Conversational repair strategies to cope with errors and breakdowns in conversations with customer service chatbots

Jasmin Joëlle Verhagen

SNR: 2091138, ANR: 481314

Master's Thesis

Communication and Information Sciences

Specialization New Media Design

Department Communication and Cognition

School of Humanities and Digital Sciences

Tilburg University, Tilburg

Supervisors: A. Braggaar and Dr. C. Liebrecht

Second reader: Dr. E. van Miltenburg

July 2023

Table of Contents

Abstract **4**

Theoretical Framework **7**

Errors *7*

Breakdowns and Conversational Repair Strategies *10*

User Satisfaction, Brand Attitude, and Trust *12*

Subquestions and Hypotheses *14*

Study I – Corpus Study **14**

Method **14**

Design *14*

Dataset Description *14*

Procedure *15*

Preliminary Results **15**

Errors *16*

Repair Strategies *16*

Preliminary Conclusion **18**

Hypotheses Study II *18*

Study II - Online Experiment **20**

Method **20**

Design *20*

Participants *20*

Stimuli *21*

Measurements *24*

Procedure *26*

Results 27

User Satisfaction 28

Brand Attitude 29

Trust 30

Ranking and thematic analysis 34

Conclusion and Discussion 36

Conclusion 36

Discussion 37

References 42

Appendix A 49

Stimulus Experiment 49

Appendix B 54

Overview Cronbach's Alpha for dependent variables 54

Appendix C 56

Survey Online Experiment 56

Appendix D 63

Outliers and assumption of normality for dependent variables 63

Appendix E 65

Thematic Analysis of preferred repair strategies 65

Abstract

The current study aims to (1) investigate what errors and conversational repair strategies appear during conversations with customer service chatbots and (2) how people perceive these errors and repair strategies in terms of user satisfaction, brand attitude, and trust. The current study consists of a corpus study (i.e., study I) and an online experiment (i.e., study II). The corpus study analyzed a dataset containing real conversations (N = 100) between customers and a customer service chatbot of a Dutch railroad company to examine what errors and conversational repair strategies appeared. The results of the corpus study revealed that the following six errors appeared: (1) Lack of information, (2) Ill-formed, (3) Excess of information, (4) Ignoring the question, (5) Non-understandable, and (6) Non-cooperativity, and the following five types of repair strategies appeared: (1) Apologetic behavior, (2) Repeat, (3) Top response, (4) Defer, and (5) Options. Regarding errors, the error type Excess of information occurred most frequently. In terms of repair strategies, Defer occurred most frequently. Subsequently, the results of study I informed the experimental materials for study II. The online experiment (N = 150) employed a 3 (error) x 3 (repair strategy) mixed design with the type of error as a between-subjects factor and repair strategy as a within-subjects factor. The results of the online experiment revealed statistically significant main effects of brand attitude and trust for the different repair strategies. The repair strategy Options scored significantly higher on brand attitude and trust than the repair strategy Repeat, and the repair strategy Defer scored significantly higher on brand attitude and trust than the repair strategies Options and Repeat. Unexpectedly, there was no statistically significant main effect of user satisfaction for the different repair strategies, nor a statistically significant interaction effect between error type and repair strategy on user satisfaction, brand attitude, or trust.

Keywords: customer service chatbots, errors, conversational breakdowns, repair strategies, user satisfaction, brand attitude, trust

Conversational repair strategies to cope with errors and breakdowns in conversations with customer service chatbots

In recent years, chatbots rapidly gained popularity as they can serve as alternatives to live customer support (Li et al., 2020). Chatbots are artificial agents emulating human-like interactions allowing users to access information and services through natural language dialogue (Asisof, 2022; Følstad et al., 2020). Implementing chatbots for customer service purposes allows organizations to enable faster, more efficient, and 24/7 service provision, as is expected by today's customers (Følstad & Skjuve, 2019; Gnewuch et al., 2017; Kvale et al., 2019). Moreover, customer service chatbots are time and cost-saving for organizations since fewer employees are required (Chung et al., 2020; Gnewuch et al., 2017). Examples of well-known customer service chatbots that Dutch organizations have implemented are Billie from Bol.com and Pennenveer from the amusement park Efteling. Considering the benefits for organizations, it is not remarkable that the "global chatbot market size was estimated at USD 641.1 million in 2022 and is expected to expand at a compound annual growth rate of 25.7% from 2022 to 2030" (Grand View Research, 2023, Report Summary).

Although chatbot technologies have evolved, customer service chatbots have yet to learn to operate error-free (Ashktorab et al., 2019; Toader et al., 2019). Chatbot users frequently experience conversational breakdowns during human-chatbot conversations since chatbots regularly fail to interpret (complex) requests correctly or understand the intended meaning (Følstad et al., 2018; Kvale et al., 2019; Li et al., 2020; Matsumoto et al., 2022; Sheehan et al., 2020). These conversational breakdowns emerge since chatbots cannot yet avoid errors in natural language understanding resulting in the chatbot responding with an answer that does not match the customer's request or with a message that he does not understand the request (Asisof, 2022; Følstad & Taylor, 2019). Conversational breakdowns can cause users frustration, annoyance, and confusion, leading to less trust in chatbots

(Benner et al., 2021; Følstad & Taylor, 2019; Toader et al., 2019) and reduced user satisfaction and willingness to continue using a chatbot (Ashktorab et al., 2019).

However, chatbots have an opportunity to restore the conversation and overcome the breakdown by using appropriate conversational repair strategies (Ashktorab et al., 2019). Repair is a well-known phenomenon in human-human conversations where individuals take action to regain communication after a breakdown occurs, and it can be described as "replacing an error or mistake with what is correct" (Ashktorab et al., 2019; Schegloff et al., 1977). These actions (e.g., clarifying, repeating) are performed subconsciously, rely on a high degree of mutual understanding, and are often a combination of facial expressions, gestures, and speech (Ashktorab et al., 2019). Contrary to repair in human-human conversations, repair strategies in human-chatbot conversations require more explicit communication and guidance as the chatbot depends on algorithms to process the users' input, which is considered a 'black box' for the users (Ashktorab et al., 2019; Brinton et al., 1986).

Although users' preferences for conversational repair strategies are investigated in various studies (Ashktorab et al., 2019; Benner et al., 2021; Reinkemeier & Gnewuch, 2022), interestingly, few to no studies investigated context-dependent repair strategies. In other words, whether the context in which the conversational breakdown occurred (e.g., the type of error) affects which repair strategy is most appropriate. Therefore, further research is needed to investigate context-dependent repair strategies. Moreover, further research is required to investigate when a handover to a human service employee (i.e., Defer) is needed and whether this affects user satisfaction, brand attitude, and trust (Benner et al., 2021). This is also true for other conversational repair strategies since little research has been done on which repair strategies can mitigate trust loss and dissatisfaction (Seeger & Heinzl, 2021).

The Present Study

This research consists of two parts: study I and study II. In Study I, a corpus study will identify which errors and conversational repair strategies occur in conversations with a real customer service chatbot. For this purpose, the following research question was formulated: RQ1: *What errors and conversational repair strategies occur during conversations with a customer service chatbot?* In Study II, an online experiment will explore how people perceive these errors and conversational repair strategies in terms of user satisfaction, brand attitude, and trust. For this purpose, the following research question was formulated: RQ2: *To what extent do the type of error and conversational repair strategy used by a customer service chatbot affect user satisfaction, brand attitude, and trust?*

The present study aims to understand people's preferences for conversational repair strategies and whether that depends on the type of error. Contrary to most studies, the present study focuses on context-dependent repair strategies that recover the breakdown and ensure users can accomplish the task. The present study also aims to fill the knowledge gaps by discovering insights to strengthen conversational repair strategies for customer service chatbots to minimize dissatisfaction and distrust (Benner et al., 2021). Understanding which conversational repair strategies are most preferred and suitable for specific errors or contexts is crucial to enhance user satisfaction, brand attitude, and trust. The above adds to the existing literature on errors and miscommunication in conversations with customer service chatbots and allows organizations to improve their chatbot's conversational repair strategies.

Theoretical Framework

Errors

During human-chatbot conversations, it happens that the system (i.e., chatbot) fails to understand the user or that the user fails to understand the system (Cuadra et al., 2021). These events are known as errors and can be defined as follows: "any event that might have a negative impact on the flow of the interaction, and more in general on its quality, potentially resulting in conversation breakdowns […] which may also result in misinterpretations of any kind" (Sanguinetti et al., 2020, p. 151). The study by Sanguinetti et al. (2020) developed an annotation scheme describing the characteristics and definitions of errors that occurred during conversations with a chatbot of an Italian telecommunication service provider. The scheme is inspired by the Gricean maxims, as the maxim of Quantity, Relation, and Manner represent the error classes (see Table 1). However, it does not include the fourth maxim (i.e., Quality). The Gricean maxim is a set of norms that is part of the Cooperative Principle that explains how effective communication can be achieved (Grice, 1989).

The first error class (i.e., Quantity) represents that the message needs to address the intent but does not need to be more informative than required (Sanguinetti et al., 2020; Setlur & Tory, 2022). The second error class (i.e., Relation) represents that the message needs to be relevant (Setlur & Tory, 2022), while the third error class (i.e., Manner) illustrates that the message needs to be concise and clear (Setlur & Tory, 2022). According to Grice's definition of this maxim, it is not to focus on what is said but on how it is said (Grice, 1989). Thus, if the message violates this maxim, it is included in this error class (Sanguinetti et al., 2020).

The study by Sanguinetti et al. (2020) found that the error tags Ignoring question (20.94%) and Ill-formed (13.09%) most frequently occurred among customer messages in conversation with the chatbot of the Italian telecommunication service provider. In chatbot messages, the error tag Topic change' (30.37%) and Repetition (23.70%) most frequently occurred. Surprisingly, little is known about which errors appear in Dutch human-chatbot interactions or other service contexts. Hence, Study I will examine which errors occur during Dutch human-chatbot conversations in another service context (i.e., public transportation).

Table 1

Error tags according to their corresponding classes

Error class	Error tag	Description	Customer	Chatbot
Quantity	Lack of information	The message contained an insufficient amount of information.	✓	✓
	Excess of information	The message exceeds the maximum number of characters by containing excessive and redundant information.	✓	✓
Relation	Ignoring question/feedback	The customer/chatbot did not consider any information in earlier messages.	✓	✓
	Repetition	The message contains the same information as in earlier messages.	✓	✓
	Straight wrong response	The message is irrelevant to what was discussed earlier in the conversation.	✓	✓
	Topic change	The chatbot proposed to switch to another topic.		✓
	Answering with question	The customer replies with a question to a request by the chatbot.	✓	
Manner	Ill-formed	The message does not correspond to the chatbot's expected form. This usually applies to cases where the customer does not select the offered options.	✓	
	Indirect question	The message contains an implicit request/question.	✓	
	Non-understandable	The customer's message is obscure and incomprehensible, preventing the chatbot from understanding the request.	✓	
	Grammatical error	The message contains a grammatical error.[1]	✓	✓
Generic	Non-cooperativity	The customer is not cooperating by refusing to provide the chatbot with information.	✓	
	Other	Errors that did not fall into one of the other categories.	✓	✓

Note. Adapted from Sanguinetti et al. (2020). The two columns on the right show a check mark if the error tags belong to messages of the customer, chatbot, or both.

[1] Although chatbots use a set of manually created responses that is nearly always grammatically correct, the chatbots' messages sometimes still contain grammatical errors (Sanguinetti et al., 2020).

Breakdowns and Conversational Repair Strategies

Errors may result in conversational breakdowns or misinterpretations (Sanguinetti et al., 2020). When a conversational breakdown occurs, it negatively affects users since it causes frustration, annoyance, and confusion, which in turn leads to reduced satisfaction and trust in chatbots (Ashktorab et al., 2019; Benner et al., 2021; Følstad & Taylor, 2019). However, chatbots can use conversational repair strategies to restore the conversation, overcome the conversational breakdown and mitigate the negative effects (Ashktorab et al., 2019). Several studies have highlighted the importance of using appropriate repair strategies since not every strategy gives the desired result (Ashktorab et al., 2019; Reinkemeier & Gnewuch, 2022).

The study by Benner et al. (2021) described six categories for repair strategies: (1) Confirmation, (2) Information, (3) Disclosure, (4) Social, (5) Solve, and (6) Ask. The first category (i.e., Confirmation) is an attempt by the chatbot to continue the conversation without truly recovering the breakdown. The chatbot rejects the error by responding, '*I do not know*' or '*I do not understand.*' or ignoring the error, assuming the customer attempts to recover the conversational breakdown (Benner et al., 2021). The second strategy (i.e., Information) concerns the chatbot giving a help message to ensure the conversation can continue if the chatbot fails to understand the request. The third category (i.e., Disclosure) is closely related to the second category (i.e., Information). However, the customer service chatbot does not address the conversational breakdown in this third category and only gives help messages. The fourth category (i.e., Social) evolves around human behavior. The customer service chatbot shows empathy and apologizes for the conversational breakdown. The fifth category (i.e., Solve) is mostly goal-oriented since the customer service chatbot tries to provide alternatives, for example, a list of options (Benner et al., 2021). Lastly, the sixth category (i.e., Ask) concerns the chatbot asking the customer to (1) rephrase the question, (2) repeat the question, or (3) input to provide the right answer (Benner et al., 2021).

The study by Ashktorab et al. (2019) also provided valuable insights regarding repair strategy preferences for conversational breakdowns in human-chatbot interactions. Eight conversational repair strategies were presented: (1) Out-of-vocabulary explanation, (2) Keywords confirmation explanation, (3) Keyword highlight explanation, (4) Defer, (5) Repeat, (6) Confirmation, (7) Options, and (8) Top. The results showed that people preferred conversational repair strategies in which the breakdown was explicitly acknowledged (e.g., *Sorry, I did not understand*) and when the right action was suggested. Furthermore, chatbots frequently use handovers to human customer service agents (i.e., Defer). An example of what the chatbot might then say is: '*Sorry, I did not understand. Would you like me to redirect you to a colleague?'* (Ashktorab et al., 2019). This repair strategy partially overlaps with the category 'Confirmation' and 'Social' in the study of Benner et al. (2021) since the chatbot explicitly indicates a failure and apologizes. One of the strengths of this repair strategy is that human interactions are faster and more likely to solve the issue than chatbots (Ashktorab et al., 2019). On the other hand, the human customer service agent's availability must be considered since redirecting may also cause an increased waiting time for the customer. Moreover, the intervention of a human customer service agent is not always necessary (Ashktorab et al., 2019). This raises the question of whether this affects user satisfaction.

Another instance of a repair strategy is Options, a repair strategy with evidence of a breakdown. In that case, the customer service chatbot could suggest several options for potential intents (Ashktorab et al., 2019). An example of what the customer service chatbot might say is: '*Sorry, I did not understand. Can you use the buttons below to indicate what I can help you with?'*. The user then clicks via the buttons the chatbot provided what he initially meant. This repair strategy partially overlaps with the overarching categories 'Confirmation,' 'Solve,' and 'Social' in the study of Benner et al. (2021) since the chatbot explicitly indicates a failure, apologizes, and provides options. This strategy has the advantage that the conversational breakdown can be resolved fast (Ashktorab et al., 2019). A third instance of a repair strategy is ‘Repeat.’ This repair strategy partially overlaps with the overarching categories ‘Confirmation’ and ‘Ask’ in the study of Benner et al. (2021) since the chatbot explicitly indicates a failure and asks the customer to rephrase the question.

The existing literature on miscommunication and conversational breakdowns in chatbot conversations primarily focused on the effects of breakdowns, distinguishing the different repair strategies, and investigating users’ preferences for repair strategies (Ashktorab et al., 2019; Benner et al., 2021; Reinkemeier & Gnewuch, 2022). For example, research demonstrated that repair strategies explicitly acknowledging breakdowns and suggesting the right action are users' most preferred conversational repair strategies (Ashktorab et al., 2019; Reinkemeier & Gnewuch, 2022). People also liked the repair strategy Options (i.e., the chatbot lists options for potential intents) since it was considered efficient and effortless. The repair strategy Defer (i.e., redirecting to a human employee) was only preferred with unsuccessful outcomes (Ashktorab et al., 2019. However, little is known regarding the effects of these repair strategies on user satisfaction, brand attitude, and trust.

User Satisfaction, Brand Attitude, and Trust

User satisfaction is commonly used to evaluate chatbots (Maroengsit et al., 2019). It is, however, a complex concept that is difficult to define (Griffiths et al., 2007; Lindgaard & Dudek, 2003). User satisfaction is “the sum of one’s feelings or attitudes toward various factors affecting that situation.” (Bailey & Pearson, 1983, p. 531). However, the existing definitions of user satisfaction leave room for interpretation (Thong & Yap, 1996). According to Hsiao and Chen (2021), users’ satisfaction is based on comparing the expected and actual performance. Thus, the customer will be satisfied if the users’ expectations are met. User satisfaction is affected by the negative emotions that users experience and predicts customers' intention to continue to use a product or service (Ashktorab et al., 2019; Følstad & Taylor, 2019; Kvale et al., 2021; Liao et al., 2009). Research by Chiou (2006) also demonstrated that customer loyalty and retention are heavily influenced by user satisfaction.

Brand attitude is also an important factor to consider since it is the overall evaluation summary of judgment to any brand-related information (Keller, 2003). A study by Pavone et al. (2022) demonstrated that when customers interact with a customer service chatbot, they blame the organization for negative outcomes (e.g., frustration and annoyance). However, they do not blame the organization for negative outcomes when interacting with human customer service employees (Pavone et al., 2022). This also highlights the importance of measuring brand attitude after the chatbot used a conversational repair strategy.

Regarding the importance of measuring trust, several studies indicated that trust is negatively affected by the negative emotions users encounter after exposure to conversational breakdowns (Følstad & Taylor, 2019; Kvale et al., 2021). Trust can be defined as: “one party (trustor) is willing to rely on the actions of another party (trustee), and the situation is directed to the future” (Przegalinska et al., 2019, p. 789). In the context of chatbots, trust can be defined as “the customer is willing to rely on the chatbot’s actions.” (Przegalinska et al., 2019). While trust is typically studied in the context of society, interpersonal relations, or organizations (Rousseau et al., 1998), it is increasingly being studied in the context of new technologies (Følstad et al., 2018). Since “trust is about expectations of the future” (Shneiderman, 2000, p. 58), the occurrence of the breakdown disconfirms the trustee’s positive expectations resulting in their trust being damaged (Seeger & Heinzl, 2021). While the study by Ashktorab et al. (2019) points out that some repair strategies (i.e., Repeat) appear less intelligent and therefore result in less trust, the current knowledge regarding users’ trust after the repair strategy is used is limited (Følstad et al., 2018).

Subquestions and Hypotheses

As elaborated in the theoretical framework, there is a knowledge gap regarding which error tags (most frequently) occur in Dutch human-chatbot conversations. Sanguinetti et al. (2020) described that these errors could pertain to customer messages, chatbot messages, or both. Study I will only analyze the customer's messages on errors since the chatbot's messages are analyzed on which repair strategies are used. For this purpose, the following two sub-questions will be investigated in study I: (RQ1a) *Which errors most frequently occur during conversations with a customer service chatbot?* And (RQ1b) *Which repair strategies are most frequently used by the customer service chatbot?* The hypotheses of study II are proposed after the first study is conducted since the findings of Study I will inform the experimental materials of Study II (see Preliminary conclusion).

Study I – Corpus Study

Method

Design

In study I, a dataset containing real human-chatbot conversations is analyzed to gain insight into (1) which errors occurred during the conversations, (2) the chatbots' used repair strategies, and (3) which customer errors and conversational repair strategies most frequently occurred. For this purpose, a thematic analysis with a deductive approach was conducted.

Dataset Description

The dataset contained real conversations between customers and the customer service chatbot of a Dutch railroad company that took place in 2021. Due to the time frame of this study, it would not be feasible to analyze all conversations. Another researcher accordingly reduced the dataset to 100 conversations. The researcher used the following selection criteria: all conversations were in Dutch, and each consisted of at least two messages from the customer service chatbot and two from the customer. These messages could be a free input or a chosen button. Only conversations between customers and the customer service chatbot were used; conversations for training purposes were excluded from the dataset. Moreover, messages sent after the customer was redirected to a human customer service employee (i.e., Defer) were not analyzed.

Procedure

The thematic analysis had a deductive approach (i.e., preconceived themes) and was conducted as follows: first, two codebooks were created to be able to analyze the conversations and determine to which category the errors (see Table 1) and repair strategies (see Table 3) belonged. The codebook on errors is based on a previous study by Sanguinetti et al. (2020). The codebook of the repair strategies is based on multiple studies (Ashktorab et al., 2019; Benner et al., 2021; Eberhart et al., 2022). Following the codebook, each conversation was coded for the different error types and repair strategies, and the frequencies of the errors and repair strategies were recorded. One researcher manually coded the dataset.

Preliminary Results

The corpus was analyzed through a thematic analysis with a deductive approach to investigate which errors and repair strategies occurred during conversations with the customer service chatbot of a Dutch railroad company (RQ1), which customer errors most frequently occurred (RQ1a), and which repair strategies were most frequently used by the customer service chatbot (RQ1b). In 41% of all conversations (N = 100), the customers' questions were about subscriptions, and in 25% of the conversations, the questions were about refunds. More topics covered in the conversations include customers' questions about bicycles (4%), travel information (6%), invoices (7%), and other less common topics (12%).

Errors

The corpus study's findings revealed that the following six errors occurred in the customers' messages: (1) Excess of information, (2) Ill-formed, (3) Lack of information, (4) Ignoring the question, (5) Non-understandable, and (6) Non-cooperativity. In 43.24% of all cases, there was no error or a chatbot error. As well as, in 11.71% of the cases, the error was unclear because (part of) sentences were filtered since they contained privacy-sensitive information (see Table 2). The error that most frequently occurred was the error tag Excess of information, meaning that the customer fed the customer service chatbot with too much (detailed) information that prevented the chatbot from answering (Sanguinetti et al., 2020).

Table 2

Frequencies errors

Error tag	Frequency	Percent
Ill-formed	10	9.01%
Excess of information	19	17.12%
Lack of information	7	6.31%
Ignoring question	4	3.6%
Non-understandable	7	6.31%
Non-cooperativity	3	2.7%
No error or error chatbot	48	43.24%
Unclear error (filtered)	13	11.71%
Total	111	100%

Note. $N = 100$

Repair Strategies

A conversational breakdown occurred in 69 out of 100 conversations. In those 69 conversations, 117 repair strategies were employed. This number translates to multiple occurrences of using more than one repair strategy during one conversation. The customer service chatbot's most frequently used repair strategy was Defer (48.71%), meaning that the chatbot cannot resolve the request and redirects the customer to a human customer service employee (Ashktorab et al., 2019). The second most frequently used repair strategy was Repeat (26.50%), meaning that the chatbot recognizes and explicitly indicates a conversational breakdown and asks the customer to repeat and/or rephrase the initial prompt (Ashktorab et al., 2019; Benner et al., 2021; Eberhart et al., 2022). The frequencies of all repair strategies and real-life examples from the corpus study can be found in Table 3.

Table 3

Codebook, frequencies, and examples of repair strategies

Repair strategy	**Description**	**Example codebook**	**Example Corpus**	**Frequency and percent**	
Repeat	The chatbot recognizes and explicitly indicates a potential breakdown, then asks the customer to repeat the initial prompt, which can include rephrasing (Ashktorab et al., 2019; Benner et al., 2021; Eberhart et al., 2022).	I don't understand. Can you repeat the question? I don't know what you mean.	"Ik begrijp helaas niet helemaal wat je bedoelt. Wil je de vraag nog een keer stellen in andere woorden?"	31	26.50%
Apologetic behavior	The chatbot shows empathy and apologizes (Benner et al., 2021)	Sorry, I cannot help you with this question/ request.	"Sorry, ik geloof dat ik je nog niet kan helpen."	8	6.84%
Top response	The chatbot outputs a response that does not match the users' requests. The user must initiate a repair after seeing the wrong response (Ashktorab et al., 2019).	The customer indicates that the chatbot's response does not match the question/request.	"Als ik het goed begrijp heb je een vraag over je abonnement."[2]	13	11.11%
Options	The chatbot lists options for potential intents (Ashktorab et al., 2019).	Please use the buttons below to let me know how I can help you.	"Wil je je abonnement tijdelijk stopzetten of wijzigen om kosten te besparen in verband met Covid-19? Klik op onderstaande knop."	8	6.84%
Defer	The chatbot cannot resolve a request and transfers the request to a human service agent (Ashktorab et al., 2019).	Unfortunately, I am not able to help you with this question. I will redirect you to a colleague.	"Zal ik je doorverbinden met mijn collega?"	57	48.71%

[2] The response of the chatbot did not match the users' request. The user indicated that the chatbot's answer was not correct after seeing the wrong response.

Preliminary Conclusion

The corpus study investigated which errors and repair strategies occurred during conversations with the customer service chatbot of a Dutch railroad company, which customer errors most frequently occurred (i.e., Excess of information and Ill-formed), and which repair strategies were most frequently used by the chatbot (i.e., Repeat, Options, and Defer). Regarding the errors, the error tag Excess of information was used as input for Study II; however, the error tag Ill-formed was not used since it is a complex error that is not easily simulated. Instead, the error tag Lack of information was used error. The third used error for Study II is the 'Unsolvable question' error. This error is compiled from the literature on why chatbots fail, which describes the importance for customers to understand the chatbot's capabilities, for example, the subjects he can help with (Janssen et al., 2021). Thus, whenever a customer asks a question unrelated to the company, it can be assigned to this error.

Regarding repair strategies, the repair strategies Defer and Repeat were used as input for the experimental materials of Study II. Although Top response was the third most frequently used repair strategy, it was not used for Study II since it requires the customer to initiate the repair. Instead, the fourth most frequently used repair strategy (e.g., Options) was selected for the experiment. Ultimately, three errors (Excess of information, Unsolvable question, Lack of Information) and repair strategies (Repeat, Options, Defer) were used as input for the experiment's stimuli.

Hypotheses Study II

After an error and conversational breakdown occur, it can negatively affect users (e.g., they get frustrated and annoyed), user satisfaction, brand attitude, and trust (Ashktorab et al., 2019; Benner et al., 2021; Følstad & Taylor, 2019). However, no indications are found in the literature that there might be a difference in user satisfaction, brand attitude, or trust between error types, especially since the three errors pertain to customer messages. Therefore, the following hypothesis will be tested in Study II**:**

H1 a/b/c: *There are no differences in user satisfaction (a), brand attitude (b), and trust (c) between the three error types.*

Regarding the conversational repair strategies, it is considered a major weakness of the repair strategy Repeat that it makes the chatbot appear less intelligent since it lacks resources for the user to repair the conversation, which could also result in less trust in the chatbot (Ashktorab et al., 2019). Moreover, Pavone et al. (2022) found that customers blame organizations for negative outcomes (e.g., frustration or annoyance) when interacting with chatbots. Thus, it can be argued that customers have lower levels of user satisfaction, brand attitude, and trust when the repair strategy Repeat is used compared to other repair strategies. Therefore, the following hypothesis will be tested in Study II**:**

H2 a/b/c: *User satisfaction (a), brand attitude (b), and trust (c) are lowest when the chatbot uses the repair strategy Repeat compared to the repair strategy Options and Defer.*

After a conversational breakdown occurred, it is reasonable to assume that customers would prefer to talk to a human service employee instead of the chatbot. Specifically, since (complex) questions are more likely to be solved by human service employees (Ashktorab et al., 2019). The error Excess of information deals with a customer's message that contains excessive (detailed) information (Sanguinetti et al., 2020), which is most likely an indication of a relatively complex question compared to other errors. Thus, if the error Excess of information occurs, it can be assumed that the repair strategy Defer has a more positive effect on user satisfaction (satisfaction with the repair strategy) since the customers' request is most likely to be solved by the human service employee (Ashktorab et al., 2019).

Regarding brand attitude, customers do not blame the organization for negative outcomes when interacting with human service employees (Pavone et al., 2022); therefore, it can be assumed that when the error Excess of information occurs, the repair strategy Defer has a more positive effect on brand attitude compared to the repair strategy Repeat and Options. Finally, even though the chatbot is not able to resolve the request itself when the repair strategy Defer is used, the chatbot indicates the breakdown and provides a helping solution in terms of redirecting the customer to a human service employee. As a result, the repair strategy Defer is also expected to have a more positive effect on trust since the customer can feel that the chatbot is honest and trustworthy by indicating that he is not (yet) able to help the customer with his request. Concluded, it can be argued that there is an interaction effect between the error types and repair strategies on user satisfaction, brand attitude, and trust. Therefore, the following hypothesis will be tested in Study II:

H3 a/b/c: *After the error excess of information occurs, the repair strategy Defer has a more positive effect on user satisfaction (a), brand attitude (b), and trust (c) compared to the repair strategy Repeat and Options.*

Study II - Online Experiment

Method

Design

The findings of study I informed the experimental materials for study II. Through an online experiment, it was investigated how people perceived the errors and repair strategies in terms of user satisfaction, brand attitude, and trust. The online experiment employed a 3 (error) x 3 (repair strategy) mixed design with error as a between-subjects factor and repair strategy as a within-subjects factor and was conducted in Dutch.

Participants

Initially, 184 people were recruited via convenience sampling. All participants had to be 18 or older and speak Dutch since the experiment was conducted in Dutch. Other than that, there were no special requirements to participate in the online experiment. Ultimately, 150 people fully completed the experiment (50 per condition). Of these participants, 64.7% were women ($N = 97$), and 35.3% were men ($N = 53$). Their age ranged from 18 to 70 ($M = 33.56$, $SD = 15.07$). Of all participants, 31.3% completed or were currently enrolled in a master's degree at university, 20% in a bachelor at university, 28.7% in higher education (HBO), 3.3% in an HBO master's degree, 12% in MBO, and 4.7% in high school or lower. The three conditions did not differ in the distribution of gender (χ^2 (2) = 3.74, $p = .155$) and educational level (χ^2 (10) = 8.43, $p = .587$). However, age ($F(2,147) = 3.543$, $p = .031$) was not evenly distributed between the conditions since participants in condition 3 were younger ($M = 29.04$, $SD = 12.72$) than participants in conditions 1 ($M = 35.34$, $SD = 15.48$) and 2 ($M = 36.30$, $SD = 16.03$). Finally, 88.7% of the participants indicated familiarity with chatbot technologies, 67.4% indicated using customer service chatbots frequently, and 44.6% indicated that their previous experiences with customer service chatbots were generally positive.

Stimuli

Every participant was asked to observe three screenshots of fictional conversations between a customer and a customer service chatbot. All conversations contained one error and one conversational repair strategy used by the customer service chatbot to overcome the conversational breakdown. The experiment used three types of errors (i.e., Excess of information, Unsolvable question, and Lack of information) and three repair strategies (i.e., Repeat, Options, Defer). Thus, three error types were formulated so all participants in the same condition saw the same error. In total, nine screenshots were designed since the experiment consisted of three conditions with different errors in each condition (see Table 4).

The three conversations in condition 1 were about a customer who wanted to cancel a subscription, condition 2 was about a customer requesting a refund, and condition 3 was about a customer's question about an invoice. The conversations from the corpus study inspired the topics for the experiment. In all conversations, the chatbot's messages were structured the same (e.g., greeting, closing message). Furthermore, the same names were used for the customer (i.e., Sofie) and the human service employee (i.e., Sanne). These aspects were considered to prevent that differences in the conversations could affect the results.

Table 4

Conditions online experiment

Condition	Type of error	Repair strategy	Screenshots conversation
Condition 1	Excess of information	Repeat, Options, Defer	Conversations 1, 2 and 3
Condition 2	Unsolvable question	Repeat, Options, Defer	Conversations 4, 5 and 6
Condition 3	Lack of information	Repeat, Options, Defer	Conversations 7, 8 and 9

Note. Participants are randomly and evenly assigned to one of the three conditions.

The nine screenshots of the conversations were designed using the online design platform Canva (see Appendix A). Efforts were made to emulate a real conversation with a customer service chatbot. No brand logo was used for the chatbot, as this could possibly affect the results. Instead, an image of a robot was used, making the screenshot seem more realistic. Figures 1 and 2 show a screenshot of conversations 3 (i.e., condition 1) and 8 (i.e., condition 3). The reason screenshots of conversations were used instead of implementing a pre-programmed chatbot or a link to an external chatbot is that they tend to get stuck often or cause Qualtrics to be slow, possibly resulting in participants quitting the experiment in an early stage. Moreover, since every participant had to observe three conversations, using a pre-programmed chatbot or redirecting to an external chatbot would be too time-consuming and makes it more difficult to control the conversation (e.g., the occurrence of unexpected errors).

Figure 1

Screenshot conversation 3

Note. Error Excess of information and Repair strategy Defer.

Figure 2

Screenshot conversation 8

Note. Error Lack of information and Repair strategy Options.

Measurements

User satisfaction

The variable user satisfaction was measured with a scale by Chung et al. (2020). The scale consisted of six items. However, the three items, 'I was satisfied with the experience of talking with the service agent,' 'I am content with the service agent,' and ' I am happy with the service agent,' were removed from the scale since there was some overlap with the other items. User satisfaction was measured on a 7-point Likert scale ranging from 1 = strongly disagree to 7 = strongly agree because studies pointed out that a 7-point scale is the most preferred rating scale since it is the most accurate and easiest to use (Preston & Colman, 2000; Taherdoost, 2019). The three remaining items have been partially modified to fit this study; the word 'service agent' has been replaced by 'chatbot' to avoid confusion among participants so that the same terms are used throughout.

Furthermore, all three items have been modified to focus specifically on satisfaction with the repair strategies used by the chatbot rather than referring to the chatbot itself. For example: '*I am satisfied with the service agent*' has been adapted to '*I am satisfied with the repair strategy used by the chatbot*.' Subsequently, all items were translated into Dutch to fit the experiment (see Table 5) and were reviewed by a second person to ensure the quality of the translation. The scale's reliability was good (ranging from Cronbach's α = .813 to Cronbach's α = .940) in all conditions (see Appendix B, Table 1).

Brand attitude

The variable brand attitude was measured with a scale by Liebrecht and Van Der Weegen (2019). The original scale consisted of eight items; however, this study used only five items since there were too many items to keep the experiment within 15 minutes. For each item, participants could rate to what extent they agreed or disagreed using a 7-point Likert scale ranging from 1 = strongly disagree to 7 = strongly agree. The items were originally written in Dutch and translated into English since this thesis is written in English (see Table 5). The scale's reliability was good (ranging from Cronbach's α = .874 to Cronbach's α = .932) in all conditions (see Appendix B, Table 2).

Trust in chatbots

The variable trust was measured with a scale by Jiang et al. (2022) and consisted of four items. For each statement, participants could rate to what extent they agreed or disagreed using a 7-point Likert scale (1 = strongly disagree to 7 = strongly agree). All items were translated into Dutch to fit the experiment (see Table 5) and were reviewed by a second person to ensure the quality of the translation. The scale's reliability was good (ranging from Cronbach's α = .796 to Cronbach's α = .942) in all conditions (see Appendix B, Table 3).

Demographic Questions and Ranking

The demographic information collected consisted of (1) gender, (2) age, (3) level of education, and (4) two statements adapted from Ashktorab et al. (2019) about participants' experience with customer service chatbots: "*I am familiar with chatbot technologies*" and "*I use customer service chatbots frequently."* Lastly, participants were asked to rate the following statement*: 'In general, my previous experiences with customer service chatbots have been positive.* Participants could rate the statements on a 7-point Likert scale ranging from 1 = strongly disagree to 7 = strongly agree. The complete questionnaire of the online experiment can be found in Appendix C.

Table 5

Items measuring user satisfaction, brand attitude, and trust.

Construct	Item in English	Item in Dutch
User satisfaction	I am satisfied with the repair strategy used by the chatbot	Ik ben tevreden over de reparatiestrategie die de chatbot heeft gebruikt
	The chatbot did a good job restoring the conversation	De chatbot heeft het gesprek goed hersteld
	The chatbot did what I expected to restore the conversation	De chatbot deed wat ik verwachtte om het gesprek te herstellen
Brand attitude	The brand is communicative	Het merk is communicatief
	The brand is respectful	Het merk is respectvol
	The brand is empathetic	Het merk is inlevend
	The brand is sympathetic	Het merk is sympathiek
	The brand is friendly	Het merk is vriendelijk
Trust	I believe that this chatbot is honest	Ik vind dat deze chatbot oprecht is
	I believe that this chatbot is trustworthy	Ik vind dat deze chatbot betrouwbaar is
	I believe that this chatbot is dependable	Ik vind dat deze chatbot deugdelijk is
	I believe that this chatbot is reliable	Ik vind dat deze chatbot geloofwaardig is

Note. Items measuring user satisfaction were adapted from Chung et al. (2020), items brand attitude from Liebrecht and Van Der Weegen (2019), and items trust from Jiang et al. (2022).

Procedure

The experiment was conducted with an online questionnaire through Qualtrics. The researcher generated a link that directed participants to the online questionnaire and was subsequently distributed through the researcher's network. People could participate anywhere with internet access with a smartphone, laptop, or desktop. The online experiment started with a brief introduction stating the purpose of the experiment, the approximate duration of completing the survey, and information about the participants' privacy rights, such as that participation was voluntary and that participants could withdraw at any time during the experiment. Moreover, the key terms conversational breakdown, repair strategy, and error were explained. All participants were required to read, understand, and sign the informed consent form before they could proceed to participate in the experiment.

After obtaining the consent, participants were asked to answer the three demographic questions (i.e., gender, age, education level) and three statements about customer service chatbots. Subsequently, participants were randomly and evenly assigned to one of the three conditions. All participants were shown three screenshots of fictional conversations between a customer and a customer service chatbot, where each conversation applied a different repair strategy. This meant all participants were shown all three conversational repair strategies but only one type of error, depending on their condition. After viewing each of the three screenshots of the conversations, participants were asked to rate 12 statements to measure user satisfaction (3 items), brand attitude (5 items), and trust (4 items). Finally, participants were asked to rank the three conversational repair strategies (i.e., Repeat, Options, Defer) based on which appeals to them the most. After that, participants could describe why they chose that particular order in an open text box. Following this open-ended question, the experiments' debriefing was shown, and participants were asked if they had any comments, questions, or concerns about the study. Contact details of the researcher were provided so that participants could contact the researcher if necessary. This was followed by a message thanking the participant for their time participating in the study.

Results

The experiment's data was analyzed in two parts to investigate how people perceived the three errors (i.e., Excess of information, Unsolvable question, Lack of information) and conversational repair strategies (i.e., Repeat, Options, Defer) resulting from the corpus study in terms of user satisfaction, brand attitude, and trust. Firstly, a mixed ANOVA with repeated measures was performed separately for each dependent variable: user satisfaction, brand attitude, and trust. Secondly, a thematic analysis was performed to gain insight into people's preferences for conversational repair strategies and the reasoning behind their ranking.

User Satisfaction

A mixed ANOVA with repeated measures was performed to test if the type of error and repair strategy affect user satisfaction, separately and together. Error type was measured between subjects, and repair strategy was measured within subjects. The data was first checked for outliers and normality (see Appendix D). Subsequently, the other assumptions of the mixed ANOVA with repeated measures were tested[3].

The main effect of error type showed that there was no statistically significant difference in user satisfaction between the errors $F(2, 147) = 1.43$, $p = .242$, partial $\eta^2 = .02$. The participants in the Excess of information condition had a mean user satisfaction score of 5.64 ($SD = 0.84$), in the Unsolvable question condition a mean score of 5.73 ($SD = 0.63$), and in the Lack of information condition a mean score of 5.48 ($SD = 0.78$). Since hypothesis 1a stated no difference in user satisfaction between the errors, H1a is accepted.

The mixed ANOVA also showed no statistically significant main effect of repair strategy on user satisfaction $F(1.83, 269.122) = 2.08$, $p = .131$, partial $\eta^2 = 0.14$. Although the participants' mean user satisfaction (H2a) score for the repair strategy Repeat ($M = 5.51$, $SD = 1.07$) was lower than the mean score for the repair strategy Options ($M = 5.61$, $SD = 1.08$), and Defer ($M = 5.73$, $SD = 1.10$), H2a is rejected since the results were not significant.

[3] There was homogeneity of variances and homogeneity of covariances, as assessed by Levene's test of homogeneity of variance ($p > .05$) and Box's test of equality of covariance matrices ($p = .055$). Mauchly's test of sphericity indicated that the assumption of sphericity was violated for the two-way interaction, $\chi^2(2) = 14.16$, $p = < .001$. The degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity.

Finally, there was no statistically significant interaction effect between error type and repair strategy on user satisfaction $F(3.662, 269.122) = .797$, $p = .518$, partial $\eta^2 = .01$ (see Table 6 for an overview of all descriptive statistics). Since hypothesis 3a stated that the repair strategy Defer has a more positive effect on user satisfaction after the error Excess of information occurs compared to the repair strategy Repeat and Options, H3a is rejected.

Brand Attitude

A mixed ANOVA with repeated measures was performed to test if the type of error and repair strategy affect brand attitude, separately and together. Error type was measured between subjects, and repair strategy was measured within subjects. The data was first checked for outliers and normality (see Appendix D). Subsequently, the other assumptions of the mixed ANOVA with repeated measures were tested[4].

The main effect of error type showed no statistically significant difference in brand attitude between the errors $F(2, 147) = 0.317$, $p = .728$, partial $\eta^2 = .004$. The participants in the Excess of information condition had a mean brand attitude score of 5.25 ($SD = 0.84$), in the Unsolvable question condition a mean score of 5.16 ($SD = 0.84$), and in the Lack of information condition a mean score of 5.13 ($SD = 0.75$). Since hypothesis 1b stated no difference in brand attitude between the errors, H1b is accepted.

The mixed ANOVA did show a statistically significant main effect of repair strategy on brand attitude $F(1.730, 254.330) = 38.844$, $p = < .001$, partial $\eta^2 = .21$ (large effect). The participants' mean brand attitude score for the repair strategy Options ($M = 5.12$, $SD = 1.04$) was .311[5] 95% CI [0.133, 0.488] higher compared to the repair strategy Repeat ($M = 4.81$, $SD = 1.07$), a statistically significant difference, $p = < .001$. Subsequently, the mean brand

[4] There was homogeneity of variances, as assessed by Levene's test of homogeneity of variance ($p > .05$). There was homogeneity of covariances, as assessed by Box's test of equality of covariance matrices ($p = .056$). Mauchly's test of sphericity indicated that the assumption of sphericity was violated for the two-way interaction, $\chi^2(2) = 24.76$, $p = < .001$. The degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity.

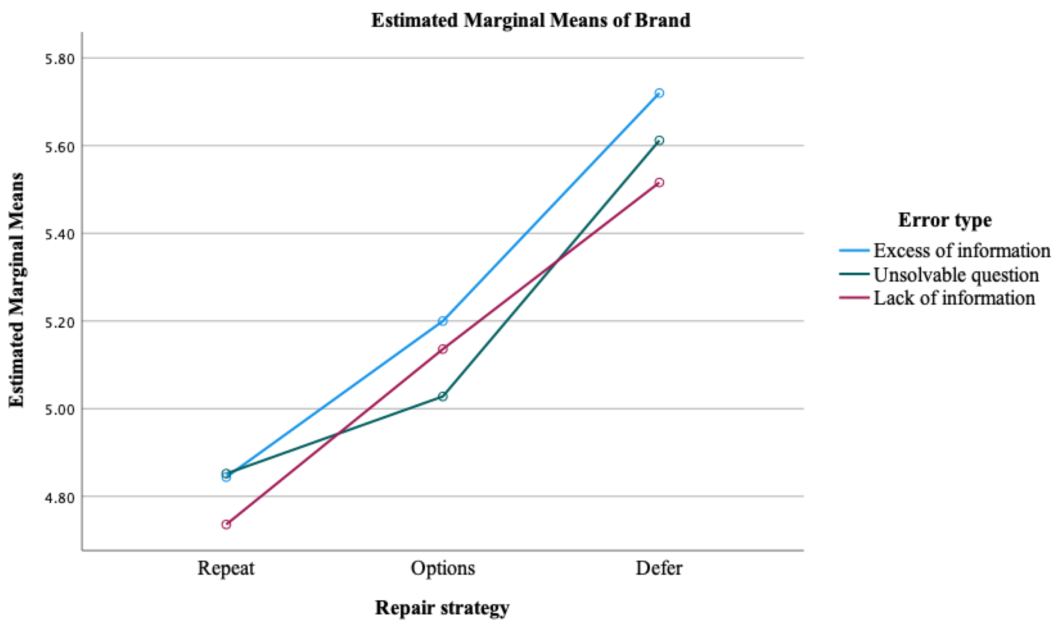
[5] Mean difference according to the pairwise comparisons.

attitude score for the repair strategy Defer (M = 5.62, SD = 1.00) was .805, 95% CI [0.550, 1.061] higher compared to the repair strategy Repeat, a statistically significant difference, p = < .001. Lastly, the mean score for the repair strategy Defer was .495 CI [0.265, 0.724] higher compared to the repair strategy Options, p = < .001 (see Figure 3). Since H2b stated that brand attitude would be lowest when the customer service chatbot uses the repair strategy Repeat compared to the repair strategy Options and Defer, H2b is supported.

Finally, there was no statistically significant interaction effect between error type and repair strategy on brand attitude $F(3.460, 254.330) = 0.377$, $p = .798$, partial $\eta^2 = .01$ (see Table 6 for an overview of all descriptive statistics). Since H3b stated that the repair strategy Defer has a more positive effect on brand attitude after the error Excess of information occurs compared to the repair strategy Repeat and Options, H3b is rejected.

Figure 3

Profile plot for the variable brand attitude



Trust

A mixed ANOVA with repeated measures was performed to test if the type of error and repair strategy affect trust separately and together. Error type was measured between subjects, and repair strategy was measured within subjects. The data was first checked for outliers and normality (see Appendix D). Subsequently, the other assumptions of the mixed ANOVA with repeated measures were tested[6].

The main effect of error type showed no statistically significant difference in trust between the errors $F(2, 147) = 0.26$, $p = .773$, partial $\eta^2 = .004$. The participants in the Excess of information condition had a mean trust score of 5.22 (*SD* = 0.89), a mean trust score of 5.34 (*SD* = 0.85) in the Unsolvable question condition, and a mean trust score of 5.30 (*SD* = 0.69) in the Lack of information condition. Since hypothesis 1c stated no difference in trust between the errors, H1c is accepted.

The mixed ANOVA did show a statistically significant main effect of repair strategy on trust $F(1.689, 248.327) = 20.737$, $p < .001$, partial $\eta^2 = 0.12$ (medium effect). The participants' mean trust score for the repair strategy Options (*M* = 5.28, *SD* = 0.96) was .262, 95% CI [0.09, 0.43] higher compared to the repair strategy Repeat (*M* = 5.02, *SD* = 0.99), a statistically significant difference, *p* = < .001. Subsequently, the participants' mean trust score for the repair strategy Defer (*M* = 5.57, *SD* = 1.09) was .553, 95% CI [0.31, 0.80] higher compared to the repair strategy Repeat (*M* = 5.02, *SD* = 0.99), a statistically significant difference, *p* = < .001. Lastly, the participant's mean trust score for the repair strategy Defer was .292 CI [0.09, 0,48] higher than the repair strategy Options (*M* = 5.28, *SD* = 0.96), *p* = .002 (see Figure 4). Since hypothesis 2c stated that trust in chatbots would be lowest when the customer service chatbot uses the repair strategy Repeat compared to the repair strategy Options and Defer, it can be concluded that H2c is supported.

---

[6] There was homogeneity of variances, as assessed by Levene's test of homogeneity of variance (*p* > .05). There was homogeneity of covariances, as assessed by Box's test of equality of covariance matrices (*p* = .014). Mauchly's test of sphericity indicated that the assumption of sphericity was violated for the two-way interaction, $\chi^2(2) = 29.67$, *p* = < .001. The degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity.

Finally, there was no statistically significant interaction effect between error type and repair strategy on trust $F(3.379, 248.327) = 1.554$, $p < .196$, partial $\eta^2 = .02$ (see Table 6 for an overview of all descriptive statistics). Concluded, since hypothesis 3c stated that the repair strategy Defer has a more positive effect on trust after the error Excess of information occurs compared to the repair strategy Repeat and Options, H3c can be rejected.

Figure 4

Profile plot for the variable trust

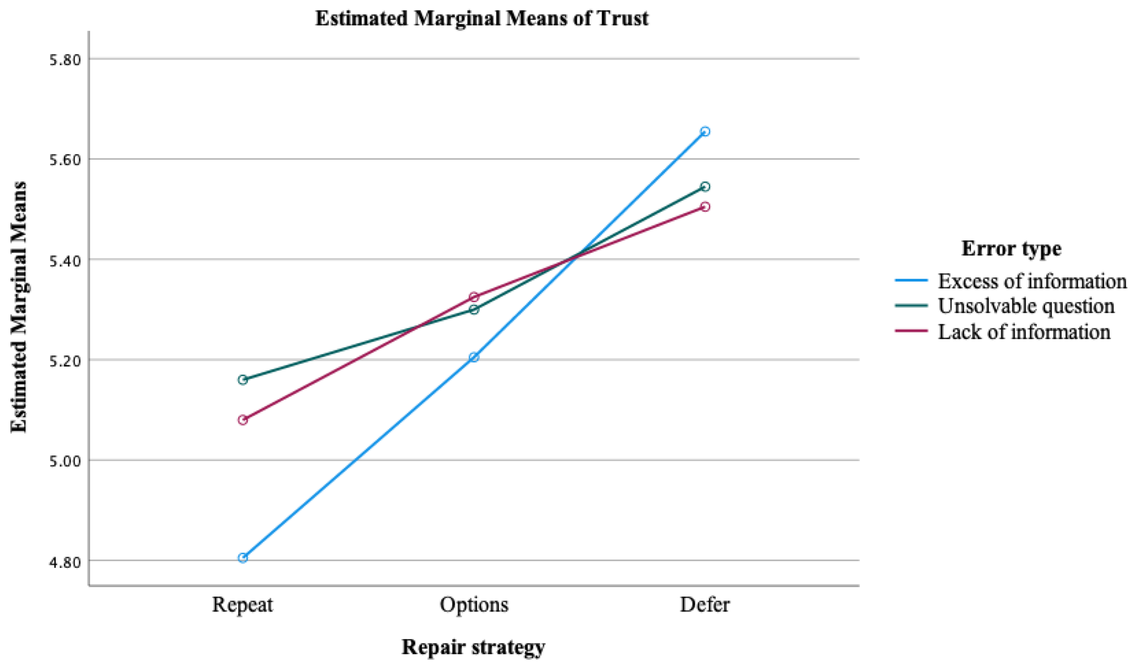


Table 6

Descriptive statistics for error type, repair strategy, and interaction

	User satisfaction		**Brand attitude**		**Trust**	
Error type	***M***	***SD***	***M***	***SD***	***M***	***SD***
Excess of information	5.64	0.84	5.25	0.84	5.22	0.89
Unsolvable question	5.73	0.63	5.16	0.84	5.34	0.85
Lack of information	5.48	0.78	5.13	0.75	5.30	0.69
Repair strategy						
Repeat	5.51	1.07	4.81	1.07	5.02	0.99
Options	5.61	1.08	5.12	1.04	5.28	0.96
Defer	5.73	1.10	5.62	1.00	5.57	1.09
Interaction						
Excess of information and Repeat	5.53	1.17	4.84	1.04	4.81	1.09
Excess of information and Options	5.55	1.21	5.20	1.03	5.21	1.01
Excess of information and Defer	5.85	1.13	5.72	0.97	5.66	1.17
Unsolvable question and Repeat	5.71	0.79	4.85	1.16	5.16	0.91
Unsolvable question and Options	5.68	0.84	5.03	1.06	5.30	0.97
Unsolvable question and Defer	5.79	1.00	5.61	0.96	5.55	1.08
Lack of information and Repeat	5.27	1.17	4.74	1.02	5.08	0.94
Lack of information and Options	5.61	1.18	5.14	1.04	5.33	0.92
Lack of information and Defer	5.55	1.17	5.52	1.08	5.51	1.02

Note. $N = 150$

Ranking and thematic analysis

All participants ($N = 150$) were asked to rank the three previously shown repair strategies (i.e., Repeat, Options, Defer) to examine how they perceived these repair strategies. Subsequently, the participants were asked to explain the reasoning behind their ranking to gain more insight into why they prefer specific repair strategies. The results of the ranking question revealed that the repair strategy Options were ranked first by 42% of the participants, and the repair strategy Defer was ranked first by 41.33%. The repair strategy Repeat was ranked first by 16.67% of the participants, as opposed to 54.67% who ranked that repair strategy third (see Table 7). This implies that the repair strategy Options and Defer were considered the best compared to the repair strategy Repeat.

A thematic analysis with an inductive approach was performed to analyze the results of the open-ended question asked after the ranking (see Appendix E). For the repair strategy Repeat, participants shed light on the negative aspects of this repair strategy and why they chose to rank this strategy third. Firstly, participants indicated that the repair strategy Repeat is ineffective and annoying. It was mentioned that not all people could easily rephrase their questions and that there is a possibility of ending up in an endless cycle of rephrasing questions and the chatbot asking to rephrase them again. On the other hand, participants also shed light on the positive aspect of the repair strategy Repeat, such as it could provide faster answers instead of being redirected to a human customer service employee (i.e., Defer).

Regarding the repair strategy Defer, participants frequently mentioned that redirecting to a human customer service employee can be slow and is one major disadvantage of this repair strategy. Participants indicated that when using a chatbot, it is preferred that the request is resolved quickly. However, when dealing with complex questions, participants preferred the repair strategy Defer since they had more trust in the human employee to resolve the request. One participant said: ” *It is nicer to communicate with a real employee for more* *complex questions."* [69]. Some participants revealed that they strongly prefer talking to a person instead of a chatbot and find it more customer-friendly than talking to a chatbot. For example, someone said: "*Nice to communicate with a real person*." [114].

Regarding the repair strategy Options, participants shed light on the positive aspects of this repair strategy and why they chose to rank it first or second. They indicated that this repair strategy is easy and fast. However, participants also mentioned the negative aspects of this repair strategy. They stated that there are limited options which can mean that the correct answer is not listed or that the listed options do not fully match the customer's request. For example, one participant stated: "*Choice options can be limiting; I prefer to talk to an employee if the chatbot does not understand it.*" [104].

Table 7

Participants preferred repair strategies

Repair strategy	**Ranking**	**Percent**
Repeat	1	16.67%
	2	28.67%
	3	54.67%
Options	1	42%
	2	38.67%
	3	19.33%
Defer	1	41.33%
	2	32.67%
	3	26%

Note. N = 150

Conclusion and Discussion

Conclusion

Study I (i.e., corpus study) investigated what errors and repair strategies (most frequently) occurred during conversations with a customer service chatbot, and Study II (i.e., online experiment) investigated how people perceived these errors and repair strategies in terms of user satisfaction, brand attitude, and trust. The corpus study's findings demonstrated that the following six error tags occurred during the conversations with the customer service chatbot of the Dutch railroad company: (1) Lack of information, (2) Ill-formed, (3) Excess of information, (4) Ignoring the question, (5) Non-understandable, and (6) Non-cooperativity. The error tag Excess of information (i.e., the customer provides the chatbot with excessive and redundant information) most frequently occurred. The corpus study's findings also revealed that the customer service chatbot alternately used the following five types of repair strategies to overcome the conversational breakdown: (1) Apologetic behavior, (2) Repeat, (3) Top response, (4) Defer, and (5) Options. Of these, Defer (i.e., the chatbot is unable to resolve the request and redirects the customer to a human service agent) was the customer service chatbot's most frequently used repair strategy.

Although the online experiment's findings ($N = 150$) demonstrated no statistically significant main effect of error types nor an interaction between error type and repair strategy on user satisfaction, brand attitude, or trust, this study did find statistically significant main effects of trust and brand attitude for the different repair strategies. However, no statistically significant main effect was found on user satisfaction for the various repair strategies. Firstly, the repair strategy Options (i.e., the chatbot lists options for potential intents) scored significantly higher on trust and brand attitude than the repair strategy Repeat (i.e., the chatbot asks the customer to repeat and rephrase the initial prompt). Secondly, the repair strategy Defer scored significantly higher on trust and brand attitude than the repair strategy Options and Repeat. The ranking of the three repair strategies (i.e., Repeat, Options, Defer) demonstrated that people considered the repair strategy Options and Defer to be the best repair strategies compared to the repair strategy Repeat. The thematic analysis of the open-ended question also revealed the reasoning behind the participant's preferences for repair strategies. Participants indicated that a longer waiting time is a major disadvantage of the repair strategy Defer. However, they also shed light on the advantages, such as that human service employees are more capable and trusted to solve complex requests than the chatbot.

Discussion

The purpose of Study I was to investigate what errors and conversational repair strategies occurred during conversations with a customer service chatbot (RQ1), which errors most frequently occurred (RQ1a), and which repair strategies were most frequently used by the customer service chatbot (RQ1b). The corpus study's findings revealed that there occurred six different errors and that the customer service chatbot used five different repair strategies to restore the conversation. These errors correspond with the error tags described in the literature (Sanguinetti et al., 2020). This is also true for the five repair strategies found (Ashktorab et al., 2019; Benner et al., 2021; Eberhart et al., 2022).

Study II aimed to investigate how people perceived these errors and repair strategies in terms of user satisfaction, brand attitude, and trust. In line with what was hypothesized, Study II did not demonstrate a significant difference in user satisfaction (H1a), brand attitude (H1b), and trust (H1c) between the error types. Further, Study II did demonstrate a significant difference in brand attitude and trust between the repair strategies. The repair strategy Options scored significantly higher on brand attitude and trust than the repair strategy Repeat, and the repair strategy Defer scored significantly higher on brand attitude and trust than the repair strategy Options and Repeat. These findings supported H2b and H2c, which hypothesized that brand attitude (H2b) and trust (H2c) are lowest when the chatbot uses the repair strategy

Repeat compared to repair strategy Options and Defer. Regarding the higher levels of brand attitude for the repair strategy Defer, this could be explained by the findings of a study by Pavone et al. (2022), which revealed that customers do not blame the organization for negative outcomes (e.g., frustration) when interacting with a human service agent (e.g., Defer). Regarding the higher levels of trust for the repair strategy Defer, it could be possible that participants perceived the intervention of the human service employee as necessary to help the customer with the request. Consequently, the findings provided no support for H2a, which hypothesized that user satisfaction (H2a) is lowest when the chatbot uses the repair strategy Repeat compared to the repair strategy Options and Defer. This unexpected finding might be explained by the fact that participants noticed that the customer's request was resolved in all cases, allowing for increased user satisfaction regardless of the repair strategy used. These results correspond with the data obtained in ranking the three repair strategies, whereas the repair strategy Defer and Options were preferred over the strategy Repeat.

Contrary to what was hypothesized, this study has been unable to demonstrate a statistically significant interaction between error type and repair strategy on user satisfaction, brand attitude, and trust. These findings suggest that it does not matter for the use of the repair strategies (i.e., Repeat, Options, and Defer) which error occurred. Consequently, the findings provided no support for H3, which hypothesized that after the error excess of information occurs, the repair strategy Defer has a more negative effect on user satisfaction (H3a), brand attitude (H3b), and trust (H3c) compared to the repair strategy Repeat or Options.

The analysis of the open-ended question asking participants why they preferred certain repair strategies revealed that repair strategy Defer was preferred if participants were dealing with a complex request, even though being redirected to a human customer service employee is often a slower process than other repair strategies. This finding suggests a possible connection exists between the type of request (simple request versus complex request) and the most suitable repair strategy for customer service chatbots to prevent or mitigate the negative consequences of a conversational breakdown.

Theoretical and Practical Implications

Theoretically, the corpus study's findings complement the existing literature on what errors and repair strategies occur in conversations with a Dutch customer service chatbot. The experiment's findings complement the current literature on conversational breakdowns and repair strategies. While studies by Ashktorab et al. (2019) and Benner et al. (2021) indicated that it is still unknown how the repair strategy Defer affects user's overall satisfaction and trust compared to other repair strategies, this study showed that the repair strategy Defer scored significantly higher on brand attitude and trust than other repair strategies. However, this study could not find a significant effect of the repair strategy Defer on user satisfaction.

Based on the existing literature, it can be stated that the occurrence of an error (resulting in a breakdown) compared to no error negatively affects users (through frustration, annoyance, or confusion), trust in the chatbot and can be harmful to the organization (Ashktorab et al., 2019; Benner et al., 2021; Følstad & Taylor, 2019; Pavone et al., 2022). Therein, the results of this study suggest that there is no need to consider the type of error and adjust the repair strategy accordingly, as there was no interaction effect between error type and repair strategy on user satisfaction, brand attitude, and trust. However, it is important to use the right repair strategy to prevent or mitigate the negative consequences of a breakdown (Ashktorab et al., 2019). The results demonstrated that repair strategies Defer and Options were considered the best for customer service chatbots.

Lastly, this study's findings also provide practical information for organizations and conversational A.I. developers for designing effective conversational repair strategies for customer service chatbots to prevent or mitigate the negative consequences (e.g., frustration) of conversational breakdowns and enhance brand attitude and trust in chatbots.

Limitations and Future Research

Although this study provides valuable insights into what errors and repair strategies occurred during conversations with a customer service chatbot, how people perceived them, and how they affected user satisfaction, brand attitude, and trust, several limitations might have influenced the results. Regarding the corpus study, there is one limitation. Since the corpus study examined the chatbot conversations of only one organization (i.e., a Dutch Railroad company), the results may be unrepresentative for customer service chatbots in other industries. Therefore, further research is needed in different sectors.

Furthermore, there are several limitations concerning the online experiment. Firstly, all participants were recruited via convenience sampling, resulting in relatively young, highly educated, and mostly female participants. This meant that the participants were not fully representative of the population, and further research should therefore focus on a more diverse group of participants. Secondly, the conversations used in the online experiment were fictitious. Although the real conversations from the corpus study inspired the conversations, this may have affected the results. Likewise, the experiment used screenshots of the conversations and not a real chatbot system or a pre-programmed chatbot that simulated an interactive experience of a real conversation. Future research could consider using a real chatbot system where participants can interact with the chatbot. However, researchers must consider that interacting with a real chatbot is more time-consuming for participants.

Thirdly, participants were not part of the conversation but had to assess user satisfaction based on their ability to empathize with the situation. Although most people have traveled by public transportation at some point and can probably empathize with the situation, it could potentially affect the results. Researchers should therefore consider conducting research with existing customers. Fourthly, the reader should bear in mind that this study examined only three types of errors and three repair strategies. It was beyond this study's scope to investigate other errors and repair strategies described in the existing literature. Hence, additional research that accounts for more errors and repair strategies is needed to fully understand which conversational repair strategies are best for customer service chatbots. Finally, this study has two additional recommendations for further research. Researchers could explore in which situations (simple versus complex requests) the repair strategy Defer is most effective, especially since the thematic analysis of the open-ended question revealed that people preferred the repair strategy Defer if they were dealing with complex requests. For future research, researchers could consider including a control condition (i.e., a conversation without an error) in the experiment to draw better conclusions.

To summarize, this study took you into the world of customer service chatbots, the benefits they can have for organizations and their customers, their shortcomings, such as miscommunication and conversational breakdowns, and finally, the use of repair strategies and its effects on user satisfaction, brand attitude, and trust. This study's findings contribute to improving and optimizing repair strategies for customer service chatbots so that you, as a customer, will leave your next chatbot conversation with a positive experience.

References

Ashktorab, Z., Jain, M., Liao, Q. V., & Weisz, J. D. (2019). Resilient Chatbots. https://doi.org/10.1145/3290605.3300484

Asisof, A. (2022). Towards a comprehensive repair framework for human-chatbot interaction. Proceedings of the 22nd ACM International Conference on Intelligent Virtual Agents. https://doi.org/10.1145/3514197.3549641

Bailey, J., & Pearson, S. W. (1983). Development of a Tool for Measuring and Analyzing Computer User Satisfaction. Management Science, 29(5), 530–545. https://doi.org/10.1287/mnsc.29.5.530

Benner, D., Elshan, E., Schöbel, S., & Janson, A. (2021). What do you mean? A Review on Recovery Strategies to Overcome Conversational Breakdowns of Conversational Agents. Proceedings of the 42nd International Conference on Information Systems (ICIS).https://www.researchgate.net/publication/354812835_What_do_you_mean_A_Review_on_Recovery_Strategies_to_Overcome_Conversational_Breakdowns_of_Conversational_Agents

Brinton, B., Fujiki, M., Loeb, D. F., & Winkler, E. (1986). Development of Conversational Repair Strategies in Response to Requests for Clarification. *Journal of Speech Language and Hearing Research*, *29*(1), 75–81. https://doi.org/10.1044/jshr.2901.75

Chiou, J. (2006). Service Quality, Trust, Specific Asset Investment, and Expertise: Direct and Indirect Effects in a Satisfaction-Loyalty Framework. Journal of the Academy of Marketing Science, 34(4), 613–627. https://doi.org/10.1177/0092070306286934

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

Cuadra, A., Li, S., Lee, H., Cho, J. H. D., & Ju, W. (2021). My Bad! Repairing Intelligent Voice Assistant Errors Improves Interaction. *Proceedings of the ACM on Human-computer Interaction*, *5*(CSCW1), 1–24. https://doi.org/10.1145/3449101

Eberhart, Z., Bansal, A., & McMillan, C. (2022). A Wizard of Oz Study Simulating API Usage Dialogues With a Virtual Assistant. *IEEE Transactions on Software Engineering*, *48*(6), 1883–1904. https://doi.org/10.1109/tse.2020.3040935

Følstad, A., & Skjuve, M. (2019). Chatbots for customer service. Proceedings of the 1st International Conference on Conversational User Interfaces - CUI ’19. https://doi.org/10.1145/3342775.3342784

Følstad, A., & Taylor, C. (2019). Conversational Repair in Chatbots for Customer Service: The Effect of Expressing Uncertainty and Suggesting Alternatives. Springer International Publishing EBooks, 201–214. https://doi.org/10.1007/978-3-030-39540-7_14

Følstad, A., Araujo, T., Papadopoulos, S., Law, E. L., Granmo, O., Luger, E., & Brandtzaeg, P. B. (2020). Chatbot Research and Design: Third International Workshop, CONVERSATIONS 2019, Amsterdam, The Netherlands, November 19–20, 2019,

Følstad, A., Nordheim, C. B., & Bjørkli, C. A. (2018). What Makes Users Trust a Chatbot for Customer Service? An Exploratory Interview Study. In Lecture Notes in Computer Science (pp. 194–208). Springer Science+Business Media. https://doi.org/10.1007/978-3-030-01437-7_16

Følstad, A., Nordheim, C. B., & Bjørkli, C. A. (2018). What Makes Users Trust a Chatbot for Customer Service? An Exploratory Interview Study. Springer International Publishing EBooks, 194–208. https://doi.org/10.1007/978-3-030-01437-7_16

Gnewuch, U., Morana, S., & Maedche, A. (2017, December 10). Towards Designing Cooperative and Social Conversational Agents for Customer Service. ResearchGate. https://www.researchgate.net/publication/320015931_Towards_Designing_Cooperative_and_Social_Conversational_Agents_for_Customer_Service

Grand View Research. (2023). Chatbot Market Size, Share & Trends Analysis Report By End Use (Large Enterprises, Medium Enterprises), By Application, By Type, By Product Landscape, By Vertical, By Region, And Segment Forecasts, 2022 - 2030. In Grand View Research (GVR-1-68038-598-4). Retrieved February 20, 2023, from https://www.grandviewresearch.com/industry-analysis/chatbot-market#

Grice, P. (1989). *Studies in the Way of Words*. Harvard University Press.

Griffiths, J. R., Johnson, F., & Hartley, R. (2007). User satisfaction as a measure of system performance. Journal of Librarianship and Information Science, 39(3), 142–152. https://doi.org/10.1177/0961000607080417

Hsiao, K., & Chen, C. (2021). What drives continuance intention to use a food-ordering chatbot? An examination of trust and satisfaction. Library Hi Tech, 40(4), 929–946. https://doi.org/10.1108/lht-08-2021-0274

Janssen, A., Grützner, L., & Breitner, M. H. (2021). Why do Chatbots fail? A Critical Success Factors Analysis. *ResearchGate*. https://www.researchgate.net/publication/354811221_Why_do_Chatbots_fail_A_Critical_Success_Factors_Analysis

Jiang, Y., Yang, X., & Zheng, T. (2022). Make chatbots more adaptive: Dual pathways linking human-like cues and tailored response to trust in interactions with chatbots. Computers in Human Behavior, 138, 107485. https://doi.org/10.1016/j.chb.2022.107485

Keller, K. L. (2003). Brand Synthesis: The Multidimensionality of Brand Knowledge. Journal of Consumer Research, 29(4), 595–600. https://doi.org/10.1086/346254

Kvale, K., Freddi, E., Hodnebrog, S., Sell, O. A., & Følstad, A. (2021). Understanding the User Experience of Customer Service Chatbots: What Can We Learn from Customer Satisfaction Surveys? In Springer eBooks (pp. 205–218). https://doi.org/10.1007/978-3-030-68288-0_14

Kvale, K., Sell, O. A., Hodnebrog, S., & Følstad, A. (2019). Improving Conversations: Lessons Learnt from Manual Analysis of Chatbot Dialogues. Lecture Notes in Computer Science, 187–200. https://doi.org/10.1007/978-3-030-39540-7_13

Li, T. J., Chen, J., Xia, H., Mitchell, T. M., & Myers, B. A. (2020). Multi-Modal Repairs of Conversational Breakdowns in Task-Oriented Dialogs. User Interface Software and Technology. https://doi.org/10.1145/3379337.3415820

Liao, C., Palvia, P., & Chen, J. (2009). Information technology adoption behavior life cycle: Toward a Technology Continuance Theory (TCT). International Journal of Information Management, 29(4), 309–320. https://doi.org/10.1016/j.ijinfomgt.2009.03.004

Liebrecht, C., & van der Weegen, E. (2019). Menselijke chatbots: een zegen voor online klantcontact? Het effect van conversational human voice door chatbots op social presence en merkattitude. Tijdschrift voor Communicatiewetenschap,47(3-4), 217-238

Lindgaard, G., & Dudek, C. (2003). What is this evasive beast we call user satisfaction? Interacting With Computers, 15(3), 429–452. https://doi.org/10.1016/s0953-5438(02)00063-2

Maroengsit, W., Piyakulpinyo, T., Phonyiam, K., Pongnumkul, S., Chaovalit, P., & Theeramunkong, T. (2019). A Survey on Evaluation Methods for Chatbots. https://doi.org/10.1145/3323771.3323824

Matsumoto, K., Sasayama, M., Yoshida, M., Kita, K., & Ren, F. (2022). Emotion Analysis and Dialogue Breakdown Detection in Dialogue of Chat Systems Based on Deep Neural Networks. Electronics, 11(5), 695. https://doi.org/10.3390/electronics11050695

Pavone, G., Meyer-Waarden, L., & Munzel, A. (2022). Rage Against the Machine: Experimental Insights into Customers' Negative Emotional Responses, Attributions of Responsibility, and Coping Strategies in Artificial Intelligence–Based Service Failures. Journal of Interactive Marketing, 58(1), 52–71. https://doi.org/10.1177/10949968221134492

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. Acta Psychologica, 104(1), 1–15. https://doi.org/10.1016/s0001-6918(99)00050-5

Przegalinska, A., Ciechanowski, L., Stroz, A., Gloor, P. A., & Mazurek, G. (2019). In bot we trust: A new methodology of chatbot performance measures. Business Horizons, 62(6), 785–797. https://doi.org/10.1016/j.bushor.2019.08.005

Reinkemeier, F., & Gnewuch, U. (2022). Designing Effective Conversational Repair Strategies for Chatbots. ResearchGate. https://www.researchgate.net/publication/360085227_Designing_Effective_Conversational_Repair_Strategies_for_Chatbots

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. F. (1998). Not So Different After All: A Cross-Discipline View Of Trust. Academy of Management Review, 23(3), 393–404. https://doi.org/10.5465/amr.1998.926617

Sanguinetti, M., Mazzei, A., Patti, V., Scalerandi, M., Mana, D., & Simeoni, R. B. (2020). Annotating Errors and Emotions in Human-Chatbot Interactions in Italian. Proceedings of the 14th Linguistic Annotation Workshop, 148–159. https://www.aclweb.org/anthology/2020.law-1.14/

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, *53*(2), 361–382. https://doi.org/10.1353/lan.1977.0041

Seeger, A., & Heinzl, A. (2021). Chatbots often Fail! Can Anthropomorphic Design Mitigate Trust Loss in Conversational Agents for Customer Service? In European Conference on Information Systems. https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1011&context=ecis2021_rp

Setlur, V., & Tory, M. (2022). How do you Converse with an Analytical Chatbot? Revisiting Gricean Maxims for Designing Analytical Conversational Behavior. In *CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3491102.3501972

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

Shneiderman, B. (2000). Designing trust into online experiences. Communications of the ACM, 43(12), 57–59. https://doi.org/10.1145/355112.355124

Taherdoost, H. (2019). What Is the Best Response Scale for Survey and Questionnaire Design; Review of Different Lengths of Rating Scale / Attitude Scale / Likert Scale. HAL (Le Centre Pour La Communication Scientifique Directe).

Thong, J. Y., & Yap, C. (1996). Information systems effectiveness: A user satisfaction approach. Information Processing and Management, 32(5), 601–610. https://doi.org/10.1016/0306-4573(96)00004-0

Toader, D., Boca, G., Toader, R., Măcelaru, M. V., Toader, C., Ighian, D. C., & Radulescu, A. (2019). The Effect of Social Presence and Chatbot Errors on Trust. *Sustainability*, *12*(1), 256. https://doi.org/10.3390/su12010256

Appendix A

Stimulus Experiment

Figure 1

Conversation 1 (error 1 x Repeat)

Figure 2

Conversation 2 (error 1 x Options)

Figure 3

Conversation 3 (error 1 x Defer)

Figure 4

Conversation 4 (error 2 x Repeat)

Figure 5

Conversation 5 (error 2 x Options)

Figure 6

Conversation 6 (error 2 x Defer)

Figure 7

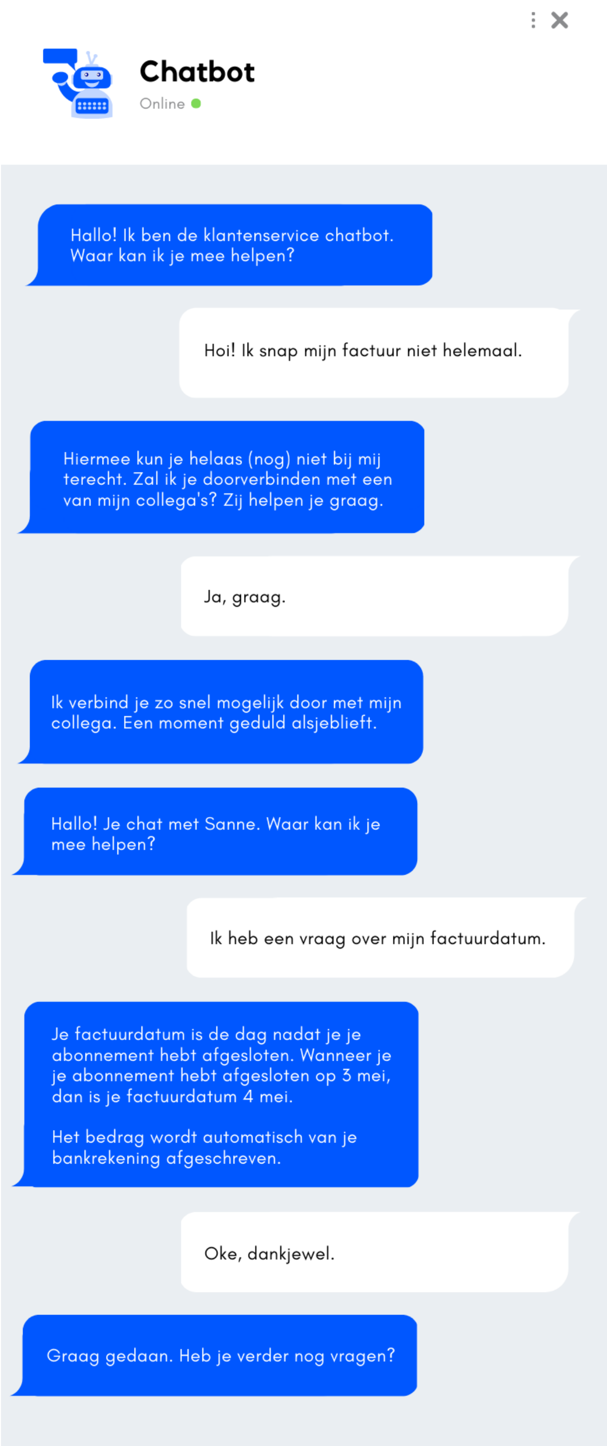
Conversation 7 (error 3 x Repeat)

Figure 8

Conversation 8 (error 3 x Options)

Figure 9

Conversation 9 (error 3 x Defer)



Appendix B

Overview Cronbach's Alpha for dependent variables

Table 1

Cronbach's alpha for user satisfaction in all conditions

Error type	Repair strategy	Cronbach's Alpha
Error 1	Repeat	.844
	Options	.882
	Defer	.869
Error 2	Repeat	.836
	Options	.813
	Defer	.911
Error 3	Repeat	.911
	Options	.940
	Defer	.861

Note. Number of Items = 3

Table 2

Cronbach's alpha for brand attitude in all conditions

Error type	Repair strategy	Cronbach's Alpha
Error 1	Repeat	.904
	Options	.902
	Defer	.926
Error 2	Repeat	.896
	Options	.894
	Defer	.932
Error 3	Repeat	.874
	Options	.918
	Defer	.932

Note. Number of Items = 5

Table 3

Cronbach's alpha for trust in all conditions

Error type	Repair strategy	Cronbach's Alpha
Error 1	Repeat	.829
	Options	.908
	Defer	.942
Error 2	Repeat	.796
	Options	.887
	Defer	.937
Error 3	Repeat	.891
	Options	.905
	Defer	.910

Note. Number of Items = 4

Appendix C

Survey Online Experiment

Start of Block: Briefing

Briefing Beste participant,
Bedankt voor uw interesse om deel te nemen aan dit online onderzoek. Dit onderzoek is onderdeel van mijn masterscriptie in Communicatie- en Informatiewetenschappen aan Tilburg University. Het is mogelijk om deel te nemen via uw smartphone, tablet of computer/laptop met internetverbinding. Het experiment duurt ongeveer 10 minuten.

Waar gaat deze studie over?
Dit onderzoek gaat over miscommunicatie en fouten tijdens gesprekken met klantenservice chatbots. Een klantenservice chatbot werkt op basis van kunstmatige intelligentie (A.I.) en is 24/7 beschikbaar om klanten met vragen te kunnen helpen. Een voorbeeld van een klantenservice chatbot is Billie van de online webwinkel Bol.com.

Contactgegevens
Voor vragen met betrekking tot dit onderzoek kunt u contact opnemen met Jasmin via: j.j.verhagen@tilburguniversity.edu.

Voorwaarden
Voordat u kunt deelnemen aan dit onderzoek is het van belang dat u onderstaande informatie goed doorleest:
- Ik begrijp dat deelname aan dit onderzoek vrijwillig is
- Ik begrijp dat ik op elk moment tijdens deelname aan dit onderzoek kan stoppen
- Ik begrijp dat mijn antwoorden volledig anoniem zijn en vertrouwelijk worden bewaard
- Ik begrijp dat er geen risico's zijn verbonden aan dit onderzoek

Als u akkoord gaat met deelname aan dit onderzoek selecteert u: 'ik ga akkoord met de voorwaarden'. Wanneer u niet wil deelnemen selecteert u 'ik ga niet akkoord met de voorwaarden' waarna de vragenlijst automatisch wordt afgesloten.

Selecteer hieronder of u wel of niet akkoord gaat met de voorwaarden

o Ik ga akkoord met de voorwaarden (1)

o Ik ga niet akkoord met de voorwaarden (2)

End of Block: Briefing

Start of Block: Demografische gegevens

Q3 Wat is uw geslacht?

o Vrouw (1)

o Man (2)

o Non-binair (3)

o Zeg ik liever niet (4)

o Anders (5) ____

Q4 Wat is uw leeftijd?

____

Q5 Wat is het hoogste opleidingsniveau dat u heeft voltooid of op dit moment volgt?

o Middelbareschooldiploma of lager (1)

o MBO (2)

o HBO (3)

o HBO master (4)

o Bachelor aan de universiteit (5)

o Master aan de universiteit (6)

End of Block: Demografische gegevens

Start of Block: Stellingen

Q6 Beoordeel onderstaande stellingen

	Volledig oneens (1)	Oneens (2)	Enigzins oneens (3)	Niet mee eens of oneens (4)	Enigzins mee eens (5)	Mee eens (6)	Volledig mee eens (7)
Ik ben bekend met chatbot technologieën (1)	O	O	O	O	O	O	O
Ik maak vaak gebruik van klantenservice chatbots (2)	O	O	O	O	O	O	O
Over het algemeen zijn mijn eerdere ervaringen met klantenservice chatbots positief (3)	O	O	O	O	O	O	O

End of Block: Stellingen

Start of Block: Conditie 1[7]

Zometeen krijgt u drie screenshots te zien van gesprekken tussen een klantenservice chatbot van een OV-bedrijf en een klant. In elk gesprek past de chatbot een reparatiestrategie toe. Na ieder screenshot wordt gevraagd om stellingen te beoordelen op een 7-punts Likert schaal. Om deze stellingen te kunnen beantwoorden is het van belang dat u begrijpt wat een reparatiestrategie is. Lees daarom onderstaande definitie goed door.

Wat is een reparatiestrategie?

Regelmatig ontstaan er gespreksonderbrekingen tijdens gesprekken met klantenservice chatbots. Deze gespreksonderbrekingen ontstaan doordat chatbots niet altijd in staat zijn om (complexe) vragen van de klant correct te interpreteren of de bedoelde betekenis van een vraag te begrijpen. De chatbot kan dan antwoorden met een bericht dat hij de vraag niet begrijpt en zal proberen om het gesprek voort te zetten door een reparatiestrategie te gebruiken. Voorbeeld: de chatbot geeft aan dat hij met dit onderwerp (nog) niet kan helpen en onderneemt dan actie om het gesprek voort te zetten.

---

[7] In all, there were three conditions in the experiment. Since only the content of the screenshots differed and the set-up was the same in all conditions, only condition 1- screenshot 1 is shown here.

Conditie 1 - screenshot 1 (error 1 x repair strategy repeat)

Hieronder ziet u een screenshot van een chatbotgesprek met reparatie. Bekijk deze goed en beantwoord dan de vragen.

Screenshot 1 displayed

Beoordeel de volgende statements over het screenshot die u zojuist heeft gezien.

Statements user satisfaction[8]

	Volledig oneens (1)	Oneens (2)	Enigzins oneens (3)	Niet mee eens of oneens (4)	Enigzins mee eens (5)	Mee eens (6)	Volledig mee eens (7)
Ik ben tevreden over de reparatiestrategie die de chatbot heeft gebruikt (1)	O	O	O	O	O	O	O
De chatbot heeft het gesprek goed hersteld (2)	O	O	O	O	O	O	O
De chatbot deed wat ik verwachtte om het gesprek te herstellen (3)	O	O	O	O	O	O	O

[8] Titles of constructs (e.g., user satisfaction) were not visible to participants.

Beoordeel de volgende statements over het screenshot die u zojuist heeft gezien.

Statements trust

	Volledig oneens (1)	Oneens (2)	Enigzins oneens (3)	Niet mee eens of oneens (4)	Enigzins mee eens (5)	Mee eens (6)	Volledig mee eens (7)
Ik vind dat deze chatbot oprecht is (1)	O	O	O	O	O	O	O
Ik vind dat deze chatbot betrouwbaar is (2)	O	O	O	O	O	O	O
Ik vind dat deze chatbot deugdelijk is (3)	O	O	O	O	O	O	O
Ik vind dat deze chatbot geloofwaardig is (4)	O	O	O	O	O	O	O

Uitleg brand attitude

De klantenservice chatbot staat in dienst van een merk, in dit geval een OV-bedrijf. De volgende statements gaan over dit merk.

Beoordeel de volgende statements over het screenshot die u zojuist heeft gezien.

Statements brand attitude

	Volledig oneens (1)	Oneens (2)	Enigzins oneens (3)	Niet mee eens of oneens (4)	Enigzins mee eens (5)	Mee eens (6)	Volledig mee eens (7)
Het merk is communicatief (1)	O	O	O	O	O	O	O
Het merk is respectvol (2)	O	O	O	O	O	O	O
Het merk is inlevend (3)	O	O	O	O	O	O	O
Het merk is sympathiek (4)	O	O	O	O	O	O	O
Het merk is vriendelijk (5)	O	O	O	O	O	O	O

Conditie 1 - screenshot 2 (error 1 x repair strategy options)

Hieronder ziet u een screenshot van een chatbotgesprek met reparatie. Bekijk deze goed en beantwoord dan de vragen.

Screenshot 2 displayed

Beoordeel de volgende statements over het screenshot die u zojuist heeft gezien.

→ Statements user satisfaction
→ Statements trust
→ Statements brand attitude

Conditie 1 - screenshot 3 (error 1 x repair strategy defer)

Hieronder ziet u een screenshot van een chatbotgesprek met reparatie. Bekijk deze goed en beantwoord dan de vragen.

Screenshot 3 displayed

Beoordeel de volgende statements over het screenshot die u zojuist heeft gezien.
→ Statements user satisfaction
→ Statements trust
→ Statements brand attitude

End of Block: Conditie 1

Start of Block: Ranking reparatiestrategieën

Zojuist heeft u drie screenshots gezien van een gesprek tussen een klantenservice chatbot van een OV-bedrijf en een klant. In ieder van deze gesprekken heeft de klantenservice chatbot een reparatiestrategie gebruikt. Hieronder kunt u de drie reparatiestrategieën ranken op basis van welke reparatiestrategie u het beste vindt (1 = beste reparatiestrategie). U kunt de volgorde veranderen door de reparatiestrategie te slepen en op een andere plek neer te zetten.

______ Strategie: opties. De chatbot geeft aan dat hij niet begrijpt wat de klant bedoelt en vraagt de klant via buttons te vertellen waar de vraag over gaat. (1)
______ Strategie: herhalen. De chatbot geeft aan dat hij niet begrijpt wat de klant bedoelt en vraagt of de klant de vraag wil herhalen. Daarbij geeft hij de tip hoe de vraag het best geformuleerd kan worden (kort en bondig). (2)
______ Strategie: doorverwijzen naar een medewerker. De chatbot geeft aan dat hij de klant niet kan helpen met deze vraag en vraagt de klant of hij doorverbonden wil worden met een medewerker. (3)

Q51 Wat is de reden dat u voor deze volgorde heeft gekozen?

____

End of Block: Ranking reparatiestrategieën

Appendix D

Outliers and assumption of normality for dependent variables

User satisfaction

The dataset contained eight outliers. Cases were considered outliers if the studentized residual value was less than -3 or greater than 3. Two outliers were found for user satisfaction and repair strategy ‘repeat,’ with studentized residual values of -3.44 and -3.05. There were also two outliers for user satisfaction and repair strategy ‘options,’ with studentized residual values of -4.27 and -3.65. Finally, there were four outliers for user satisfaction and repair strategy ‘defer,’ with studentized residual values of -4.17, -3.53, -3.53, and -3.48. These outliers are neither the result of a measurement error nor a data entry error and are retained in the dataset since they represent natural variations. The data was not normally distributed for each combination of the levels of the between- and within-subjects factors, as assessed by Shaprio-Wilk’s test ($p < .05$). However, ANOVA is fairly ‘robust’ to deviations from normality, especially since there were 50 participants in each condition.

Brand attitude

The dataset contained four outliers. There was one outlier for brand attitude and repair strategy ‘options,’ with a studentized residual value of -4.01. There were three outliers for brand attitude and repair strategy ‘defer,’ which had studentized residual values of -3.73, -3.63, and -3.13. These outliers are neither the result of a measurement error nor a data entry error and are retained in the dataset since they represent natural variations. The data was not normally distributed for each combination of the levels of the between- and within-subjects factors, as assessed by Shaprio-Wilk’s test ($p < .05$).

Trust

The dataset contained four outliers. There were two outliers for trust in chatbots and repair strategy ‘repeat,’ with studentized residual values of -3.91 and -3.17. One outlier was found for trust in chatbots and repair strategy ‘options,’ with a studentized residual value of -3.97. Finally, there was one outlier for trust in chatbots and repair strategy ‘defer,’ which had a studentized residual value of -3.28. These outliers are neither the result of a measurement error nor a data entry error and are retained in the dataset since they represent natural variations. The data was not normally distributed for each combination of the levels of the between- and within-subjects factors, as assessed by Shaprio-Wilk’s test ($p < .05$).

Appendix E

Thematic analysis of preferred repair strategies

Table 1

Quotes and corresponding themes (*N* = 129)

P	**Quote**	**Theme**
1	Ik denk dat de eerste optie uiteindelijk het snelst is, ook bij complexe vraagstukken	Faster
2	Eerst zal de chatbox je zoveel mogelijk helpen en als dat niet lukt, uiteindelijk zal er een medewerker je toch te woord staan	
3	soms is het nodig om een mederwerker er bij te halen maar niet in alle gevallen	Efficient
4	Via een medewerker voelt het vertrouwd.	Trust in human employee (defer)
5	Ik denk dat opties bieden een goede manier is om snel mensen te helpen, contact met een medewerker staat op 2 omdat het frustrerend kan zijn om helemaal geen contact te kunnen krijgen met een medewerker.	Faster
6	via buttons begrijp de chatbot je beter. Daarnaast communiceer ik met de chatbot zodat ik geen medewerker hoef te spreken. Vandaar deze volgorde.	Preference for options
7	Als er een medewerker bij geroepen wordt heb ik het idee dat het probleem directer opgelost word, want een echt persoon heeft je geholpen.	Faster
8	Voelt persoonlijker.	Personal
9	Klant eerst zelf zijn vraag laten verbeteren lijkt effectiever	Effective
10	Ik vind het altijd fijner om geholpen te worden door een medewerker, omdat de kans best groot is dat de chatbot het de tweede keer ook niet helemaal gaat begrijpen	Trust in human employee (defer) Possibly an endless circle (repeat)
11	Herhalen beste keuze. Buttons kunnen misschien wat afstandelijk overkomen. Doorverwijzen is ook een goede, maar duurt misschien langer omdat er een medewerker bij moet komen. Bij doorverwijzen kan het misschien ook wat 'onprofessioneel' overkomen, omdat het lijkt alsof de chatbot niet goed werkt	Not personal (options) Slow (defer) Chatbots are not yet well developed (defer)
12	Met optie 1 kan iemand meteen door (indien de juist opties worden aangeboden). Met optie 2 moet je even wachten totdat iemand je verder kan helpen. Optie 3 vind ik het minst prettig omdat de kans	Faster (options) Slow (defer) Possibly an endless circle

	bestaat dat na de herformulering de chatbot het nog steeds niet begrijpt.	(repeat)
13	Ik zou het liefst met het minste moeite mijn vraag beantwoord willen hebben en dat is met deze orde	Easy
14	Persoonlijk contact vind ik belangrijk. Doorverbinden met een medewerker is het meest persoonlijk.	Personal (defer)
15	Persoonlijk heb ik moeite met het kiezen van de juiste trefwoorden, daarom heb ik liever een voorgestelde keuze dan dat ik mijn eigen korte woorden moet formuleren. Bovendien geef ik de voorkeur aan het contact met een echt mens boven het zelf schrijven van de korte woorden, maar ik zet het alleen op de tweede positie omdat het meer tijd kost	Faster (repeat) Trust in human employee (defer) Slow (defer) Possibly an endless circle (repeat)
16	Omdat ik vaak het gevoel heb dat chatbots niet goed helpen kunnen met complexe onderwerpen	Complex question (defer)
17	vond die buttons niet echt overzichtelijk. en fijn om toch doorverbonden te worden met een persoon wanneer de chatbot niet in staat is om te helpen.	Not clear (options) Trust in human employees (defer)
18	De eerste zorgt ervoor dat ik het idee heb dat het merk me sowieso helpt. De tweede doet heel netjes helpen met het probleem maar mist die menselijke touch. De laatste voelt vrij koud aan als ik het zo lees.	Not personal (options)
19	opties maakt het voor klanten makkelijker om aan te geven wat hun vraag is.	Easy (options)
20	Opties zijn top, sommige mensen zullen niet weten hoe effectief te praten met een bot, dus dit kan enorm helpen. Doorverwijzen vind ik normaal vervelend, want ik wil het zelf uitzoeken.	Annoying (defer)
21	Ik zat the twijfelen tussen 1 en 2 welke beste is, ligt er aan per vraag. Uit ervaring makelijk vragen vind ik button fijnere bij moeilijker vragen vind ik fijn doorgestuurd worden	Complex question (defer) Easy question (options)
22	Meest efficiënte in mijn ogen	Efficient
23	Dat vind ik zo	
24	Persoon/medewerker voorkeur	Trust in human employee (defer)
25	Ik vind het leuk als er een mogelijkheid is om doorverbonden te worden met een medewerker als de chatbot niet kan helpen. Op de tweede plaats de strategie met opties omdat je dan de mogelijkheid hebt om voor een optie te kiezen die de chatbot kan beantwoorden. En dan de strategie met herhalen op de derde plaats, al vind ik het ook hier heel positief dat de chatbot een tip geeft met hoe de vraag het best geformuleerd kan worden.	

26	medewerkers lijkt me altijd het beste, maar niet altijd mogelijk. van daar zijn de opties een goede vervanger	Trust in human employee (defer) Effective (options)
27	ik vind de opties die de chatbot aangaf erg handig, dan kan je kijken of je daarmee uit de voeten kan en anders doorverbonden kan worden. De eerste optie was het meest onhandig omdat je dan eeuwig door kan blijven vragen in sommige gevallen.	Not effective (repeat) Possibly an endless circle (repeat)
28	Bij de doorverwijzing kreeg de klant een standaard antwoord terug van een medewerker waardoor ik het antwoord minder persoonlijk/inlevend vond. Hierdoor twijfelde ik aan de oprechtheid.	Not personal (defer)
29	Persoonlijke, meer mogelijkheden om gedetailleerde uitleg te geven als het gaat om de omstandigheden van het geval indien dit relevant is	Complex question (defer)
30	Behulpzaam, niet zelf de vraag kort nogmaals te hoeven herhalen!	Not effective (repeat)
31	Ik heb het meeste vertrouwen in een mens om de juiste info te geven.	Trust in human employee (defer)
32	Het fijnste is om een medewerker te spreken als een chatbot je niet kan helpen	Trust in human employee (defer)
33	Ik vind een doorverwijzing naar een medewerker altijd de beste optie om uiteindelijk een duidelijk antwoord te krijgen	Trust in human employee (defer)
34	Een ‘echte’ medewerker geeft je het gevoel dat ze je echt willen helpen. Opties geven geeft aan dat hij meedenkt. Herhalen is simpelweg irritant en geeft je het gevoel dat je voor niks aan het typen bent.	Trust in human employees (defer) Annoying (repeat)
35	Direct doorverbinden naar een collega kan soms lang duren en niet wat je wil als je kiest voor contact met een chatbot, daarom staar die op q ipv op 1. De buttons zijn altijd een goede, snelle en duidelijke oplossing vind ik zelf om snel tot het probleem te komen. De vraag anders herhalen werkt vaak slecht en is frustrerend.	Slow (defer) Faster (options) Frustrating (repeat)
36	Niet te veel tekst, lekker snel	Faster
37	Om een natuurlijk gesprek te houden, is het vragen en opnieuw formuleren het fijnst.	Preference for repeat
38	Persoonlijke communicatie, ook in de vorm van chat, geeft mij meer vertrouwen. De bereidheid om te helpen 'hoor' ik dan meer.	Personal (defer) Trust in human employees (defer)
39	Vervelend als er beperkte opties worden gegeven. Het komt voor dat mijn optie daar niet bij zit en dan kom ik niet verder. Daarom staat deze op nr. 3. Geholpen worden door echte personen heeft nog altijd mijn voorkeur, vandaar dat die op de 1e plaats staat	Limited choices (options) Preference for defer

40	Ik zou eerst voor optie 2 gaan als dat niet lukt is het fijn als de chatbot optie 1 aanbiedt.	Preference for options
41	Vind ik de fijnste optie	Preference for options
42	Een chatbox die doorverbind is niet wat je wil anders 2 vind ik gelijkwaardig	Preference for repeat and options
43	Ik spreek het liefste meteen met een medewerker. Opties is daarna de snelste optie om de juiste vragen beantwoord te hebben	Fast (options)
44	1. de chatbot reageerde hier het fijnst voor mij. Wel is het zo dat je vaak lang op een medewerker moet wachten vaak te lang. Het had opgelost kunnen worden met strategie 2 maar de chatbot had kunnen filteren abonnement opzeggen. Kunnen antwoorden met uit je bericht begrijp ik dat je je abonnement op wilt zeggen. druk voor bevestiging op de juiste button. Strategie 3 was vreselijk. Dat de chatbot een robot is betekent niet dat wij dat ook zijn. Tegelijkertijd komt het ook onvriendelijk over. Op zijn minst had de chatbot iets met dat bericht gekund. zo niet dan doorverwijzen. Aangezien het ook om belangrijke gegevens ging vind ik dat het erg onprofessioneel over kwam.	Slow (defer) Not customer friendly
45	Eerste optie geeft het snelste de gewenste info zonder dat je steeds weer vragen gesteld krijgt	Fast (options) Possibly an endless circle (repeat)
46	Omdat ik deze het prettigst ervaar. De laatste met buttons vind ik onpersoonlijk	Not personal (options)
47	Vaak na het geven van een kort antwoord krijg je nog niet de juiste opties. Ik zou dan liever keuzes hebben in wat de opties zijn of doorverbonden worden naar een collega	Wrong answer (options)
48	Bij deze voorbeelden is het na een keer corrigeren hersteld. Het gebeurd vaker dat er meerdere pogingen zijn voordat de chat bot begrijpt wat je bedoeld. De keuze voor meerkeuze of een medewerker is in deze klantvriendelijker.	Wrong answer (options, repeat) Customer-friendly (defer) Possibly an endless circle (repeat)
49	leek mij logisch	
50	Overzichtelijke manier!	Clear
51	Bij optie 1 krijg ik het gevoel echt met een medewerker te communiceren ipv een computer en dat heeft mijn voorkeur	Personal (defer)
52	Deze volgorde is naar mijn idee het fijnste	
53	Dit levert de snelste oplossing voor de klant.	Fast
54	Ik vind met een medewerker chatten fijner dan met een chatbot. Ik voel mij niet serieus genomen als je alleen met een chatbot kan communiceren.	Trust in human employees (defer)

55	Het kiezen uit opties is het makkelijkst voor de gebruiker en de minste moeite. De andere opties zorgen ervoor dat je weer meer moet typen en wachten.	Fast (options)
56	Ik denk dat het goed is om opties te geven zodat de gebruiker snel en makkelijk vooruit kan, de vraag herhalen of een medewerker erbij halen kan soms nog wel lang duren of teveel moeite zijn.	Fast (options) Too much effort (defer, repeat)
57	Persoonlijke inbreng geeft meer vertrouwen	Personal (defer) Trust in human employees (defer)
58	De medewerker probeert eerst zelf een beeld te vormen en als dat allemaal niet lukt, dan doorverwijzen naar een collega	
59	geen specifieke reden	
60	nummer 1: Als de juiste er optie er bij staat is het snel en makkelijk. Mocht dat niet zo zijn dan is het altijd fijn om een medewerker te spreken, maar dat duurt vaak even (nummer 2). Bij herhalen is het vaak zo dat je wellicht het alsnog verkeerd zegt/herhaalt.	Fast (options) Slow (defer) Possibly an endless circle (repeat)
61	Meest natuurlijke versie die in de praktijk zich voordoet.	
62	Fijn om zo direct mogelijk te worden geholpen. (Uit de screenshots blijkt niet hoelang je op Sanne moet wachten maar in mijn ervaring is dat meestal toch enige tijd).	Slow (defer)
63	Live chat medewerker als achtervang is gewenst. Herhaalvraag ook prima. Buttons zijn een zwaktebod	Preference for repeat
64	De bedoeling van de bot is om druk van de klantenservice af te nemen. Als de bot de vraag niet begrijpt en vervolgens direct je doorverwijst naar de persoonlijke klantenservice haal je niet het optimale eruit. Buttons gebruiken limiteert de keuze, want ik denk dat er wel meer mogelijk is dan de 4 keuzes. Tenzij de 4 keuzes zijn gebaseerd op wat de ai haalt uit het bericht wat hij niet volledig begrijpt. De 1e optie lijkt me het beste als er alleen nog bij de tip een voorbeeld komt	Limited choices (options) Preference for repeat
65	Ik vind de laatste het beste omdat deze toch het meest persoonlijk is	Personal
66	Via buttons snelst resultaat, doorverwijzen zekerste resultaat, herhalen irritant	Fast (options) Annoying (repeat) Trust in human employee (defer)
67	De onderste 2 vind ik gelijkwaardig, maar een medewerker overtreft nog steeds alles, zij zijn begripvoller en persoonlijk vind ik het ook een stuk fijner om een echt persoon te spreken	Trust in human employee (defer)
68	Is denk het efficiënste	Efficient

69	Het is fijner om met een echte medewerker te communiceren voor complexere vragen	Complex questions (defer)
70	Het is voor mij de beste volgorde	
71	Ik graag in contact ben met een echt persoon.	Trust in human employees (defer)
72	Ik wil snel geholpen worden en niet nog eens doorverbonden worden als het zo lukt	Fast (repeat, options)
73	2 1 3 Bij 2 persoonlijk contact Bij 1 de buttons maakt de vraag direct Bij 3 je moet dan nadenken om de vraag duidelijker te maken	Personal (defer) Harder to understand (repeat)
74	Opties weer het snelste, doorverbinden is het fijnste, herhalen werkt niet goed	Fast (options)
75	Opties geeft duidelijkheid, hiermee kan de klant duidelijk kiezen. Optie 2 geeft duidelijk aan dat er meer informatie nodig is. Optie 3 komt ongeloofwaardiger over en heeft te veel stappen nodig. Voorkeur voor 1 want minste stappen en het meest duidelijk.	Clear (options) Harder to understand (repeat)
76	Op basis wat ik zelf fijn vind. Vaak kun je vragen ook al vinden om de website. Door de eerste optie kun je vraag specifiek aangeven. Bij optie nummer twee krijg je een echte medewerker aan de lijn, dat vind ik heel erg fijn. Bij de laatste moet je eigenlijk twee keer dezelfde vraag stellen terwijl bij optie 1 en 2 gelijk het antwoord krijgt van de robot zelf of een medewerker	Trust in human employee (defer)
77	stategie 1 is het makkelijst, 2 vergt meer begrijpen, 3 kost tijd	Easy (options) Harder to understand (repeat) Slow (defer)
78	Gemak, duidelijkheid en snelheid	Easy, fast
79	Als de klanten voor een chatbot kiezen, is dat b.v. Omdat het buiten openingstijden is. Er is dan geen medewerker beschikbaar om naar door te verbinden. Dus dat kan beter een laatste redmiddel zijn. Ik vond het vragen om herhaling gewoon ‘elegant’. En het werkt volgens mij heel goed. De klant krijgt een kans om het nog eens te proberen, met een tip, en daarna gaat het gesprek goed. Dus goed opgelost. De knoppen vind ik minder leuk, omdat het toch beperkend is en het is geen gesprek meer op die manier. Bovendien krijg je 4 knoppen, maar hoe weet je of het de goede zijn, of het juiste erbij zit?	Limited choices (options)
80	Doorverwijzing is fijn als de chatbot er zelf na een herkaling van de vraag niet uitkomt	Preference for defer

81	Graag zsm met een echt mens, nu chatbots nog beperkt zijn.	Chatbots are not yet well developed (defer) Preference for defer
82	zo word je sneller geholpen voor jouw specifieke vraag	Fast
83	Een medewerker helpt toch wat sneller	Fast (defer)
84	Menselijk contact is belangrijk voor serviceverlening. Buttons is toch wat onpersoonlijk dus daarom heb ik die als laatste gezet.	Customer-friendly (defer) Not personal (options)
85	Mijn voorkeur	
86	meteen een collega inschakelen kan veel tijd kosten	Slow (defer)
87	Werkt het makkelijkste voor de klant	Easy
88	Ik vind strategie ‘opties’ het beste omdat dit voor de klant de makkelijkste optie zal zijn met snel resultaat. Dan herhalen, je kunt nog steeds door de chatbot geholpen worden. Wel vraag ik mij af of dit voor sommige mensen werkt, of het hun lukt om de vraag daadwerkelijk anders te formuleren. De laatste optie is wat mij betreft ‘een laatste redmiddel.’ Het is fijn dat je doorgestuurd wordt, maar in principe beter geweest als de chatbot het op had kunnen lossen.	Easy, fast (options) Harder to understand (repeat) Possibly an endless circle (repeat)
89	Omdat ik dat in mijn eigen ervaringen als fijn ervaren heb	
90	Gemakkelijkste	Easy
91	Eerst een nieuwe strategie gebruiken om op een open manier de vraag te stellen. Als dat nog niet lukt meteen naar mens, want met die buttons kom je er ook niet	Limited choices (options)
92	Buttons komen op mij het meest betrouwbaar over en meestal staat de oplossing er dan wel tussen. Dat gaat het snelst. Ik vind het ook fair als je na het niet kunnen beantwoorden direct aan een mens wordt gekoppeld, maar dat is dan weer arbeidsintensief en toont dat chatbots nog niet ver genoeg ontwikkeld zijn. Vragen moeten herhalen vind ik minder goed overkomen op de klant en duurt het langst in mijn ogen.	Slow (repeat) Fast (options) Chatbots are not yet well developed (defer)
93	Het meest persoonlijk, toch mis ik een antwoord op het tweede deel vd vraag of het abonnement nu wel of niet voor de gewenste datum opgezegd wordt!	Personal
94	Als hij het niet begrijpt wil ik direct juiste communicatie met een echt persoon. Daarnaast is het daadwerkelijk niet begrijpen irritant en is het heter om de mogelijke opties te geven	Annoying (options)
95	Ik vind het niet erg als de chatbot gewoon de vraag opnieuw stelt. Als deze dan hierna het goede antwoord of hulpt dient.	Preference repeat
96	Ik spreek het liefst een medewerker	Trust in human employee

		(defer)
97	Het snelste antwoord op je vraag	Fast
98	Bij een vraag wil ik een reactie van we gaan u helpen.	
99	Ik vind het best, lijkt mij het snelste werken en mocht met herhaling niet werken zou ik doorgaan op strategie 2 zodat je zeker weten goed geholpen wordt door een medewerker. Ik vind de 3e strategie minder goed, omdat je niet altijd de optie krijgt die je bedoelt.	Fast (repeat) Limited choices (options) Trust in human employees (defer)
100	Gerichtere keuze	
101	Bij opties kan ik gelijk aangeven wat de chatbot voor opties heeft welke die wel kan beantwoorden, dat mis ik bij herhalen. In dit geval ging herhalen goed, maar uit eigen ervaring kan ik spreken dat de chatbot vaak blijft herhalen. Doorverwijzen naar een medewerker vind ik vaak fijn bij complexere vragen die de chatbot sws niet kan beantwoorden	Complex questions (defer) Chatbots are not yet well developed (repeat)
102	Buttons scheelt typen, mits de beoogde optie er daadwerkelijk bij staat. Herhalen van formuleren klinkt onnodig, je hebt de vraag al gesteld dus het kost onnodig veel tijd. Als je wordt doorverwezen naar een medewerker voelt de chatbot nutteloos aan.	Faster (options) Slow (repeat) Useless chatbot (defer)
103	Meteen doorverwijzen als er geen antwoord is. Buttons vind ik soms lastig te kiezen	Limited choices (options)
104	Keuze mogelijkheden kunnen beperkend zijn, ik praat het liefste met een medewerker als chatbot het niet begrijpt	Limited choices (options)
105	Via buttons is het snel en makkelijk om te fixen, aangezien het makkelijk op te lossen is. Wellicht is het probleem dat de bot moet oplossen wel op te lossen als je de vraag anders formuleert. Een medewerker de beste optie, dat blijft toch het makkelijkst. (Voornamelijk voor ouderen die het niet zo heel snel snappen). Aan de andere kant zorgt dit laatste voor hogere kosten die uiteindelijk wel bij de klant terecht móeten komen.	Fast, easy (options) Easy (defer)
106	Persoonlijk maken van chat box	Personal
107	De eerste twee zijn voor mij redelijk gelijk, maar het blijft fijner om met een echt persoon te chatten als deze optie er is. Opties heb ik als laatste geplaatst omdat de kans er is dat jouw vraag er niet tussen staat	Limited choices (options)
108	De buttons kunnen helpen om meteen duidelijk te hebben waar het over gaat. Het is alleen een ellende als de specifieke button er niet bij staat. Dan is een doorverwijzing medewerker als button heel praktisch. De derde strategie slaat meestal nergens op	Limited choices (options)

109	Direct contact is altijd beter, buttons geven opties en hoef je niet te raden. Herhalen helpt weinig.	Limited choices (options)
110	Vaka vastloopt tegen het feit dat ene xhatbot mijn vraag niet snapt en er geen oplossing gevonden kan worden.	Possibly an endless circle (repeat)
111	De eerste optie is het duidelijkst. Daarna kan altijd nog 2 opties gegeven worden. De tweede optie voelt omslachtig en inefficiënt. De derde optie voelt ongeloofwaardig: waarom is contact met een medewerker nodig bij een simpele vraag?	Not efficient
112	Veel kan doende chatbox beantwoord worden, en is prima op deze manier. Wel fijn om uiteindelijk doorverbinden te worden als de vraag onbekend is.	Preference for repeat or options. Complex question (defer)
113	Dit werkt toch het snelst en is het meest duidelijk. Vaak kan de chatbot je uiteindelijk wel helpen en dat is sneller dan de echte medewerker tijd heeft	Fast (options) Slow (defer)
114	Fijn om met een persoon te communiceren	Personal (defer)
115	Buttons zorgen ervoor dat de klant makkelijker snapt wat de opties zijn	Easy (options)
116	Liefst een directe schakel naar een echte persoon wanneer er iets misgaat, het draait om efficiëntie	Fast (defer)
117	Nuttigst gebruik	
119	Ik spreek liever met een medewerker om er zeker van te zijn dat het goed is	Trust in human employee (defer)
120	Voor mij de meest efficiënte manier om te communiceren en te bereiken wat ik wil weten of gedaan krijgen	Efficient
121	Bij de derde heb ik vaak een vraag die er niet bij staat. Een persoon spreken is dan het fijnst	Limited choices (options)
122	De eerste oplossing is praktisch en snel. Bij de tweede oplossing krijg je ook een goede oplossing door met een medewerker te spreken, maar dit is minder snel. Bij de laatste strategie is er een kans dat de chatbot het nog steeds niet begrijpt.	Fast (options) Slow (defer)
123	Herhalen is goed, als het niet meer dan 1 keer hoeft Doorverwijzen werkt goed als je vraag heel specifiek is, voor algemene of gangbare vragen hoeft dat niet voor mij Opties werkt alleen als alle opties opgesomd worden en als jouw optie er nou net niet tussen zit, watoet je dan!	Complex question (defer) Limited choices (options)
124	Strategie 1 is het meest klantvriendelijk en laat de deugdelijkheid van de chatbot zien. Strategie 2 laat de zwaktes zien van de chatbot maar is alsnog werkzaam. Strategie 3 laat zien dat de chatbot eigenlijk niet werkt en veroorzaakt 'dubbel' werk de klant.	Customer-friendly (options) (defer) Chatbots are not yet well developed (repeat)

125	Persoonlijker	Personal
126	De buttons geven in iedergeval het idee dat de bot enigzins in de juiste hoek zit. Een medewerker vind ik zelf het fijnst maar dat gaat het doel van de bot voorbij. Zeker bij zo;n simpele vraag als in het voorbeeld. Herhalen is irritant, dat werkt vaak niet.	Annoying (repeat) Preference for defer
127	Strategie met de opties leek me in eerste instantie het beste maar alleen als de optie er in staat. Misschien een idee om bij openen van chatbot de standaard opties vooraf te melden. Dus voor mij is als een chatbot het niet direct door heeft waar ik het over heb een medewerker aan de 'lijn' het beste.	Limited choices (options) Trust in human employees (defer)
128	Voor standaardvragen is het niet altijd nodig om te wachten op een medewerker. Een AI bot kan dat prima uitleggen	Slow (defer)
129	De buttons maken het gemakkelijker om een juiste keuze te maken. Daarna is het doorverbinden met een medewerker ook fijn als je er niet uit komt met een chatbot, dan heb je alsnog 'persoonlijk' contact met een medewerker van de klantenservice. Herhalen staat op 3 omdat ik meer vertrouwen heb in opties en doorverwijzen naar een medewerker.	Easy (options) Personal (defer) Trust in human employees (defer)