiate repairs of misunderstandings with agents represented as human. In: Computers in Human Behavior, vol. 58, pp. 431–442.

[19] Hendriks, F., Ou, C. X.J., Khodabandeh Amiri, A.; Bockting, S. (2020): The Power of Computer-Mediated Communication Theories in Explaining the Effect of Chatbot Introduction on User Experience, Proceedings of the 53 Hawaii International Conference on System Sciences.

[20] Rai, A. K.; Srivastava, M. (2014): Customer Loyalty: Concept, Context and Character. New York: McGraw Hill.

[21] Chen, J. V., Thi Le, H.; Tran, S. T. T. (2021): Understanding automated conversational agent as a decision aid: matching agent's conversation with customer's shopping task. In: Internet Research (4), pp. 1376–1404.

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

[23] Siau, K.; Wang, W. (2018): Building Trust in Artificial Intelligence, Machine Learning, and Robotics. In: Cutter Business Technology Journal, vol. 31 (2), pp. 47-53.

[24] Ameen, N., Tarhini, A., Reppel, A.; Anand, A. (2021): Customer experiences in the age of artificial intelligence. In: Computers in human behavior, vol. 114.

[25] Jin, R. (2010): Causal Attributions on Service Outcomes by Self-Service Technology Users, Purdue University.

[26] Kanazawa, S. (1992): Outcome or Expectancy? Antecedent of Spontaneous Causal Attribution. In: Personality and Social Psychology Bulletin, vol. 18 (6), pp. 659–668.

[27] van Vaerenbergh, Y., Orsingher, C., Vermeir, I.; Larivière, B. (2014): A Meta-Analysis of Relationships Linking Service Failure Attributions to Customer Outcomes. In: Journal of Service Research, pp. 381–398.

[28] Gelbrich, K. (2010): Anger, frustration, and helplessness after service failure: coping strategies and effective informational support. In: Journal of the Academy of Marketing Science, pp. 567–585.

[29] Kumar, V., Bhagwat, Y.; Zhang, X. (2015): Regaining "Lost" Customers: The Predictive Power of First-Lifetime Behavior, the Reason for Defection, and the Nature of the Win-Back Offer. In: Journal of Marketing, pp. 34–55.

[30] Silitonga, K. A. A., Ikhsan, R. B.; Fakhrorazi, A. (2020): Drivers of buyer retention in e-commerce: The role of transaction characteristics and trust. In: Management Science Letters, pp. 3485–3494.

[31] Pidas AG (2018): Chatbot-Studie – Die digitalen Helfer im Praxistest. Available at: https://cdn2.hubspot.net/hubfs/5893787/PIDAS_ZHAW_Chatbot_Studie_FINAL.pdf

[32] Bhattacherjee, B. (2002): Individual Trust in Online Firms: Scale Development and Initial Test. In: Journal of Management Information Systems, pp. 211–241.

Appendix

Test-statistic	Df	p-value	Alternative hypothesis	Mean in Group 1	Mean in Group 2
4.221	126	4.613e-05 ***	Two-sided	1.446	1.127

Table 5: Manipulation check.

Test-statistic	Df	p-value	Alternative hypothesis	Mean in Group 1	Mean in Group 2
−3.665	126	0.0003631 ***	Two-sided	5.333	6.246

Table 6: Validity check.

	Df	Sum Sq	Mean Sq	F-value	Pr (> F)
Service outcome	1	239.46	239.46	246.933	< 2e−16 ***
Chatbot disclosure	1	1.25	1.25	1.293	0.258
Service outcome* disclosure	1	0.33	0.33	0.341	0.560
Residuals	124	120.25	0.97		

Table 7: ANOVA analysis of service outcome, chatbot disclosure, and their interaction with trust (significance codes: 0 '***', 0.001 '**', 0.1 '*', 0.05 '.', 0.1 ' ', 1).

	Df	Sum Sq	Mean Sq	F-value	Pr (> F)
Service outcome	1	101.07	101.07	71.518	6.39e−14 ***
Chatbot disclosure	1	1.18	1.18	0.833	0.363
Service outcome* disclosure	1	0.38	0.38	0.272	0.603
Residuals	124	175.24	1.41		

Table 8: ANOVA analysis of service outcome, chatbot disclosure, and their interaction with customer retention (significance codes: 0 '***', 0.001 '**', 0.1 '*', 0.05 '.', 0.1 ' ', 1).

	Chatbot success	Chatbot failure
Chatbot non-disclosure	M = 6.61, SD = 0.66	M = 3.56, SD = 1.89
Chatbot disclosure	M = 6.45, SD = 0.73	M = 3.18, SD = 1.53

Table 9: Mean and standard deviation for the trust dimension competence.

	Chatbot success	Chatbot failure
Chatbot non-disclosure	M = 6.13, SD = 0.93	M = 3.56, SD = 1.44
Chatbot disclosure	M = 5.83, SD = 1.04	M = 3.66, SD = 1.49

Table 10: Mean and standard deviation for the trust dimension integrity.

	Chatbot success	Chatbot failure
Chatbot non-disclosure	M = 6.65, SD = 0.62	M = 3.21, SD = 1.67
Chatbot disclosure	M = 6.23, SD = 1.00	M = 3.32, SD = 1.54

Table 11: Mean and standard deviation for the trust dimension benevolence.