

Perception of chatbots in customer service

Aksel Hjerpbakk (akselhj@stud.ntnu.no),
Anders Mølster Hopland (andermho@stud.ntnu.no),
Sander Elias Høyland (sandereh@stud.ntnu.no),
Aksel Kirknes (akselki@stud.ntnu.no),
Maria Emine Nylund (mariaiva@stud.ntnu.no)

NTNU Trondheim

Abstract—Chatbots are a cost effective tool for customer support. An increasing number of businesses employ chatbots, but how the users perceive this technology is not entirely known. In this study, we investigate how this technology is perceived, with a focus on Norwegian e-commerce customers. The study uses a questionnaire to gather data. The results from 72 responses show that 77% of the respondents would rather chat with a human than a chatbot. Frustrating experiences with chatbots in the past have a correlation with the perceived usefulness of chatbots. The majority of participants believe that chatbots work well for simple questions, but are not able to respond usefully to unusual or complicated requests.

Index Terms—Chatbots, E-Commerce, Human-Machine Interaction

I. INTRODUCTION

The use of chatbots is a cost-effective way of dealing with customer service requests, as one automated support system can replace several customer service representatives [1]. One of the advantages of using chatbots is that the response to a customer query can be almost instantaneous while relying on human support agents might subject the customer to waiting in line. It is estimated that chatbots can answer up to 80% of routine customer support questions, and reduce the customer support cost by 30% [2]. Major companies such as DNB, which is Norway's largest bank, automated 51% of its online chat traffic using a chatbot [3]. However, despite technical advances, customers continue to have unsatisfactory encounters with chatbots that are based on artificial intelligence. Chatbots may provide unsuitable responses to the user requests, leading to a gap between the user's expectation and the system's performance [4]. Our research objective is to add to this body of knowledge regarding chatbots to allow businesses and organizations to make more informed decisions regarding their customer service operation. Our research question is:

- How do Norwegian e-commerce customers perceive the usefulness of chatbots in customer service?

II. BACKGROUND

The usage of chatbots has been increasing over the last couple of years. Despite this growth, chatbots still struggle to handle the complexity of customer service interactions. Problems include a high failure rate and skepticism from users [5]. In order to solve those problems, it is important to understand the customer perspective. Norman's Gulfs of

Execution and Evaluation explains the importance of bridging the gap between user expectations and the results of their action [6]. An exploratory study in the USA has been done with 14 participants to find the gap between user expectations and the system operation of chatbots. They found that users have higher expectations of what conversational agents can do than the capabilities of the agent support [4].

There has also been a study done on Norwegian e-commerce chatbots to find what makes a user trust a chatbot [7]. The researchers conducted interviews with 14 participants to find a number of factors that affect their trust. The results have shown that trust is dependent on the context of use. In addition, the functionality of the chatbot and its perceived usefulness also affects the user's experience of trust.

A number of studies have investigated the motivation behind the application of chatbots. Brandtzaeg [8] investigates why people use chatbots. The study was performed in 2017, with its respondents being from the U.S. The main reason for the application of chatbots is to increase productivity, with entertainment and curiosity of the technology as secondary motivations.

Chatbots have the potential to automate many aspects of customer service in e-commerce. Chatbot technology is being developed at a tremendous pace and research in the field can quickly become outdated. Norway is one of the world's most digitized countries [9], but few studies have investigated customers' opinions on chatbots in Norway. Our research explores the perceived usefulness of chatbots in the Norwegian market, to try to decrease the knowledge gap.

III. METHODS

A. Research Strategy

We utilized the survey strategy as our main research strategy. The requirement for data was that it can illustrate individuals' perception of chatbots. Our sampling frame was everyone in Norway that uses e-commerce services, as these were the people that were most likely to have to use a chatbot, and were also the most relevant group in regards to useful applications of our results. We employed the use of a questionnaire as our data generation method. We have included a few questions to gather demographic data, to be able to check for correlations between these demographics and biases in the results for the rest of the questionnaire. The demographics included consist

of sex, age group, and level of education, as these were the demographics we believed might have an impact on the results. Our sample size was 72.

B. Data Generation Methods

Our survey was presented to the research subjects in the form of a self-administered questionnaire. This was because we wanted to ask pre-defined questions to all participants and have them answer the questions without being biased from the researchers. The advantage of this data generation method was that the data generation required less direct focus from the researchers. We distributed the questionnaire to our research subjects, using a combination of snowball and self-selection sampling [10]. It was possible to get a larger number of respondents using the questionnaire since it was easier to share it further. It also made analyzing the data simpler, as most of the data were ordinal. We have attempted to design the questionnaire to be simple to understand, with as few questions as possible, consisting of questions that can be answered quickly. The reason for this was to ensure a low resistance to participating in the research project, hopefully resulting in more people starting and completing the entire questionnaire [11]. The questionnaire consisted of mostly closed questions in the Likert scale format, but also some open questions to provide the possibility to elaborate more on certain topics. We included open-ended questions so the research subjects could provide more detailed information in case the Likert scale questions did not capture their exact opinion about a certain topic.

C. Data Analysis Methods

The Likert scale questions were deliberately mixed so that a 5 on the Likert scale would sometimes reflect a positive sentiment towards chatbots, and sometimes reflect a negative sentiment. This was done to ensure that the subjects read the questions thoroughly before answering [12].

We used Spearman’s rank correlation coefficient to find the correlation between responses to different questions [10]. Spearman’s rank correlation was chosen because it works well for ordinal data. The variation in sentiment in regards to what a 5 on the Likert scale represents, meant we had to correct for this before calculating the coefficient. This was done by flipping the values for the Likert scale questions where the value 5 represented a negative sentiment towards chatbots (meaning the value 5 was replaced by the value 1, 4 by 2, 2 by 4, and 1 by 5). After this correction, a positive coefficient represented the correlation between a positive chatbot sentiment for one question and a positive chatbot sentiment for another question.

IV. FINDINGS

In total, 72 people completed the questionnaire. Of the respondents, 43 were male, and 29 were female. 68% of the respondents indicated that they had used a chatbot during the past 12 months.

Below is a list of some of the questions in our questionnaire, with letters to represent them through the rest of the paper:

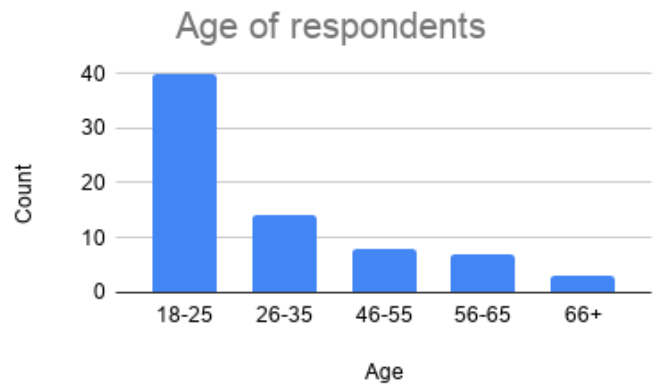


Fig. 1. Chart of age of respondents

Likert style questions

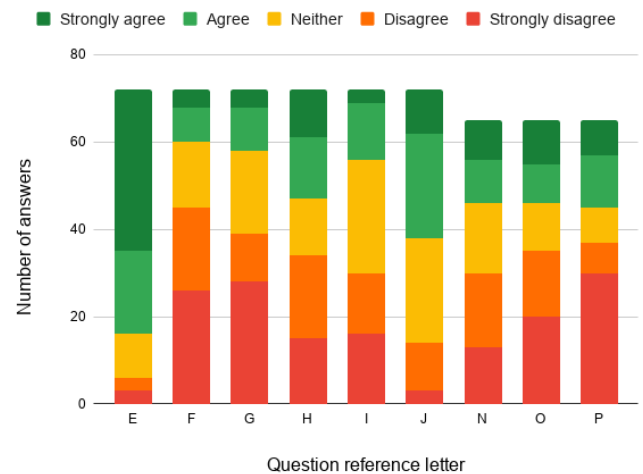


Fig. 2. Results to questions (Likert scale) from strongly disagree to strongly agree. Questions N, O and P were only answered by respondents who have used chatbots, because of the nature of the questions.

- **A:** Education level
- **B:** Age
- **E:** If I can choose between chatting with a human and a robot, I always choose to chat with a human
- **F:** If I am in a hurry, I will always choose to chat with a chatbot rather than with a human
- **G:** If I have to share sensitive information, I prefer to share information with a chatbot rather than with a human
- **H:** If a chatbot is the only available customer service, I will rather try to find the answer on my own
- **I:** If I have a question, I think a chatbot will be able to help
- **J:** I think that other people use chatbots more than me
- **N:** I have experienced that a chatbot has given me false or misleading answers
- **O:** I am satisfied with my last conversation with a chatbot
- **P:** My query was resolved the last time I used a chatbot

It is clear from the responses to our questionnaire that most of our respondents prefer to chat with humans rather than chatting with a chatbot when they have customer service needs. According to the responses from question **H**, a substantial minority of customers prefer the option of chatting with a chatbot to no customer support at all, implying that chatbots are somewhat useful. Question **F** implies that the convenience of not having to wait in line when chatting with a chatbot is not an important factor for most of the respondents.

	E	F	G	H	I	J
N	0.19	0.16	0.05	0.36	0.30	0.21
O	0.46	0.72	0.37	0.60	0.63	0.28
P	0.37	0.63	0.33	0.49	0.65	0.25

TABLE I
SPEARMAN COEFFICIENTS, OPINIONS AND EXPERIENCES

We decided to explore the correlation between experiences using chatbots and opinions about chatbots. To find the correlation between these responses, we have calculated the correlation coefficients for the response pairs using Spearman’s rank correlation coefficient measure. A correlation coefficient with an absolute value greater than 0.3 indicates a substantial correlation between two data sets [10]. As we can see from table 1, responses to questions N and J seem not to be particularly correlated to any of the compared responses. However, responses to question O and P seems to be correlated to most of the compared responses, and highly correlated with responses to questions F, H, and I. Responses to question O seems to have the highest correlation to these questions overall. Our collected data implies a strong correlation between positive past experiences with chatbots and positive sentiment in regards to chatbots.

	E	F	G	H	I	J
A	0.40	0.01	0.02	0.29	0.11	0.02
B	-0.14	0.06	0.05	0.10	0.21	-0.07

TABLE II
SPEARMAN COEFFICIENTS, OPINIONS AND DEMOGRAPHICS

We also explored the correlation between demographics and opinions about chatbots using Spearman’s rank correlation coefficient measure. There was no clear correlation between demographics and opinions about chatbots, except for a slight correlation between the level of education and a disagreement with statement **E** (higher education levels were less likely to prefer chatting with a human in all situations). As the demographics of our respondents are very skewed towards students between 18 and 25 years old, the confidence in these correlations is weak.

V. DISCUSSION

In our introduction we mentioned that “It is estimated that chatbots can answer up to 80% of routine customer support questions ...”[2]. This is not supported by our findings as most customers has recent negative experience(s) with chatbots as seen in figure 2 (question *O*). Our findings indicate that a majority of people prefer to chat with a human rather than

a chatbot in most, if not all scenarios. We would like to highlight the result of question **E** and **O** which shows a trend of discontentment related to chatbots. The findings further indicate that the confidence in chatbots’ usefulness is limited in the general population. Norwegian users prefer chatting with real humans rather than chatbots. However, the applications of chatbots are not void. Companies and organizations should strive towards deploying human support if they want higher customer satisfaction, but using chatbots is better than no active support.

An example of good use of chatbots in the context of our findings could be using them to help customers in the hours where support services are closed, or overloaded. This way the chatbots can attempt to assist with common issues and refer to human support in the event where a solution is not found. This is based on the fact that a total of 34 participants would rather talk to a chatbot than search for information by themselves. We would also argue that an example of misuse of chatbots is to use them as the default support solution in any scenario. Even if the chatbot later refers you to human support, the experience has already been tainted. A total of 30 participants strongly agreed or agreed that they had experienced getting misleading information from a chatbot, and exposing users to chatbots when human support is available could negatively impact customer satisfaction.

In our findings, we see that there is little correlation between education/age and the perception of chatbots IV. This is an interesting finding, and the reasoning behind it is unknown. The perception of chatbots might not correlate with demographics, but a likely alternative reason for this result is the sparseness of variation in regards to respondent demographics affecting the Spearman’s rank correlation coefficient.

We mentioned in section III-C that we were flipping the values of some of the raw data. The result was that the coefficient generated no longer would show a negative value, in the case of a descending monotonic behaviour. One can argue, by doing this we would no longer be capable of separating the descending and ascending monotonic behaviour and thus restrict the research. At the same time, the researchers found that it was easier to read the produced coefficients. As the intended use of the correlation coefficients was to determine what data sets to focus on, we did not find it necessary to separate the different monotonic behaviours.

Comparing our findings to the findings of Luger’s [4] paper, it appears that when chatbots are presented as a substitute for human customer service, user expectations far exceed chatbots practical capabilities. This leads to frustration. Our findings are in line with Følstad’s [7] paper, which argues that although chatbots are unable to answer complex questions, they are sometimes presented as such on websites, leading to false affordance.

VI. CONCLUSION

Our research has several limitations that can negatively impact the validity of our results. A sample size of 72 is not nearly large enough to make conclusive statements about a

large group of people, but might give an indication regarding the opinions of our sample frame. Since our sampling frame is in the order of magnitude of millions of people, ideally our sample size should be at least 1000 people, according to table 7.2 in *Researching information systems and computing* [10]. The group is also very skewed towards the younger and more well educated demographic, which could significantly impact our results. We also believe that a questionnaire has substantial limitations, such as respondents being dishonest or misremembering, leading to inaccurate data. This can be mitigated by supplementing a questionnaire with other data generation methods. Observations of someone that actively is chatting with a chatbot could serve as a confirmation method where we would collect enough data to see if there is a correlation between the findings. Due to time and resource limitations, we chose to focus on doing our research using only a questionnaire as the data generation method for the survey strategy. We believe our findings should be treated with caution, as they are not sufficient to provide conclusive insights. However, we do believe they can give an indication of opinions, attitudes and patterns of behaviour in the Norwegian population.

Our research indicates that most people are familiar with chatbots and use them on a semi-regular basis. It also makes it clear that most people prefer to chat with a human rather than a chatbot in all circumstances. People are not very confident in the abilities of chatbots, and indicate that they are mostly useful for answering simple questions that could have been found by searching the website. However, it seems that most people would rather have a chatbot available than no customer service provided at all. Negative experiences with chatbots seem to correlate with a preference for chatting with humans rather than chatbots, and people that have negative experiences with chatbots are very unlikely to prefer chatbots to humans even if they are in a hurry.

Furthermore, it would be beneficial to look further into what types of questions chatbots are able to answer and what expectations customers have when using chatbots. It would also be interesting to see how these attitudes change over time, as chatbots develop and become more advanced.

ACKNOWLEDGMENT

We would like to thank everyone who has enabled us to conduct this research, including the course staff for IT3010, our reviewers, the research participants and the people who provided feedback on our questionnaire. A big thank you to the Overleaf team, who have made a terrific program for writing journals.

REFERENCES

- [1] S. Woodford, "Why chatbots are essential to retail," Tech. Rep., 2020.
- [2] A. Martin, "Ai-based chatbots in customer service and their effects on user compliance," *Electron Markets*, 2020.
- [3] B. AI. (). "How norway's biggest bank automated 51% of its online chat traffic with ai," [Online]. Available: <https://www.boost.ai/articles/how-norways-biggest-bank-automated-51-of-its-online-chat-traffic-with-ai> (visited on 04/18/2021).
- [4] E. Luger and A. Sellen, "Like having a really bad pa" the gulf between user expectation and experience of conversational agents," in *Proceedings of the 2016 CHI conference on human factors in computing systems*, 2016, pp. 5286–5297.
- [5] A. Orłowski, *Facebook scales back ai flagship after chatbots hit 70% f-ai-lure rate*, 2017.
- [6] D. A. Norman, "Cognitive engineering," *User centered system design*, vol. 31, p. 61, 1986.
- [7] A. Følstad, C. B. Nordheim, and C. A. Bjørkli, "What makes users trust a chatbot for customer service? an exploratory interview study," in *International conference on internet science*, Springer, 2018, pp. 194–208.
- [8] P. B. Brandtzaeg and A. Følstad, "Why people use chatbots," in *International conference on internet science*, Springer, 2017, pp. 377–392.
- [9] I. W. C. Center. (). "Imd world digital competitiveness ranking 2020," [Online]. Available: <https://www.imd.org/wcc/world-competitiveness-center-rankings/world-digital-competitiveness-rankings-2020/> (visited on 04/19/2021).
- [10] B. J. Oates, *Researching information systems and computing*. Sage, 2005.
- [11] D. Stone, "Design a questionnaire.," *British Medical Journal*, vol. 307, no. 6914, pp. 1264–1266, 1993.
- [12] J. J. Barnette, "Effects of stem and likert response option reversals on survey internal consistency: If you feel the need, there is a better alternative to using those negatively worded stems," *Educational and Psychological Measurement*, vol. 60, no. 3, pp. 361–370, 2000.