

Understanding the User Experience of Customer Service Chatbots: What Can We Learn from Customer Satisfaction Surveys?

Knut Kvale¹[0000-0001-5453-8031], Eleonora Freddi¹[0000-0002-1794-5188],
Stig Hodnebrog²[0000-0001-7446-3575], Olav Alexander Sell²[0000-0002-8546-8934],
and Asbjørn Følstad³[0000-0003-2763-0996]

¹Telenor Research, Norway

²Telenor Norway, Norway

³SINTEF, Norway

knut.kvale@telenor.com

Abstract. Understanding and improving user experience is key to strengthening uptake and realizing the potential of chatbots for customer service. In this paper, we investigate customer satisfaction surveys as a source of insight into such user experience. A total of 5,687 customer satisfaction reports on users' interactions with a customer service chatbot, and the corresponding chatbot interactions, are analyzed. The findings demonstrate that customer satisfaction reports are closely associated with the degree to which the problems motivating users' chatbot interactions are resolved. Furthermore, the findings show substantial variation in the performance of different chatbot intents in terms of customer satisfaction and problem resolution. This implies that user experience varies substantially depending on the problems motivating users to interact with the chatbot. Finally, we identify key characteristics of the intents associated with particularly high or low customer experience, suggesting paths towards efficient improvement of chatbot user experience. Based on the findings, we point to key implications for theory and practice and suggest directions for future research.

Keywords: Chatbot · User Experience · Customer Satisfaction

1 Introduction

Customer service is a major chatbot application domain. Chatbots may be low-threshold channels for information and support, serving as a cost-effective and accessible supplement to manual customer service. In a recent report, Gartner [13] predicts that about one-third of the surveyed companies deploy or have near-future plans of deploying conversational platforms in customer service, and that such deployment will substantially increase operational efficiency. However, the same report also predicts that there will be substantial turnover in such conversational platforms, and that a number

of older chatbot applications will be abandoned. This suggests that, while potential benefit of chatbots for customer service is high, there is substantial risk of failure.

User experience is a key determinant of successful implementation of chatbots for customer service [5]. While an increasing number of service providers are offering customer service through chatbots, user uptake has been found to lag behind [12]. In response, substantial research and practitioner effort have been spent on strengthening user experience for chatbots in general [24], and for customer service in particular [19, 28]. Research and practitioner guidelines suggest an at times bewildering range of potential drivers of chatbot user experience, but at the same time, there seems to be a shortage of advice for how to strengthen it. Hence, service providers might find it challenging to prioritize and guide chatbot development and improvement initiatives.

In this context, we present a study that provides insights from a systematic application of customer satisfaction surveys to understand chatbot user experience. We asked a sample of users of a chatbot for customer service ($N = 5,687$) to report (a) their satisfaction with the chatbot interaction and (b) the degree to which the problem motivating their interaction had been resolved. The survey data was analyzed with regard to information on the actual user conversations, specifically on the chatbot intents triggered as part of the interaction.

The findings strongly suggest that problem resolution is highly important for a better user experience in chatbots for customer service. Furthermore, the findings indicate that the user experience depends on the quality of the support provided through the specific intents triggered in an interaction. Accordingly, a user's perception of the chatbot is shaped by the type of problem he or she want to get resolved. In particular, customers were found to have a more positive experience when the chatbot provided concrete and detailed support, guiding the user closer to problem resolution, for example, by drawing on integrations with back-end systems for user information and transactions on behalf of the user.

The findings contribute to the current body of knowledge on chatbot user experience, demonstrating the importance of efficient and effective problem resolution. They also provide implications for practice, suggesting how customer satisfaction data may be used to improve the user experience in existing chatbots for customer service.

2 Background

In this section, we present the background concerning chatbots for customer service, chatbot user experience, and – finally – customer satisfaction, its relation to user experience, and how customer satisfaction surveys are used in research and industry to assess and improve service quality.

2.1 Chatbots for Customer Service

Chatbots are increasingly important for customer service. A recent CapGemini report [32] finds that, in retail banking and insurance, 49% of the top 100 organizations employ chatbots. In consumer products and retail, the corresponding number is 23%. In response to the

current interest, there is a wide range of available chatbot platforms for customer service [13], such as IBM Watson, IPsoft, Microsoft Bot Framework, Nuance, and boost.ai.

The increased interest in chatbots over the last few years is motivated by users and service providers' uptake of chat platforms, as well as by improved language processing capabilities due to advances in artificial intelligence and machine learning [7]. In particular, enhanced access and availability of chat-based customer service, as well as cost-efficiency and upgraded user experiences, have increased service providers' engagement in chatbot for customer service [27]. Users are reported to appreciate the opportunity for immediate and low-threshold access to support provided by chatbots for customer service [6]. However, chatbots for customer service have also been found to entail challenges concerning, for example, the ability to correctly predict users' intents and a lack of capabilities for handling complex requests [11].

Key components of chatbots for customer service include language processing and intent handling, exception handling, context awareness, analytics, and integration with back-end systems [13]. Users typically enter their requests as free text messages, often in conjunction with the option of selecting predefined answer alternatives through buttons or quick replies. The user requests that the chatbot may recognize and respond to are typically structured in intent hierarchies, where language processing techniques and machine learning models are employed to predict user intents on the basis of the users' free text input [30]. As soon as an intent is correctly predicted, the chatbot often provides the user with navigation support to move within the intent hierarchy. For example, if a chatbot predicts user requests corresponding to an intent on invoicing, the chatbot may ask users if they want to see an invoice, ask about a particular invoice, or pay an invoice. These options are meant to facilitate easy navigation in the intent hierarchy.

Chatbots for customer service typically need to support a large number of request types. For example, chatbots in the banking sector may require several thousand intents to cover a sufficient breadth of user requests [9]. A larger number of intents widens chatbot coverage, but also increases the potential for false positives, that is, responses which do not adequately address the user request [10]. Establishing and maintaining the needed intent hierarchies in chatbots for customer service is demanding [19]. The key parts of this work are to identify needed intents, create content and actions associated with intents, and train machine learning models to improve intent prediction.

2.2 Chatbot User Experience

Following the surge of interest in chatbots, chatbot user experience has recently become a topic and focus of substantial research. User experience is a complex construct, including factors such as pragmatic quality, hedonic quality, aesthetic appeal, and goodness [20]. In the international standard of human-centered design, ISO 9241-210 [18], user experience is defined as users' perceptions and responses from use and anticipated use including "emotions, beliefs, preferences, perceptions, physical and psychological responses, behaviours and accomplishments" (p. 3).

Because of the complexity of the construct, user experience is investigated from a wide range of perspectives and in consideration to a large number of determinants. This holds both for chatbots in general and for chatbots for customer service in particular. General

chatbot user experience has been, for example, studied from the perspectives of pragmatic and hedonic quality [8], social presence [21] and anthropomorphism [23]. Furthermore, chatbot user experience has been investigated as determined by chatbot personality [31], chatbot gender [22], human likeness [3], and conversational design [14].

User experience of customer service has been shown to depend on effectiveness and efficiency in problem resolution [5] but also on factors such as expectation management [26], courtesy and friendliness [16], and responding adequately to the customer's mood or tone of voice [17]. This varied set of factors affecting user experience is reflected also in research on chatbots for customer service. Here, user experience has been studied in relation to factors such as trust [25], user emotion and rapport [28], anthropomorphism, and social presence [1]. User experience for customer service chatbots has been found to be determined by such factors as types of dialogue scripts [28] and conversational repair [10], as well as information quality, system quality, and service quality [34]. However, industry reports suggest that users' main reasons for not using chatbots for customer service include chatbots not being sufficiently skilled for complex requests and challenges concerning interpretation and goal achievement [6, 11]. This may imply that the user experience of chatbots for customer service is likely to be strongly determined by the chatbot's ability to provide accessible help. Likewise, motivations for using chatbots in general have been found to be strongly determined by pragmatic motivations [2].

2.3 Customer Satisfaction, User Experience, and Service Improvement

Given the broad variety of user experience perspectives and measurements available, it will be important for service providers to identify those that are of particular relevance for continuous improvement of chatbot implementations [19].

To provide insight into users' perceptions of service interactions, a much used measurement in the service industries is customer satisfaction surveys [4]. In such surveys, users report their satisfaction with a brand, a service, or particular service interactions. Temkin [33] finds that 85% of surveyed companies measure customer satisfaction at the level of interactions. Customer satisfaction reports linked to particular service interactions, while not direct measures of the entire user experience construct, are likely to reflect the overall perceived quality of the user experience. That is, low customer satisfaction scores may indicate poor user experience, while high customer satisfaction scores indicate good user experience. However, satisfaction scores alone cannot provide insight into the details of the user experience [29].

Customer satisfaction has been shown to substantially covary with company performance and with other customer feedback measures [4]. Previously, customer satisfaction was typically measured with multi-item instruments, but studies such as that of Van Doorn et al. [35] have shown that single-item satisfaction measures have similar performance to that of multi-item measurements.

Customer satisfaction surveys also have a substantial impact more generally on service organizations, and in particular on service improvement. For transactional interactions, customer satisfaction surveys have been found to outperform other experience-oriented measurements in terms of organizational impact [33].

3 Research Questions

Motivated by the current state of the art, we acknowledge the need to strengthen insight into the key determinants of user experience in chatbots for customer service. The prominence of customer satisfaction assessment in the service literature suggests that this may be a viable approach towards such insight. Hence, our research objective is to investigate what we can learn about chatbot user experience from customer satisfaction surveys. Drawing on the presented background, three research questions are of particular relevance:

- RQ1.** How is the user experience of chatbots for customer service affected by whether or not users' problems are resolved?
- RQ2.** How is the user experience of chatbots for customer service affected by the kind of problem for which users seek help?
- RQ3.** What characterizes intents associated with positive and negative user experiences in chatbots for customer service?

Through RQ1, we investigate the degree to which problem resolution impacts the user experience in chatbots for customer service. While a broad range of drivers may impact user experience in customer service chatbots, the literature suggests that problem resolution is likely to have a substantial impact [6, 11]. Furthermore, it is relevant to investigate potential variation in user experience depending on the character of the problem and how problem resolution is implemented through related chatbot intents (RQ2). Finally, given such variation in user experience across different user problems, it is important to explore the characteristics of the problems and associated intents to identify potential underlying causes of this variation (RQ3).

4 Method

To address the research questions, we conducted a study analyzing data on customer satisfaction, problem resolution, and customer interaction with a running chatbot for customer service. In our analysis addressing RQ1, we compared customer satisfaction scores for different levels of problem resolution in the chatbot interaction. To shed light on RQ2, we compared customer satisfaction scores for chatbot interactions with different intents triggered. Investigating RQ3, we analyzed and compared the characteristics of chatbot interactions associated with high customer satisfaction scores with interactions associated with low customer satisfaction scores.

In the following, we first present the chatbot for which the customer satisfaction data were gathered as well as the different data sets included in the analysis. We then detail the analyses conducted in response to the three research questions respectively.

4.1 The Customer Service Chatbot

This study was conducted by analyzing data gathered from *Telmi*, a chatbot employed by the international telecom provider Telenor. The chatbot complements the company's customer service in its Norwegian operation. *Telmi* is provided as a separate channel

for customer service, in addition to self-service on the Telenor customer website and smartphone app, phone-based support, and assistance through social media platforms such as Facebook and Twitter.

The chatbot provides help and information in response to user free text input. The user intents are predicted on the basis of language processing through machine learning models. The chatbot is capable of predicting more than 2,700 intents structured in intent hierarchies. When an intent is identified, the dialogue can evolve by the user choosing among buttons with predefined answer alternatives or by interpreting further free text input within the context of the previous intents.

The chatbot is offered to users who log into the customer website as well as to anonymous users. It can help users by answering frequently asked questions, but it can also provide personalized information and conduct transactions associated with tasks, such as getting PUK codes, ordering and activating a SIM card, getting details on data spending, ordering extra data packages, and blocking a subscription. The chatbot was implemented in the first half of 2019 and has since been in continuous improvement and development. Intents have been added and updated; training of machine learning models has been improved, and support for more personalization has been provided through application programming interfaces (APIs) to back-end business systems.

4.2 The Data Sets

Customer Satisfaction Survey. To assess and improve the chatbot, feedback from users was gathered through a customer satisfaction survey. All users interacting with the chatbot received invitations to respond to the survey provided that (a) the user was logged into a customer website or application when using the chatbot, and (b) the user had not responded to similar surveys during the last three months. In the period February 26–May 31, 2020, $N = 5,687$ users responded to the survey, constituting a response rate of 18% of those receiving the invitation. In total, 14,8381 customers used Telmi in the analyses period, either on the open web or as logged in customers. Hence, the 5,687 customers only represented 4% of the total number of conversations.

In the customer satisfaction survey, users reported on their recent interaction with the chatbot. Primarily, they were asked to report their “satisfaction with the interaction with Telmi” on a five-point scale (1 = very dissatisfied; 5 = very satisfied). Furthermore, the customers were asked to disclose the degree to which their problem was resolved at the time of responding to the survey (fully resolved, partially resolved, not resolved). In addition, the survey contained questions about factors that are not treated in this study, such as how easy or difficult the users found the problem resolution, details on failure to resolve problems, expectations of the chatbot, and whether they had sought to solve the problem through other means of support.

Chatbot Interaction Data. Data from the users’ interactions with the chatbots are stored in a way that protects the users’ privacy, while allowing to combine the interaction data with the satisfaction survey data. Here, data which may contain personal in-

formation are handled in accordance with the European Union's General Data Protection Regulation (GDPR) and deleted within the defined retentions schedule. However, predicted intents, chatbot responses, and timestamps are stored.

Thus, for each interaction, it is possible to include in the analysis the time of the interaction, the number of triggered intents, the specific triggered intents, and the chatbot's associated responses. These data provide valuable information for training and improvement of the chatbot, as it is possible to identify interactions leading or not leading to relevant support, as well as to analyze the different intents or combinations of intents with regard to user reports of satisfaction and problem resolution.

4.3 The Analyses

On the basis of the combined data sets for the users' self-reports (the customer satisfaction survey) and the users' interactions (chatbot interaction data), we conducted analyses to investigate the three research questions.

Impact of Problem Resolution on User Experience (RQ1). In line with Temkin [33], user experience was operationalized as customer satisfaction. Clearly, customer satisfaction is not a direct measure of the entire user experience construct – as this encompasses, for example, users' emotional responses – but it is to be seen as reflecting user experience. Therefore, customer satisfaction will be a useful proxy for investigating whether the user experience was positive or negative. In particular, we compared customer satisfaction for users with different levels of self-reported problem resolution (fully resolved, partially resolved, unresolved). Comparisons among the levels of problem resolution were conducted as descriptive analyses and one-way analyses of variance (ANOVA).

Impact of the Kind of Problem for Which Users Seek Help (RQ2). The problem for which users seek help was operationalized as the specific user intent predicted by the chatbot based on users' free text requests. All intents with more than 40 observations in the data set were included, except for intents concerning social interaction (e.g., greetings and pleasantries), requests for escalation to human support, error recovery (e.g., fallback responses to user requests without a certain intent prediction), and two intents that had recently been reworked. An observation refers to a single interaction between a user and the chatbot. The included intents were compared on the basis of customer satisfaction scores. Comparisons were conducted as descriptive analyses and one-way ANOVA.

Characteristics of Intents Associated with Positive and Negative User Experience (RQ3). The most frequently predicted intents were sorted into groups according to the level of customer satisfaction. The intents associated with particularly low or high levels of satisfaction were investigated with regard to their common characteristics. The investigation was conducted in an open-ended analysis by three of the AI trainers working with improving and maintaining the chatbot.

5 Results

In this section, we present the findings associated with each of the three research questions consecutively.

5.1 Problem Resolution Impacts User Experience (RQ1)

To investigate the impact of problem resolution on user experience, we compared customer satisfaction scores for chatbot interactions associated with different levels of self-reported problem resolution. The findings suggest that problem resolution strongly impacts user experience. Among users responding that the problem had been fully resolved, 97% reported positive customer satisfaction (score 4 or 5 – satisfied or very satisfied). However, among users reporting the problem to be unresolved, 4% gave a positive customer satisfaction score. Among users disclosing partial problem resolution, 55% reported positive customer satisfaction. Details are provided in **Fig. 1**.

One-way ANOVA with self-reported problem resolution as an independent variable and customer satisfaction as a dependent variable showed significant differences among the groups with no problem resolution ($M = 1.66$, $SD = 0.87$), partial problem resolution ($M = 3.52$, $SD = 0.90$), and full problem resolution ($M = 4.55$, $SD = 0.63$), ($F(2, 13310) = 10951$, $p < 0.000$). The effect size of the differences was large ($\omega^2 = 0.62$).

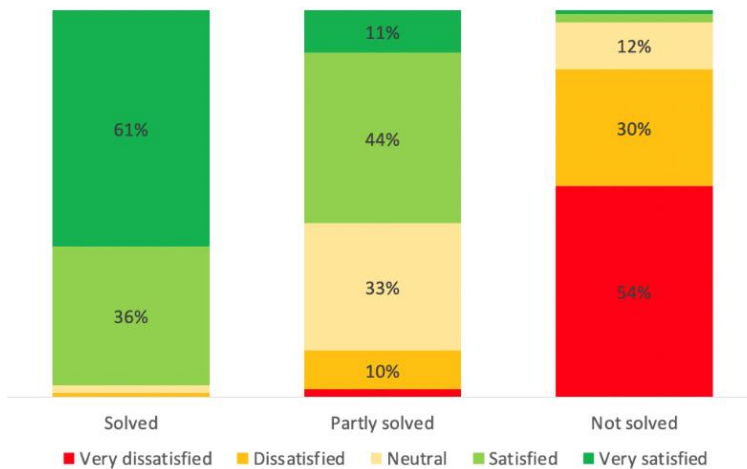


Fig. 1. Distribution of customer satisfaction scores for the three groups reflecting full, partial, and no problem resolution

5.2 User Experience Varies Substantially for Different Kinds of Problems (RQ2)

To investigate the impact of the users' type of problem on user experience, the different kinds of problems were operationalized as the distinct predicted intents in the chatbot

– where each intent corresponds to a particular problem area and level of detail – and user experience was operationalized as customer satisfaction.

The analysis was run on the 22 most frequently observed intents. Customer satisfaction was found to vary substantially among the different intents, as shown in **Table 1**. Customer service scores are provided for the five high-scoring intents and the five low-scoring intents. In this table, we also include details on the proportion of users’ self-reported problem resolution.

To investigate the significance of the variation in the customer service scores, a one-way ANOVA was conducted with the four most frequently observed intents (Problems internet, Invoice question, Update smart card, TV) as independent variables and customer satisfaction score as the dependent variable ($F(3, 878) = 20.81, p < 0.000$).

Table 1. Examples of most frequently observed chatbot intents with corresponding customer satisfaction scores and proportion of users reporting problem resolution. The examples include the five high-scoring and the five low-scoring intents out of the 22 intents in the analysis.

Predicted intent	% reporting problem resolved	Customer satisfaction score		% positive (Customer satisfaction 4 or 5)
		Mean	SD	
Forgot PIN code	65%	3.65	1.49	70%
PUK code	42%	3.56	1.51	62%
Cancel subscription	44%	3.14	1.61	56%
Activate SIM card	49%	3.36	1.71	53%
Update smart card TV	40%	3.28	1.54	52%
Help with router	3%	2.03	1.12	13%
Invoice question	5%	1.96	1.09	11%
Help with e-mail	0%	1.7	0.99	10%
Help with TV decoder	2%	1.62	1.03	9%
Error in invoice	1%	1.89	1.02	7%

5.3 Varying Characteristics for Intents Associated with Positive and Negative User Experience (RQ3)

To explore possible systematic variations in the characteristics of intents associated with positive and negative user experience respectively, intents with particularly high and low scores were investigated by three chatbot AI trainers.

As a starting point for this investigation, it was noted that high-scoring intents were also those for which the highest proportion of users reported the problem to be fully resolved. Moreover, it was noted that high-scoring intents were those that cover specific and concrete issues with a simple-to-understand answer that actually solves the problem. For example: If a customer asks “Please update my smart card” and the chatbot answers “Sure. Press confirm to update.”

For low-scoring intents, it was found that intents typically aim to deal with too many situations, thereby not being sufficiently concrete and specific. One example is “Problem with internet,” which covers a wide range of possible issues and has multiple possible directions intended to support the user. The risk is that the user does not always understand the nature of the problem and, therefore, often ends up in an information loop.

The scores for the remaining 12 intents cover the entire range of values from high-scoring to low-scoring intents and so it is harder to draw insightful conclusions across these – hence, our analysis is limited to the high and low scoring intents.

In conclusion, intents that provide concrete and relevant help bringing the user closer to problem resolution will generate a good user experience, whereas generic responses to issues that the user does not fully grasp will likely generate a poor user experience.

6 Discussion

The presented study provides insight in response to gaps in the current knowledge, as well as implications of theoretical and practical relevance. We first discuss the findings for each research question relative to the state of the art. We then detail key implications of the findings, and finally, we address limitations of our study and suggest venues for future research.

6.1 User Experience Insights from a Customer Satisfaction Survey

The analysis of customer satisfaction data, combined with data on problem resolution and the evaluation of the characteristics of the most used intents in the chatbot, provides relevant insights into the three research questions.

Problem Resolution Is Likely a Key Determinant of User Experience in Chatbots for Customer Service (RQ1). While user experience is certainly a broader construct than customer satisfaction, measures of satisfaction may indicate whether or not a service interaction has generated a positive or negative user experience. Our finding that problem resolution is strongly associated with customer satisfaction suggests that solving the user’s case is a highly important determinant of user experience. Previous research has indicated that a broad range of factors, such as trust, user emotion, and anthropomorphism, may impact user experience of chatbots for customer service [1, 25, 28]. Nevertheless, our findings concerning the impact of problem resolution resonate with the previous research on general chatbots [2], where it has been shown that most users have pragmatic motivations for utilizing chatbots – that is, chatbots are typically used because they are seen as an easy and convenient means for users to achieve their goals. Furthermore, chatbots’ ability to provide support, training or help has been found to be highly important for user experience for general chatbots [8]. Thus, it is not surprising that problem resolution may strongly predict user experience. Our findings also shed light on this relationship in chatbots for customer service. Moreover, we find the strength of the relation between problem resolution and user experience to be noteworthy. In particular, our results highlight that nearly all users with their

problem resolved reported to be satisfied, whereas nearly no customers with unsolved problem gave positive scores. These findings suggest to service providers that whether or not a chatbot for customer service actually resolves a problem – or contributes to it being solved – may be decisive for user experience.

User Experience Varies Depending on the Users' Specific Problems (RQ2). Chatbots differ from other interactive systems, such as customer websites, by providing less information with regard to their capabilities and the opportunities they offer for the user [30]. Because of this, the limited interaction between the user and the chatbot will likely be decisive for the user experience [15]. For example, in the case of a service inquiry, the content and user value provided through the few messages exchanged between the chatbot and the user during the interaction could play a key role in the overall experience. Due to the limited content offered to all users in a chatbot – as opposed to website interaction where broader ranges of content are presented to all users – different users may experience the chatbot very differently. Our study findings clearly suggest the importance of the users' problem for the user experience in the chatbot. For some problems, associated with one set of user intents, customer satisfaction scores are consistently high. For other issues, associated with other user intents, customer satisfaction is consistently low. Based on these results, it seems fair to assume that the overall design of the chatbot is less important to user experience than the actual support the user receives when triggering a particular intent. This is a highly interesting finding for service providers as it proposes a concrete means to improve user experience – by prioritizing improvements for intents with relatively low customer satisfaction scores.

Specificity in Problem Resolution May Determine the User Experience Associated with Certain Intents (RQ3). The AI trainer analysis of the intents associated with particularly high or low customer satisfaction scores further served to detail the insight that problem resolution is of key importance to user experience in customer service chatbots. It is noteworthy that intents associated with high customer satisfaction scores are those that provide concrete and direct support for the user towards problem resolution. Typically, such direct assistance is due to the chatbot drawing on information about the user from back-end systems or conducting transactions for the users directly following the user's request. Conversely, intents associated with low customer satisfaction scores were typically either presenting too generic information, not sufficiently adapting to the user's situation, or supporting problem areas that are inherently complex and bewildering to users. These findings are useful both to understand variation in user experience for one and the same chatbot and to guide practical upgrade of chatbots. Moreover, it highlights the kind of improvement efforts that are most likely to lead to an enhanced user experience.

6.2 Implications for Theory and Practice

Our study findings hold several implications for theory and practice. We see the following as having particular relevance:

Implications for Theory. The findings serve to extend and strengthen current theory on chatbots to specifically chatbots for customer service.

1. **The primacy of pragmatic quality.** In the theory of user experience, addressing both pragmatic and hedonic quality is seen as equally salient [20]. However, while hedonic quality – for example, in the form of stimulation and identity – is important to chatbots for customer service, pragmatic quality – in particular, goal achievement – is of primary importance in state of the art for chatbots.
2. **User experience as a consequence of the particular interaction.** Whereas user experience assessments are often seen as concerning the entire application or user interface [14], it may be relevant to consider user experience on a task level for customer service chatbots. This bears resemblance to the interaction effects between the task and chatbot personality found in previous research [22].

Implications for Practice. The findings also hold important implications for practical design and implementation of chatbots for customer service.

1. **Targeted improvement of chatbot intents.** Customer satisfaction surveys have substantial potential in guiding chatbot improvement and maintenance. Specifically, identifying and reworking intents that are widely used but score low on satisfaction will be important to address.
2. **Prioritization of problem resolution.** To strengthen user experience in customer service chatbots, it will be beneficial to prioritize helping users to take the concrete steps needed towards problem resolution rather than providing general information. It is likely that offering personalized help and support by drawing on data from back-end systems will be required in this regard.

6.3 Limitations and Future Work

The presented study, while providing valuable insights into user experience for customer service chatbots, also has important limitations. First, the study focuses on one chatbot within a specific market. While we assume that the findings may be generalizable also to chatbots for customer service in other markets, it is advisable to test this assumption. Hence, for future work, we foresee studies including a broader range of chatbots and markets.

Second, the study is limited to relying on customer satisfaction as the only measurement of customer experience. While valuable as a reflection of good or poor user experience, this measurement does not provide needed nuance for the customer experience construct. Thus, for future work, we anticipate studies combining customer satisfaction surveys with other measurements of relevance for user experience. Furthermore, we foresee qualitative user experience studies to gain insight into how different levels of problem resolution and different intents impact user experience.

Finally, the study is limited in that it does not include satisfaction measures for different channels. Such additional measures would have enabled cross-channel comparison and, thereby, an assessment of the relative satisfaction with the chatbot channel –

something that would be relevant both for theory and practice. Cross-channel comparison clearly is a relevant topic for future work.

In spite of the limitations, the study encourages further use of customer satisfaction surveys as a means to understand user experience in chatbots for customer service. Such surveys may provide new theoretical insight and will also have benefits for practice given the widespread uptake of customer satisfaction surveys within service providers.

Acknowledgements. The work of the four first authors was supported by Telenor Research and Telenor Norway. The work of the fifth author was supported by the Research Council of Norway through research grant no. 270940.

References

1. Araujo, T.: Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Comput. Human Behav.* **85**, 183–189 (2018). doi: 10.1016/j.chb.2018.03.051
2. Brandtzaeg, P.B., Følstad, A.: Why people use chatbots. In: *International Conference on Internet Science*, pp. 377–392. Springer, Cham (2017)
3. Ciechanowski, L., Przegalinska, A., Magnuski, M., Gloor, P.: In the shades of the uncanny valley: An experimental study of human–chatbot interaction. *Future Gener. Comput. Syst.* **92**, 539–548 (2019)
4. De Haan, E., Verhoef, P.C., Wiesel, T.: The predictive ability of different customer feedback metrics for retention. *Int. J. Res. Mark.* **32**(2), 195–206 (2015)
5. Dixon, M., Freeman, K., Toman, N.: Stop trying to delight your customers. *Harv. Bus. Rev.* **88**(7/8), 116–122 (2010)
6. Drift: The 2018 State of Chatbots Report. Technical Report, Drift (2018), <https://www.drift.com/blog/Chatbots-report/>
7. Følstad, A., Brandtzaeg, P.B.: Chatbots and the new world of HCI. *Interactions.* **24**(4), 38–42 (2017)
8. Følstad, A., Brandtzaeg, P.B.: Users’ experiences with chatbots: findings from a questionnaire study. *Qual. User Exp.* **5**, 1–14 (2020)
9. Følstad, A., Skjuve, M.: Chatbots for customer service: user experience and motivation. In: *Proceedings of the 1st International Conference on Conversational User Interfaces*, Paper no. 1. ACM, New York (2019)
10. Følstad, A., Taylor, C.: Conversational repair in chatbots for customer service: The effect of expressing uncertainty and suggesting alternatives. In: *Proceedings of CONVERSATIONS 2019*, pp. 201–214. Springer, Cham (2019)
11. Forrester Consulting: Human vs. machines: how to stop your virtual agent from lagging behind. Technical report, Forrester Consulting (2017). <https://www.amdocs.com/blog/place-digital-talks-intelligent-minds/aia-humans-vs-machines-how-to-stop-your-chatbot-from-lagging-behind>
12. Forrester: The six factors that separate hype from hope in your conversational AI journey. Technical report, Forrester (2018). <https://www.forrester.com/report/The+Six+Factors+That+Separate+Hype+From+Hope+In+Your+Conversational+AI+Journey/-/E-RES143773>
13. Gartner: Market guide for virtual customer assistants. Technical report, Gartner (2019) <https://www.gartner.com/en/documents/3947357/market-guide-for-virtual-customer-assistants>

14. Go, E., Sundar, S.S.: Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. *Comput. Human Behav.* **97**, 304–316 (2019)
15. Hall, E. *Conversational Design. A Book Apart*, New York (2018)
16. Hocutt, M.A., Bowers, M.R., Donavan, D.T.: The art of service recovery: fact or fiction? *J. Serv. Mark.* **20**(3), 199–207 (2006)
17. Hu, T. et al.: Touch your heart: A tone-aware chatbot for customer care on social media. In: *Proceedings of CHI 2018*, paper no. 415. ACM, New York (2018)
18. ISO: Ergonomics of human–system interaction — Part 210: Human-centred design for interactive systems. International Standard. ISO, Geneva (2010)
19. Kvale, K., Sell, O.A., Hodnebrog, S., Følstad, A.: Improving conversations: lessons learnt from manual analysis of chatbot dialogues. In: *Proceedings of CONVERSATIONS 2019*, pp. 187–200. Springer, Cham (2019)
20. Law, E.L.C., Van Schaik, P.: Modelling user experience—an agenda for research and practice. *Interact. Comput.* **22**(5), 313–322 (2010)
21. Luria, M., Reig, S., Tan, X.Z., Steinfeld, A., Forlizzi, J., Zimmerman, J.: Re-embodiment and co-embodiment: exploration of social presence for robots and conversational agents. In: *Proceedings of the 2019 on Designing Interactive Systems Conference*, pp. 633–644 (2019)
22. McDonnell, M., Baxter, D.: Chatbots and gender stereotyping. *Interact. Comput.* **31**(2), 116–121 (2019)
23. Morana, S., Gnewuch, U., Jung, D., Granig, C.: The effect of anthropomorphism on investment decision-making with robo-advisor chatbots. In *Proceedings of ECIS 2020* (2020)
24. Moore, R.J., Arar, R.: *Conversational UX Design: A Practitioner’s Guide to the Natural Conversation Framework*. ACM, New York (2019)
25. Nordheim, C.B., Følstad, A., Bjørkli, C.A.: An initial model of trust in chatbots for customer service—findings from a questionnaire study. *Interact. Comput.* **31**(3), 317–335 (2019)
26. Palmer, A.: Customer experience management: a critical review of an emerging idea. *J. Serv. Mark.* **24**(3), 196–208 (2010)
27. PwC. Bot.Me: a revolutionary partnership. How AI is pushing man and machine closer together. *Consumer Intelligence Series*, PwC (2018). https://www.pwc.com/it/it/publications/assets/docs/PwC_botme-booklet.pdf
28. Sands, S., Ferraro, C., Campbell, C., Tsao, H.Y.: Managing the human–chatbot divide: how service scripts influence service experience. *J. Serv. Manag.* (2020). <https://doi.org/10.1108/JOSM-06-2019-0203>
29. Schaffer, E., Lahiri, A.: *Institutionalization of UX: A Step-by-Step Guide to a User Experience Practice* (2nd ed.). Upper Saddle River, NJ: Addison-Wesley (2013)
30. Shevat, A.: *Designing bots: creating conversational experiences*. O’Reilly Media, Sebastopol, CA (2017)
31. Smestad, T.L., Volden, F.: Chatbot personalities matters. In: *International Conference on Internet Science*, pp. 170–181. Springer, Cham (2018)
32. Taylor, M.P. et al.: Smart talk: How organizations and consumers are embracing voice and chat assistants. Technical report, Capgemini SE (2019). https://www.capgemini.com/wp-content/uploads/2019/09/Report—Conversational-Interfaces_Web-Final.pdf
33. Temkin, B. (2017). The state of CX metrics, 2017. Technical report, Qualtrics XM Institute. <https://www.qualtrics.com/xm-institute/state-of-customer-experience-metrics-2017/>
34. Trivedi, J.: Examining the customer experience of using banking chatbots and its impact on brand love: the moderating role of perceived risk. *J. Internet Commer.* **18**(1), 91–111 (2019)
35. Van Doorn, J., Leeflang, P.S., Tijs, M.: Satisfaction as a predictor of future performance: a replication. *Int. J. Res. Mark.* **30**(3), 314–318 (2013)