An Initial Model of Trust in Chatbots for Customer Service – Findings from a Questionnaire Study

Cecilie Bertinussen Nordheim, Asbjørn Følstad, Cato Alexander Bjørkli

University of Oslo, Norway | SINTEF, Norway

Abstract

Chatbots are predicted to play a key role in customer service. Users' trust in such chatbots is critical for their uptake. However, there is a lack of knowledge concerning users' trust in chatbots. To bridge this knowledge gap, we present a questionnaire study ($N = 154$) that investigated factors of relevance for trust in customer service chatbots. The study included two parts: an explanatory investigation of the relative importance of factors known to predict trust from the general literature on interactive systems and an exploratory identification of other factors of particular relevance for trust in chatbots. The participants were recruited as part of their dialogue with one of four chatbots for customer service. Based on the findings, we propose an initial model of trust in chatbots for customer service, including chatbot-related factors (perceived expertise and responsiveness), environment-related factors (risk and brand perceptions), and user-related factors (propensity to trust technology).

Key words

Natural language interfaces; HCI theory, concepts and models, trust; user studies

Research highlights

- We extend the current knowledge base on natural language interfaces by investigating factors affecting users' trust in chatbots for customer service.
- Chatbot-related factors, specifically perceived expertise and responsiveness, are found particularly important to users' trust in such chatbots, but also environment-related factors such as brand perception and user-related factors such as propensity to trust technology.
- On the basis of the findings, we propose an initial model of users' trust chatbots for customer service.

1   Introduction

Advances in artificial intelligence (AI) is predicted to significantly impact the labour market, as it holds promise of automating tasks that were previously thought of as requiring human personnel (Frey & Osborne, 2017). Within this context, an interesting trend is service providers' introduction of chatbots for customer service. Chatbots are computer programs that interact with humans through natural language (Abu Shawar & Atwell, 2007). In a customer service context, chatbots may serve as a first line of support by providing an accessible and low-threshold source of help and information for frequently asked questions and support tasks. Retail, finance, and information and communication technology are examples of sectors that have begun introducing chatbots for customer service.

Chatbots for customer service are seen as holding great potential in terms of service efficiency, cost savings, and customer experience. Chatbots may also be part of service providers' digitizing efforts, needed for survival in an increasingly technology-driven market. The chatbot market is expected to have a yearly growth rate of about 25% through 2025 (Grand View Research, 2017). Major technology companies such as IBM and Microsoft have invested substantially in chatbot platforms to power customer service. Alibaba and Aliexpress are examples of early-adopting service providers that have made chatbots a key feature of their customer service offering. However, observers have noted that users' adoption of chatbots has been lower than the industry expected (e.g. Griffith & Simonite, 2018). One reason might be that chatbot development has been guided more by a technology push than a market pull; consequently, the development has lacked focus on user needs and desires (Coniam, 2014).

User trust is essential for new online technologies to be adopted in the market (Corritore, Kracher, & Wiedenbeck, 2003). The humanlike qualities of chatbots, in particular their

capability for natural language interaction, may make trust particularly important (Holtgraves, Ross, Weywadt, & Han, 2007). Although considerable research has investigated trust in the interpersonal and societal domains, and in different technologies, studies addressing trust in chatbots are scarce. Specifically, the field lacks knowledge concerning trust in chatbots for customer service. This knowledge gap is critical, as technology companies and service providers need adequate models of users' trust to improve and adapt such chatbots to meet the needs and desires of users.

Addressing this knowledge gap, we present a questionnaire study aiming to establish an initial model of users' trust in chatbots for customer service. The study took as a starting point factors important to trust in interactive systems as identified in the literature. However, recognizing the particularities of chatbots as an interactive technology, we designed the study (a) to establish which factors from the literature on trust in interactive systems have most relevance to chatbots for customer service and (b) to identify other factors of importance for such chatbots.

The contribution of our study is a model that includes three groups of factors assumed to be relevant for users' trust in chatbots for customer service: *chatbot-related factors* such as perceived expertise and responsiveness, *environment-related factors* such as risk and brand perceptions, and *user-related factors* such as propensity to trust technology. We refer to this model as an *initial* model, as we believe the factors relevant for users' trust in chatbots may change as chatbots become more commonplace.

The remainder of the paper is structured as follows. First, we present relevant background for the study. We then detail the research question and method before presenting and discussing the study results. Finally, we address study limitations and suggest future research.

2   Background

2.1   Automation in customer service

Customer service is key to how users experience service providers. The ability of customer service to provide needed information and resolve users' issues is important for users' loyalty to service providers (Dixon, Freeman, & Toman, 2010); this to the point where users experiencing service recovery through well-managed customer service may be more satisfied with a service provider than users without the need for such service recovery (de Matos, Henrique, & Rossi, 2007). At the same time, customer service is highly resource-intensive, and service providers are constantly seeking to make this area of their operations more efficient without compromising on user experience (Meuter, Ostrom, Roundtree, & Bitner, 2000).

For decades, the customer service field has been moving towards strengthening users' options for self-service, in particular through customer websites (Meuter, Bitner, Ostrom, & Brown, 2005). However, though self-service options have been continuously improved and extended, the volume of users contacting service providers through channels requiring manual handling of service requests has remained high. This seemingly paradoxical situation is due, in part, to users' deliberate use of multichannel approaches that may involve both self-service through customer websites and dialogue through phone or chat (Følstad, Kvale, & Haugstveit, 2014).

The emergence of AI-powered customer service may automate responses to service requests through channels such as chat, which are currently manned by service personnel. Service providers see chat as a preferable channel for customer support, in particular because service personnel can handle multiple responses in parallel (Tezcan & Zhang, 2014). Servion (2017) recently predicted that 95% of all customer interactions, including live telephone and online conversations, could be handled by AI applications by 2025. In their analysis of automation

by means of computers in the future labour market, Frey and Osborne (2017) found the probability of such automation for customer service to be substantial. In line with these predictions, chatbots can for example provide first-line customer support and escalate more complex service requests to human personnel.

A key challenge in automating customer support, however, is to enable flexible and satisfactory service provision while introducing automation in channels that are currently manned by service personnel. Leveraging technology for service provision is seen as a priority in service research (Ostrom, Parasuraman, Bowen, Patrício, & Voss, 2015).

2.2 Chatbots for customer service

The idea of automated conversational systems has deep roots in informatics, going back to Weizenbaum's (1966) work on ELIZA in the 1960s. The sudden current interest can largely be attributed to two developments. First, advances in AI and machine learning have made chatbots easier to train and implement, due to improved natural language processing and strengthened capabilities for identifying users' intents and sentiment (Brandtzaeg & Følstad, 2017). Second, text-based interaction with users has gained interest due to the availability and popular uptake of messaging platforms such as Facebook Messenger, Slack, WhatsApp, and WeChat (Zumstein & Hundertmark, 2017).

Chatbots proliferate on messaging platforms and are increasingly implemented as digital assistants by large technology companies. In consequence, chatbot technology is gradually becoming part of people's everyday life (Accenture, 2016). Chatbots have been introduced in a variety of online environments such as e-commerce, daily news, and delivery services. Recently, the development of chatbots for customer service and marketing has seen substantial growth (Zumstein & Hundertmark, 2017). A range of platforms, such as IBM Watson and Google's Dialogflow, support chatbots for customer service. Nascent research

has explored how AI-powered chatbots perform, for example in terms of response to customer sentiment (Xu, Liu, Guo, Sinha, & Akkiraju, 2017).

Research on users' trust in chatbots is scarce. We are aware of only one small-scale interview study addressing this topic. Følstad, Nordheim, and Bjørkli (2018) found trust in customer service chatbots to be related to factors specific to the chatbot and factors specific to the service context. Chatbot-specific factors included its ability to correctly interpret user requests and provide useful answers, as well as its human-likeness, self-presentation, and professional appearance. Factors specific to the service context included users' perceptions of the brand hosting the chatbot, security and privacy, and risk.

While research on users' trust in chatbots is scarce, a substantial body of research has investigated users' interactions with and perceptions of chatbots. For example, Hill, Ford, and Farreras (2015) compared online human–human interaction and human–chatbot interaction, finding that human–chatbot interactions lasted longer and involved shorter messages than human–human interactions. Corti and Gillespie (2016) found users to invest more effort to repair misunderstandings when the chatbot was perceived as human, compared to when the chatbot was perceived as an automated conversational agent. Murgia, Janssens, Demeyer, and Vasilescu (2016) studied human–chatbot interaction in the context of a question–answer website and found that users' perceptions of the chatbot depended on its self-presentation. The current body of research on users' perceptions and experiences of chatbots has suggested that chatbots often are seen as holding humanlike characteristics (Candello, Pinhanez, & Figueiredo, 2017) and that they may be perceived as an interaction partner approaching that of a fellow human being (Mone, 2016).

2.3   The concept of trust

The extensive body of research on trust has ranged from the fields of psychology and sociology to technology studies (Corritore et al., 2003). According to Mayer, Davis, and Schoorman (1995), trust is:

> the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other part. (p. 712)

In the area of automation, Madsen and Gregor (2000) defined trust as "the extent to which a user is confident in, and willing to act on the basis of the recommendations, actions, and decisions of an artificially intelligent decision aid" (p. 1).

Theories of interpersonal relationships have established trust as a social glue in relationships, groups, and societies (Van Lange, 2015). With the increased reliance on autonomous technologies, a rapidly growing body of research has addressed trust in technology. For example, substantial research has investigated trust in robots (Hancock et al., 2011), and work has been done on measuring trust in automated systems (Jian, Bisantz, & Drury, 2000).

As a starting point for a model of trust in chatbots for customer service, the much-cited model of Corritore et al. (2003) on online trust is promising. Although Corritore et al. (2003) developed this model in the context of informational and transactional webpages, it includes elements that are useful for understanding trust in other online technologies. It considers users' trust in online technologies as dependent on their perceptions of three main factors: ease of use, risk, and credibility. *Ease of use* concerns users' perceptions of how simple the interactive technology is to use and has been found critical for predicting technology acceptance. *Risk* is held an important determinant of trust in the literature on interpersonal relations and concerns users' perceptions of the likelihood of an undesirable outcome.

Credibility is construed as consisting of the following elements: honesty, expertise, predictability, and reputation. Of these, we considered expertise, predictability, and reputation as potentially important to the study of trust in chatbots. *Expertise* concerns users' perceptions of the knowledge, experience, and competence reflected in the interactive system, *predictability* concerns perceptions of consistency in the system's behaviour, and *reputation* concerns the recognized quality of past interaction experiences.

Corritore et al. (2003) also alluded to the potential importance of individual differences in trust in technology. Likewise, the general literature on interpersonal trust has assumed that individuals' propensity for generalized trust differs (Mayer et al., 1995). Additionally, McKnight, Carter, Thatcher, and Clay (2011) and Cheung and Lee (2001) further discussed users' propensity to trust in technology as a factor with substantial individual variation. This propensity to trust is generalized across use contexts rather than being dependent on the specific technology or usage situation. Hence, the propensity to trust technology is a factor that resides in the user.

In addition to the trust factors discussed by Corritore et al. (2003), the study of trust in chatbots arguably may benefit from including factors reflecting the humanlike character of chatbots. De Visser et al. (2016) found that what they refer to as *cognitive agents* are more resilient to breakdowns in trust as their level of human-likeness increases and determined that introducing humanlike behaviour for trust repair may strengthen trust in such agents. Given that chatbots have humanlike characteristics, in particular given their capability for interaction in natural language, users are likely to ascribe human traits to chatbots (Mone, 2016) and may even struggle to discriminate between a chatbot and a human conversational partner (Candello et al., 2017). One recent example of this development is the Google Duplex, where an automated conversational system was showcased doing humanlike calls to place bookings (Leviathan & Matias, 2018). Given the increasingly humanlike character of some interactive

systems, there has been substantial research establishing measures for human-likeness in such systems (Ho & MacDorman, 2010).

Based on the above considerations, we entered our study of trust in chatbots for customer service with the seven candidate factors presented in Table 1. The factors are structured in three high-level groups, corresponding to what Corritore et al. (2003) referred to as the object of trust, the situation, and the trustor. Hardin (2003) and Hancock et al. (2011) have proposed similar tripartite classifications of trust-related factors encompassing the trusted person or technology, the trustor, and the environment or matter at stake.

Table 1: Factors from the literature assumed to affect trust in chatbots for customer service

Group	Factor	Content	Reference
Chatbot-related factors	Expertise	Users' perceptions of the knowledge, experience, and competence reflected in the interactive system	Corritore et al. (2003)
	Predicta-bility	Users' perceptions of the consistency with which the interactive system behaves	Corritore et al. (2003)
	Human-likeness	The perceived anthropomorphic character of the interactive system	Ho and MacDorman (2010)
	Ease of use	The ease or simplicity with which the interaction with the system is accomplished	Corritore et al. (2003)
Environment-related factors	Risk	Users' perceptions regarding the likelihood of an undesirable outcome	Corritore et al. (2003)
	Reputation	Perceptions of others' assessment of the interactive system	Corritore et al. (2003)
User-related factors	Propensity to trust technology	Individual variation in users' generalized tendency to enter a trusting relating with an interactive system	McKnight et al. (2011)

3   Research question and hypothesis

The field currently has a knowledge gap concerning users' trust in chatbots. In particular, it is important to gain knowledge regarding trust in chatbots for customer service, which is predicted to become a major arena for chatbot applications. Hence, the following research question was formulated for this study:

> *Which factors explain users' trust in chatbots for customer service?*

This research question presents a dual set of challenges: an explanatory challenge and an exploratory challenge.

The explanatory challenge concerns the lack of knowledge regarding which factors for trust, as identified from the general literature on interactive systems, may explain variation in users' trust in chatbots for customer service. We may assume that trust in chatbots for customer service is affected by variations in the levels of all the factors listed in Table 1; this hardly requires a research investigation. What is important for theory and practice, however, is to identify those factors that explain the most variation in such trust.

The exploratory challenge concerns the identification of factors particular to trust in chatbots that have not previously been identified in the general literature on trust in interactive systems. In their small-scale interview study, Følstad et al. (2018) found indications that factors beyond those identified in the literature may be critical for users' trust in customer service chatbots. For example, the interviewed participants suggested that such trust may be dependent on the perceived quality in the chatbot's interpretation and the perceived characteristics of the chatbot owner's brand. For the exploratory challenge, we did not make any assumptions concerning the factors to be identified. Rather, we viewed this as requiring an open-ended exploration to identify factors not previously highlighted in the literature.

By addressing these explanatory and exploratory challenges, we intend to establish an initial model of trust in chatbots for customer service. This initial model should be useful to guide future research and practice.

4   Method

The dual challenges associated with the research question motivated a research design allowing for (a) quantitative assessment of the explanatory power of factors identified from the literature and (b) qualitative exploration of relevant factors not identified in the general

literature on trust in interactive systems. In response to these needs, we chose to conduct a questionnaire study incorporating quantitative and qualitative questionnaire items.

4.1   Study participants and context

To ensure that all participants had recently experienced chatbot interaction, we invited chatbot users to participate in the study as part of their interaction with one of four chatbots for customer service. Each participant received an invitation upon the conclusion of a dialogue with a chatbot, and the participants were given access to the questionnaire immediately after accepting the invitation. As incentive, all participants were eligible to join a lottery to win a tablet computer.

The chatbots through which the participants were recruited were part of the customer service offerings of four Norwegian consumer service providers. We chose to recruit participants through different chatbots to improve the generalizability of the study findings. We also determined the Norwegian context to be beneficial. It is relatively easy to get an overview of the Norwegian market of chatbots for customer service. Furthermore, the Norwegian population is relatively advanced in terms of its uptake of smartphones and mobile Internet, a main driver of chatbot services; nearly all Norwegians in the age group 16-79 (92%) were users of smartphones as of 2018 (Statistics Norway, 2018). Also, Norwegian service providers are relatively advanced in their use of chatbots for customer service purposes; leading banking services, telecom providers, as well as a substantial number of municipalities offer customer service through chatbots on their corporate websites. These context characteristics suggest that findings from the Norwegian context may be seen as a pointer towards future developments in markets less advanced in terms of chatbot uptake.

4.2   Study material

The questionnaire used in the study consisted of measurement instruments to support the quantitative explanatory analysis as well as an open-ended free-text item to support the qualitative exploratory analysis.

Measurement instruments

No instruments previously existed to measure trust in chatbots and related factors. As a result, we adapted instruments from the literature to establish the applied measurement instruments. The questionnaire included instruments to measure *trust*, the dependent variable of the study, and seven factors from the literature assumed to explain variation in trust: *expertise*, *predictability*, *human-likeness*, *ease of use*, *risk*, *reputation*, and *propensity to trust technology*. In addition, we gathered data on participants' age, gender, education, and previous chatbot experience.

The dependent variable and the related factors were measured through five Likert-scale items with scales ranging from 1 (disagree completely) to 7 (agree completely). Following data collection, we conducted a principal component analysis to check whether the items intended to measure each factor indeed loaded on this and whether it was necessary to exclude some items before continuing the analysis. The principal component analysis suggested a factor structure in line with expectations for most of the items, but with some exceptions. In consequence, we removed two items from the measurement instrument for *ease of use* and one item from the measurement instrument for *reputation*. We also identified some examples of weak factor loadings and cross-loading, in particular for the *trust* measurement instrument. As trust is a broad construct, this was not unexpected and hence not considered a critical problem for subsequent analysis.

Inter-item reliability was adequate for *trust* and for each of the seven related factors (Cronbach's alpha > .70). For a full overview of the questionnaire items for the measurement instruments, please consider Appendix 1. Appendix 2 presents the pattern matrix for the exploratory factor analysis.

Open-ended free-text item for exploration

The questionnaire also included an open-ended free-text item that formed the basis for the exploratory part of study. Participants were encouraged to use their own words to describe what they saw as affecting their trust in the chatbot they had used for customer service immediately prior to answering the questionnaire. The question was phrased as follows: "What makes you experience trust in this chatbot?"[1] The questionnaire concluded by asking participants to answer a second open-ended question about suggested changes for the chatbot. Due to space limitations, we have not addressed responses to this second free-text question in this paper.

Ethical considerations

The research design was reviewed by the relevant institutional body to check compliance with privacy requirements for research. Participation followed informed consent. No personal data was collected, apart from the participants' email addresses for the participant lottery. These email addresses were collected in a separate form with no connection or back-link to the questionnaire form.

4.3   Analysis

Quantitative analysis – preparation

The final dataset was prepared to included participants who met five criteria. Specifically, participants (a) were 18 years or older, (b) had not given the same score for more than 90% of

---

[1] This question text was translated from Norwegian by the first author.

the questionnaire items, (c) had completed more than 50% of the questionnaire, and (d) provided a meaningful answer for at least one of the two open-ended questions. Finally, (e) each participant's answers reflected that the participant had recognized three reversed questionnaire items as such.

The raw data included the responses from 175 chatbot users. Of these, 21 were excluded from the final dataset in line with the criteria above. Four were under the age of 18; one had no variation in the answers; seven had completed less than 50% of the questionnaire; four had provided nonsense, joke, or no answers to the open-ended questions; and five had answers indicating that they had not recognized the three reversed questions were reversed. The answers of the remaining 154 participants were included in the subsequent analysis.

Quantitative analysis

In the explanatory part of the study, we conducted a multiple regression analysis to identify the factors of highest relevance for explaining variation in users' trust in chatbots for customer service. This analysis was conducted as a linear regression (enter) on IBM SPSS Statistics 25. Furthermore, we conducted an additional independent samples t-test (gender) and a series of regressions (age) to check for any effects due to gender and age differences.

Qualitative analysis

For the exploratory part of the study, we conducted a thematic analysis on the free-text data from the open-ended questionnaire item. The approach followed Ezzy (2002), where we allowed coding categories to emerge from the data rather than applying predefined themes. Each participant response could be coded as reflecting none, one, or several coding categories.

To validate the coding categories, two analysts (one of these being the first author) coded the data independently to check inter-coder agreement. For the final set of coding categories, all categories were found to have adequate inter-coder agreement (Cohen's kappa > .70).

5   Results

5.1   Participant characteristics

In total, the final dataset included 154 participants who were evenly distributed across gender (48% women; 52% men; one participant did not report gender). Participants ranged in age from 18 to 83 with a mean age of 41 ($SD = 13.9$); however, 20 participants did not report on age. Most participants reported to have higher education; specifically, 37% reported one to three years of higher education, and 42% reported four or more years.

While our research design ensured that all participants had used chatbots for customer service immediately prior to their participation, most participants still were fairly new to chatbots. In total, 57% reported to have used chatbots one to three times before, and only 19% had used chatbots more than 10 times.

5.2   Findings from the explanatory analysis

An overview of the studied variables

This section presents the results from the explanatory part of the study, where respondents answered questionnaire items measuring trust and related factors identified from the literature on trust in interactive systems.

Table 2 provides an initial overview of the scores for the dependent variable *trust* and the seven factors assumed to be of relevance for trust in customer service chatbots. The scores for all variables were calculated as the average of the associated item scores. Mean scores for most variables ranged between 4 and 6 (on a scale from 1 to 7), except for *ease of use* with

relatively higher scores (M = 6.3; SD = 1.1) and *risk* with relatively lower scores (M = 2.1; SD = 1.3). These two factors also had higher skew than the other variables, assumed to be due to the relatively high and low mean scores for these variables.

Table 2: Sample size (N), mean (M), standard deviation (SD), and skewness (skew) for trust and the seven related factors

Variable	N	M	SD	Skew
Trust	154	5.57	1.19	-.95
Expertise	154	5.32	1.72	-1.06
Predictability	153	5.54	1.27	-.67
Human-likeness	154	4.78	1.66	-.58
Ease of use	154	6.29	1.10	-2.21
Risk	154	2.12	1.34	1.42
Reputation	149	4.25	1.40	-.05
Propensity to trust technology	154	5.12	1.28	-.56

Multiple regression analysis – explaining variation in trust

We conducted a multiple regression analysis to identify which of the seven factors were most important for explaining variation in *trust*. The results of the regression analysis indicated that the seven factors explained more than half of the variance in *trust*, adjusted R^2 = .60, F (7) = 32.9, $p < .001$. Of the seven predictors, the analysis revealed three to be particularly important to explain variation in *trust*: *expertise* (β = .31, $p < .001$), *risk* (β = -.17, $p < .05$), and *propensity to trust technology* (β = .18, $p < .05$). These three factors made a significant contribution to the regression model. In addition, *human-likeness* bordered significance (β = .17, p = .05). See Table 3 for details.

Table 3: Standardized regression coefficients (β) for the seven factors from the literature assumed to predict trust

Factor	β	t	p-value
Expertise	.31	3.54	< .01
Predictability	.12	1.69	.09
Human-likeness	.17	1.96	.05
Ease of use	.13	1.76	.08
Risk	-.17	-2.62	< .05
Reputation	.01	.19	.85

			.18	3.00	< .01

Propensity to trust technology .... .18 .... 3.00 .... < .01

Note. Dependent variable *trust.*

Checking for effects of gender and age

We conducted an independent samples *t*-test to check for gender differences in trust. The results indicated no such gender difference, $t(151) = -.28$, $p = .78$. We also conduced a linear regression analysis with trust as the dependent variable to check for any age effects. This regression analysis indicated no such effects. Furthermore, a line of best fit was established to investigate linear ($R^2 = .006$), quadratic ($R^2 = .0005$), and cubic ($R^2 = .011$) relations.

5.3   Findings from the exploratory analysis

This section presents the results from the exploratory part in the questionnaire where respondents could freely write their thoughts in response to an open-ended question: "What makes you experience trust in this chatbot?" Table 4 presents the final set of categories identified in the thematic analysis. The table also presents the high-level group for each category and offers a short description, frequency, and Cohen's kappa for each category.

Table 4: Categories from the thematic analysis of responses to the question: "What makes you experience trust in this chatbot?"

Group	Category	Explanation	Frequency	Cohen's kappa
Chatbot-related categories	Expertise - Correct answer	Accuracy and relevance of information provided by chatbot reported to be important for trust.	41 (28.1%)	.83
	Expertise - Interpretation	The chatbot's correct interpretation and understanding of the user's request, as well as clear expressions of what it does not understand, reported to be important for trust.	14 (9.6%)	.96
	Expertise - Concrete answer	Concrete, clear, and easily understandable answers provided by the chatbot reported to be important for trust.	24 (16.4%)	.77
	Expertise - Eloquent answer	The chatbot providing professionally phrased answers reported to be important for trust.	14 (9.6%)	.84
	Responsiveness	Quick responses from the chatbot reported to be important for trust.	27 (18.5%)	.89

	Human-likeness	The chatbot's humanlike characteristics, including pleasantries and politeness, reported to be important for trust.	9 (6.2%)	1
	Absence of marketing	Absence of marketing, and a sense of the chatbot putting the customer first, reported to be important for trust.	6 (4.1%)	.83
Environment-related categories	Brand	Trust in the chatbot reported to be dependent on the participant's trust in the service provider for which the chatbot operates.	25 (17.1%)	.83
	Risk	Low levels of perceived risk, for example due to not needing to provide personal or sensitive information, reported to be important for trust.	9 (6.2%)	1
	Complementarity	Reports suggesting that having an opportunity to get transferred to a human operator is important for trust.	7 (4.8%)	.92
User-related categories	Not trust-relevant/ No trust	Participants reporting not to consider trust as relevant, or reporting not to have trust in chatbots due to their limited capabilities.	11 (7.5%)	.78
	Miscellaneous	Diverse answers with no direct fit to the other themes, not forming distinct themes.	15 (10.3%)	.76

Note. Frequency is calculated based on the 146 respondents that provided a meaningful answer to this question.

Chatbot-related categories

Results from the thematic analysis revealed chatbot-related categories as important for users' trust in chatbots for customer service. This importance emerged in particular in the four categories *expertise*, *responsiveness*, *human-likeness*, and *absence of marketing*. Expertise was further divided into four sub-categories: *correct answer*, *interpretation*, *concrete answer*, and *eloquent answer*; all reflecting important aspects of the expertise category.

Expertise – correct answer **(28.1%).** The participant reports strongly suggested that the correctness and relevance of the chatbots' answers were important for trust. This sub-category

of expertise was the most frequent with 28% of the participants' answers interpreted as reflecting this category. The following statements reflect the sub-category *correct answer*[2]:

I experience trust when the chatbot gives a correct answer to my question (P126)

It gives me the answer I am looking for. (P9)

***Expertise – interpretation* (9.6%).** The participant reports indicated that the chatbots' interpretation of the question was important for their trust in the chatbot. This sub-category of expertise reflected the participants' desire for the chatbot to immediately understand their question and thereby provide support. The participants also reported it as important that the chatbot provided means for avoiding confusion and misinterpretation, such as by clarifying which questions it could and could not answer. The statements below exemplify the sub-category *interpretation*:

[…] it perceives the problem and has several solutions (P29)

It is honest about saying what it can and cannot do (P74)

***Expertise – concrete answer* (16.4%).** According to the participants, the chatbots' ability to provide concrete answers was of substantial importance to trust. This sub-category of expertise concerned users' appreciation of short, precise, and clear-cut answers from the chatbot. A concrete answer seemed to be important due to users' wish for answers that were not misleading or that could not easily be misunderstood. Examples of answers exemplifying the sub-category *concrete answer* include:

Its answers to my question are short and precise (P48)

---

[2] All participant quotes were translated from Norwegian by the first author.

Clear answers that do not allow for misinterpretation [...] (P59)

***Expertise – eloquent answer* (9.6%).** The final sub-category of expertise concerned whether the chatbot formulated the answers in an eloquent manner. Comments from this sub-category suggested that users appreciated answers that appeared professional, credible, and reasonable when forming trust perceptions. The responses below exemplify the sub-category *eloquent answer*:

> *I got good answers to my questions. The robot's style of presentation seems professional and knowledgeable* (P121)

> *[...] there is logic behind the answers I get, even if I don't ask bank-related questions* (P91)

***Responsiveness* (18.5%).** The perceived expertise in the chatbot was not the only chatbot-related category to be identified. The participant reports also strongly indicated that users' trust depended on getting fast response from the chatbot. Participants appeared to appreciate such responsiveness because it made the chatbot an efficient means of support. Some participants compared the customer service chatbot with human customer service and noted that the chatbot could be a quicker way to get the requested support which, in turn, may affect trust perceptions. The following statements illustrate *responsiveness*:

> *It did right away understand what I wanted and could quickly help me* (P133)

> *Quick answer. Do not have to wait in queue* (P1)

***Human-likeness* (6.2%).** The participant reports indicated that some users appreciated humanlike characteristics in the chatbot and felt these were important for trust. These characteristics included providing polite answers or using colloquial expressions. A humanlike answer may be important to trust because customers are accustomed to such

responses when chatting with human customer service personnel. The participants may therefore have experienced the humanlike character of the chatbot as representing a familiar form of conversation. Some even reported they would not have recognized the chatbot as a machine, if it had not reported its nature itself. The following examples illustrate the category *human-likeness*:

> *It is good at answering questions, and it is very polite* (P132)

> *[...] the chatbot also responded thank you when I said thanks for the answer. This makes it more human and may induce trust* (P51)

***Absence of marketing* (4.1%).** Some participants accentuated that the appearance of the chatbot as objective and not as part of a marketing initiative was important for trust. These participants valued the ability to get answers directly related to their question rather than being led to other issues or possible offers. *Absence of marketing* is reflected in the following statements:

> *[...] I trust that the chatbot answers me objectively. No buying pressure :-)* (P10)

> *Provides information that is experienced fact-based, not normative or selling* (P104)

Environment-related categories

Three environment-related categories emerged during the thematic analysis: *brand*, *risk*, and *complementarity*.

***Brand* (17.1%).** The participant reports strongly suggested that their trust in chatbots was dependent on their trust in the service provider, that is, their brand perception. This implies that customers' previously established positive relationship with the brand may be crucial to their trust in chatbots for customer service. The participants reported that they believed the brand, which they trusted, had done a good job in implementing the chatbot. They also

mentioned that they perceived the brand as high quality and as offering only secure solutions. Examples of participant reports coded for this category are:

> *I'm trusting the service provider I contact, and then I also expect the chatbot to reflect that credibility* (P154)

> *I'm trusting the ones who have programmed the chatbot to gives it good input and, in that way, give good answers* (P105)

***Risk* (6.2%).** For some participants, trust depended on a perceived low risk when seeking support through the chatbot. For example, the participants noted that, since they were not required to provide personal or sensitive information in the conversation, the risk level was low, which made it easier to trust the chatbot. Some participants reported that risk would be perceived as higher for more personal questions, which could negatively affect their willingness to trust the chatbot. Others reported that information about how security was managed in the chatbot would strengthen trust. Furthermore, some participants expressed satisfaction with being reminded not to provide sensitive information when not being in a private chat. The following statements illustrate responses from this category:

> *When I contacted the chatbot, I had only a general question, which did not concern me personally. The trust was high since the case wasn't directly personal; the case could have been different if my need was of a more personal character* (P60)

> *Because I do not specify any personal information* (P57)

***Complementarity* (4.8%).** The participant reports indicated that some viewed trust as dependent on the chatbot for customer service being a complementary part of an integrated customer support systems where human operators were never far away. These participants specifically reported to value the opportunity to be transferred to a human operator if the

chatbot could not help, and they expressed that this would be important for their trust in a customer service chatbot. Below are statements exemplifying this category:

> *[…] and that the answers refer to relevant URLs and customer service if necessarily* (P90)

> *[…] and transferred to a human when required* (P65)

User-related categories

User-related categories represent individual differences in factors impacting users' trust in chatbots for customer service. The only category appearing in this group was *not trust relevant/no trust*.

***Not trust relevant/no trust* (7.5%).** Some participants expressed scepticism about relating the concept of trust to the use of chatbot technology. These reported not to consider trust as relevant or not to trust chatbots in their current level of maturity. In consequence, some participants reported to see chatbots as having limited usefulness and stated they would rather contact human customer service. Examples of answers coded to this category are:

> *Trust is a concept I will not use on not-living things. But I think that "easy" tasks, like in my case finding an IBAN-number (International Bank Account Number), can be well suited for chatbots* (P8)

> *I do not have trust in this chatbot. I want to talk to humans. I do not like robots, and I want to have human contact* (P21)

6   Discussion

The objective of this study was to identify factors that may explain users' trust in chatbots and to establish an initial model for predicting variation in such trust. We pursued the objective through a dual approach involving a quantitative explanatory investigation of factors assumed

to be of relevance for users' trust in the general literature on interactive systems and a qualitative exploration of factors particular to chatbots for customer service. This dual approach provided findings on two levels: (a) findings regarding the degree to which factors from the general literature explained variation in users' trust in chatbots for customer service and (b) findings regarding other factors of potential relevance for such trust. In the following, we discuss the findings and use these as basis for proposing our initial model for trust in chatbots for customer service.

6.1 A chatbot perspective on factors from the trust literature

The importance of expertise

Among the factors suggested by Corritore et al. (2003), we identified *expertise* as the most important factor for explaining trust in customer service chatbots. In the explanatory part of the study, expertise was the factor contributing most to explaining variation in *trust* in the multiple regression model. Furthermore, expertise was also the most frequently applied coding category for the participant reports in response to the open-ended free-text questionnaire item. Clearly, whether the chatbot is seen as providing answers reflecting knowledge, competence, and experience is highly important to trust. The importance of expertise is hardly a surprise. Expertise and related constructs such as ability, competence, and knowledge have long been discussed as critical to trust and perceived trustworthiness (e.g. Butler, 1991). Muir and Moray (1996) argued that trust in automated machinery mainly is a consequence of the users' perceptions of the expertise of a machine. Likewise, Fogg and Tseng (1999) regarded expertise as key to computer credibility. Also in the domain of interpersonal trust, researchers have accentuated expertise as a key determinant of trust (Mayer et al., 1995).

Our findings, however, suggest that expertise has particular connotations in the context of chatbots for customer service. In particular, the exploratory part of the study helped to

identify key aspects of expertise for such chatbots. Indeed, chatbot expertise concerns the provision of accurate and relevant information; in short, a correct answer. This understanding resembles definitions of expertise (e.g. Corritore et al., 2003) that accentuate competence, experience, and knowledge as defining characteristics. The importance of correct answers complies with previous findings that have identified the perceived ability of software agents as a strong predictor of trust (Detweiler & Broekens, 2009) and have shown that automation errors negatively affect trust (Merritt & Ilgen, 2008) - more so than do similar errors from human experts (Dietvorst, Simmons, & Massey, 2015).

However, the exploratory analysis also found that chatbot expertise is dependent on three other aspects: the chatbot's ability to provide concrete answers, to answer in an eloquent manner, and to correctly interpret the user's requests. That is, the answers not only have to be correct, but they also need to adequately fulfil the users' needs and expectations for a good answer. This finding supports the indications reported in the small-scale interview study by Følstad et al. (2018), which suggested trust in a chatbot was determined by the chatbot's interpretation of requests, its self-presentation, and its professional appearance.

One possible reason that perceived expertise in chatbots consists of adequately adapting the communication to fit the user's needs could be that current chatbots for customer service may not yet boast fully adequate communication skills. Rather, users may experience chatbots that misinterpret their requests, that provide partial answers to their requests, and that are presented in a style that is not sufficiently adapted to the dialogue context (Luger & Sellen, 2016).

Strengthening users' perceptions of expertise in customer service chatbots will be important to service providers. To achieve this, service providers may need to consider not only the correctness of a chatbot's answers, but also whether the answers fit the users' needs and are appropriate in the specific contexts of their interactions. From the general literature on

customer service, it is well known that effective and efficient resolving of users' issues is critical to their satisfaction both with customer service, and also with the service provider (Dixon et al., 2010). The perceived expertise of a chatbot is likely to have similar effects on customer satisfaction.

The effect of risk perceptions

The multiple regression analysis also suggested *risk* as a significant predictor for users' trust in chatbots for customer service. Researchers have reached substantial agreement on the importance of perceived risk for trust (e.g. Corritore et al., 2003; Rousseau, Sitkin, Burt, & Camerer, 1998); some also argue that trust is not relevant in situations not characterized by risk (e.g. Mayer et al., 1995). Because chatbots are an emerging technology, it should be no surprise that an increase in users' perceptions of risk regarding chatbot interactions may lead to a decrease in trust. When the risk of something going wrong increases, users tend to drift towards well-tried alternatives.

The findings from the exploratory analysis also supported the findings from the explanatory analysis. Though relatively few users (6%) reflected on risk levels as related to trust, the users who did specifically noted that the low risk associated with current customer service chatbots made them easier to trust. For service providers, this finding may offer important insight. Current chatbots for customer service typically have limited functionality; specifically, many simply provide answers to frequently asked questions rather than engaging in personal customer service. Additionally, interacting with a chatbot may not require users to log in to their personal accounts. This limitation in current chatbots is in large part due to immaturity in technology integration. However, our findings suggest that this limitation may be beneficial in terms of providing an opportunity for users to familiarize themselves with chatbots for customer service in a low-risk environment.

When maturing technology allows for more personalized and potentially sensitive interaction with chatbots, it may be important for service providers not to introduce features that increase risk perceptions faster than users are ready for. It is also noteworthy that participants in the exploratory part argued for the benefit of transparency concerning security. This suggestion aligns with participants' concern for security and privacy as determinants of trust (Følstad et al., 2018).

Individual differences in trust

The general literature on interactive systems has revealed that users differ in their general propensity to trust technology (e.g. Muir & Moray, 1996). Unsurprisingly, our study replicated this finding. In the explanatory analysis, we found that users' propensity to trust technology significantly contributed to explaining variation in trust. Likewise, in the exploratory analysis, we interpreted a small proportion of user responses to reflect non-trusting sentiment towards chatbots.

This individual variation in propensity to trust is highly noteworthy for both research and practice. Chatbots are an emerging technology that has not yet enjoyed widespread popular uptake. Hence, users of current chatbots will often have relatively little experience in interacting with computers through natural language dialogue. This is, for example, reflected in the level of experience of our participants: More than 50% of our sample had used chatbots only 1–3 times. Hence, at present we may be in a period in which early adopters (Rogers, 2010) constitute a larger share of the chatbot user population than would be the case with a more established type of interactive system. In consequence, the current chatbot user population may have different characteristics than the user population when chatbots have reached widespread uptake. The current user population includes individuals who are likely more characterized by openness and willingness to take risks than would the user population of future chatbots.

While the chatbot user population may be relatively trustful and willing to take risks in their explorations of chatbots for customer service, later chatbot user populations may not exhibit the same characteristics. In consequence, it will likely be important for service providers to pay close attention to reducing perceived risk in future chatbot versions for users who are more risk-averse.

An overrepresentation of innovators and early adopters in the current chatbot user population would also explain the lack of effects on trust found for age and gender in our study. For future user populations, age and gender effects may be more prevalent, as they have been for other aspects of user experience demonstrated in more controlled studies of chatbots (e.g. Shah, Warwick, Vallverdú, & Wu, 2016).

Is human-likeness important to trust?

Nass, Moon, and Carney (1999) fond that users follow social convention, for example by applying politeness, when interacting with machines. The literature has also suggested that users tend to bestow human characteristics on chatbots, given their human-likeness following from their natural language capabilities (Candello et al., 2017; Mone, 2016). In consequence, we considered the perceived human-likeness of chatbots as a likely predictor of users' trust in chatbots and included this as one of our seven factors.

To some degree, *human-likeness* seems to explain variation in trust. It was a less important predictor in our multiple regression model, only bordering significance. Likewise, only 6% of the participants who responded to the open-ended free-text item mentioned human-likeness as a possible factor contributing to trust. Hence, while human-likeness may have some impact on trust, our findings suggest that it hardly has the influence that chatbots' humanlike character might suggest. Possibly, human-likeness may be more influential for other aspects of user experience besides trust. This finding resonates with the study by Følstad et al. (2018), as

their interview participants reflected a mixed set of views on the potential role of human-likeness as a driver of trust in chatbots.

The importance of other factors from the literature

In addition to the factors discussed above, our initial set of factors from the literature also included some that were found to be of secondary importance for explaining variation in trust. Specifically, this concerned *ease of use*, *predictability*, and *reputation*, which we will briefly discuss below.

In the explanatory analysis, participants provided high ratings for *ease of use*, indicating that users find chatbots for customer service easy to use. This ease is likely because users typically are familiar with chat user interfaces and see dialogue in natural language as easy and convenient. The low variation in users' assessments of ease of use makes this a less effective differentiator for trust. Quite possibly, perceptions of low ease of use would be detrimental to trust. However, as this does not currently seem to be an issue, ease of use may be of lesser importance for explaining trust in current chatbots for customer service. That said, one could argue that the interpretational issues discussed for perceived expertise in the chatbot may be seen as reflecting issues pertaining to ease of use. Note, however, that *ease of use* here is construed as concerning users' perceptions of the interactive system and its user interface, not a more comprehensive understanding of ease of use as incorporating the entire construct of usability or user experience.

We also found that *predictability* contributed little to explain variation in trust. Predictability was substantially correlated with expertise ($r = .61$, $p < .001$), and hence did not make an individual contribution to the regression model. Therefore, predictability in this context seems redundant to the expertise construct. Furthermore, as nothing resembling the predictability construct appeared as a distinct category in the exploratory analysis of the participants' free-

text responses, it seems adequate to focus our attention on perceived expertise rather than perceived predictability when considering users' trust in chatbots for customer service.

Finally, *reputation* did not contribute independently to explain trust in the regression model. Reputation was construed as the users' impression of the interactive system based on assessments or judgements made by others. The chatbot is an emerging technology with relatively low uptake in the intended user population, which might explain why this factor did not appear relevant. Likely, many users had not heard about the chatbots they were using from others. Possibly, reputation may be more important to trust in chatbots in the future, when chatbots have a more widespread uptake.

6.2   Factors not initially identified in the literature

In the exploratory part of our study, we aimed to identify factors of relevance for trust in chatbots beyond those suggested by existing literature. The most noteworthy of these were *brand*, *responsiveness*, *complementarity*, and *absence of marketing*.

Brand perception

The exploratory analysis revealed users' brand perceptions to be important for trust. About one-sixth of the participants mentioned this category in their free-text responses; in fact, only *expertise* was more frequently reported. Brand perception concerns the environment of the chatbot; that is, it concerns the user's experiences with and sentiment regarding the service provider. While online trust have also been proposed as a determinant of brand trust in the general literature (Ha, 2004), the findings in our study clearly suggested the opposite relationship: The participants mentioning brand perception typically reported that their trust in the chatbot was due to their established trust in the service provider.

Brand perceptions may be related to the reputation construct from Corritore et al. (2003). However, whereas Corritore et al. considered reputation as pertaining to the past performance

of the interactive system, brand perception concerns the service provider. In this sense, the construct of brand perception may be closer to reputation as defined by other researchers. For example, Jarvenpaa, Tractinsky, and Saarinen (1999) considered trust in an e-commerce context to be dependent on the service provider's reputation, rather than its website in isolation.

For service providers, it is clearly relevant to note the importance of brand perception to explain variation in trust. In particular, this suggests the potential benefit of clearly associating the customer service chatbot with a trusted brand.

Responsiveness

A new category emerging from the thematic analysis was users' perceptions of responsiveness. Participants likely highlighted this factor because they see it as critical to trust that they perceive the chatbot as an effective way to get help. This echoes Brandtzaeg and Følstad (2017) who found productivity to be the most expressed reason for humans to take advantage of chatbot technology. They argued that most chatbot users seek quick and consistent feedback when they need assistance or help. In the customer service domain, this need makes sense. Some participants' statements also noted that the chatbot was a faster option than human customer support.

Also concerning responsiveness, some participants mentioned the immediate chatbot response as an indicator of the chatbot being a robot. Some even suggested that a brief delay in the answer could be beneficial for users who wish the chatbots to have more humanlike behaviour.

Other factors not initially identified in the literature

In addition to brand perception and responsiveness, the exploratory analysis suggested two other factors to be of relevance for trust in chatbots: *complementarity* and *absence of*

marketing. However, fewer than 5% of the participants mentioned these issues, which could indicate that they are of less importance to trust than the factors already discussed.

Complementarity concerns users' wish to be transferred to a human if necessary. Absence of marketing concerns users' desire to be provided help and support without marketing intentions. While potentially important for some, these factors do not seem to be among the key factors that service providers and chatbot developers should consider with regard to how chatbots are presently implemented. However, if chatbots in the future fully replace rather than complement customer service by humans, or if they serve as tireless marketers rather than supporting users in reaching their goals, these two factors may become more important.

6.3   An initial model of trust in chatbots for customer service

On the basis of our study findings, we propose an initial model of trust in chatbots for customer service. This initial model includes those factors from the literature shown in the explanatory analysis to explain the most variance in trust and the most frequently mentioned factors from the exploratory study. Following the structure applied earlier in the study, we group these factors as (a) chatbot-related, (b) environment-related, and (c) user-related.

Chatbot-related factors

If only one factor were to be chosen to explain trust in chatbots for customer service, it would be *expertise*. This was by far the factor explaining most variation in trust in the regression model of the explanatory analysis. Furthermore, the exploratory analysis revealed expertise to be the one factor that users typically considered to determine their trust in the customer service chatbot they had just used. Therefore, we suggest that perceived expertise may not only covary with trust, but that it may determine trust.

In addition to expertise, the exploratory analysis revealed *responsiveness* to be a potentially important factor in explaining trust. Nearly one-fifth of the respondents made reports coded as

alluding to the responsiveness category. Responsiveness is a potential benefit of chatbots for customer service, as compared to customer service from human operators. This responsiveness could possibly be used to strengthen trust in customer service systems where chatbots and humans closely collaborate, for example with chatbots as a first line of support. While responsiveness was not included among the factors drawn from the literature, we find the findings from the explanatory analysis to be sufficiently substantial so as to suggest this to be included as a factor in future explanatory analyses.

Other chatbot-related factors considered relevant for inclusion in the initial model included *ease of use*, *predictability*, *human-likeness*, and *absence of marketing*. While these factors may be relevant for some, they were found to be of lesser importance than the two chatbot-related factors included in the conducted analyses. As a result, we have left them out of this initial model. However, we expect that they may nevertheless be found relevant in other contexts, such as in a future situation when chatbots for customer service are more commonplace and the user group differs from the population in this study.

Environment-related factors

We found two environment-related factors (*risk* and *brand*) to be highly relevant for an initial model of trust in chatbots for customer service. We identified risk as relevant on the basis of both the explanatory and exploratory analysis, while we identified brand as relevant based on the exploratory analysis alone. In the exploratory analysis, the participant reports reflected that these factors were both seen as drivers of trust. We consider the findings from the exploratory analysis to be sufficiently substantial so as to suggest brand to be included in future explanatory analyses. However, preceding such future analyses, work will be needed to develop adequate measurement instruments both for *brand* as well as for *responsiveness* discussed above.

Other environment-related factors that we considered for inclusion in the initial model were *reputation* and *complementarity*. However, we determined they were not sufficiently prominent to be included. Reputation, which concerns the perceptions of the interactive system on the basis of others' assessments or experiences, may indeed be more important in the future when chatbots are taken up by the mass market as a key channel for customer service. Complementarity, or the use of chatbots as part of an integrated system for customer service in which human support personnel are also easily available, may also be more important if the implementation of chatbots for customer service is at the expense of necessary contact with human personnel.

User-related factors

We identified only one user-related factor in the study: *propensity to trust technology*. Both the explanatory and exploratory analyses showed the importance of this factor. It may be assumed that this factor is a determinant of trust, as it is a general characteristic of the user.

Additionally, we see individual variation in propensity to trust technology as potentially important for the changes that may be foreseen for a chatbot trust model. Specifically, we suggest that, as the character of the chatbot user population evolves, there may be a need to update the trust model.

An initial model

To summarize the above, Figure 1 presents our proposal for an initial model of trust in chatbots for customer service.
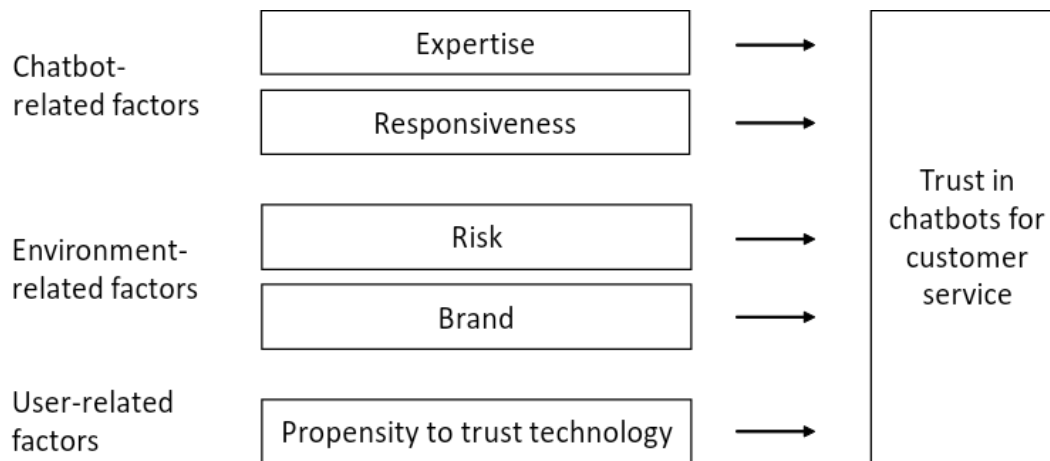
Figure 1: An initial model of users' trust in chatbots for customer service

We consider this model *initial*, as we foresee that the factors related to users' trust in chatbots may change as chatbots become more commonplace. Currently, a much smaller proportion of users of customer service are chatbot users than what is expected for the future. Given that individual user characteristics in part determine who has already at this point become a chatbot user, it is likely that the characteristics of the current and the future population of chatbot users do not fully overlap. For example, in line with Rogers' (2010) theory on the diffusion of innovation, it may be that the current population of chatbot users has a relatively larger proportion of early adopters than what may be expected in the future population of chatbot users,. These demographics will likely change as chatbots are taken up by larger user groups; in consequence, the factors relevant for users' trust in chatbots may change. Furthermore, as suggested by the high proportion of participants in our sample having relatively little experience with chatbots, users' perceptions and experiences of this technology will likely change over time as they become more experienced.

As the maturity of chatbots for customer service and their uptake increases, we anticipate increases in the importance of factors such as reputation and complementarity. It is also conceivable that some factors will reduce their importance; for example, the importance of expertise may be reduced in a future situation where most chatbots for customer service are

seen as holding reasonable levels of quality in their interpretation of and responses to users' requests.

While we anticipate future changes to the trust model, we nevertheless see this initial model as a valuable contribution to chatbot research and practice. The identified factors will be useful as a basis for future research. Furthermore, the future development and successful diffusion of chatbots for customer service will depend on adequate knowledge regarding users' trust.

6.4   Limitations and future research

As the presented study is an early attempt at closing the knowledge gap on users' trust in chatbots, some limitations are to be expected. In the following, we will discuss four limitations and point out opportunities for future research.

First, while a questionnaire study allowed us to gather data from a substantial number of users, this data collection method also carries some limitations. In particular, the participants' responses may diverge from their behaviour. Furthermore, this questionnaire study was structured such that the exploratory analysis was merely correlational, so it cannot sustain claims on causal relations. While the exploratory part of the study provided some insight into the participants' perceptions of factors determining trust, future research is needed to fully establish causal relations. One relevant avenue of future research would be to design studies for investigating causal relations between assumed determinants of trust in chatbots for customer service. Such studies could be set up either as longitudinal studies or as experimental designs.

Second, the participant sample had important limitations. Most participants had little experience with chatbots and had tested chatbots only 1–3 times before. While most users are relatively novice in the current chatbot context of use, it will likely be necessary to replicate

this kind of study and update the initial trust model when chatbots have been established as commonplace technology. An interesting string of future research would be to follow the developments of chatbot uptake and associated changes in user experience and trust perceptions.

Third, the measurement instruments used in this study were not perfect. Prior to our study, no measurement instruments existed for the study of trust in chatbots and related factors. Hence, we established the applied measurement instruments by combining and adjusting scales from other instruments. The exploratory factor analysis revealed some instances of cross-loading and low factor loading. For future research, it could be beneficial to further validate the applied measurement instruments.

Fourth, this study is limited as it concerned only the Norwegian market and a small number of chatbots. Given the characteristics of this market, in particular its maturity in terms of general mobile technology uptake as well as chatbot applications, findings made in this market likely are relevant for similar markets. Furthermore, while only four chatbots were used for recruiting the study participants, the fairly high scores given to these chatbots in terms of, for example, expertise and ease-of-use, suggests that the findings may be representative for chatbots that provide adequate customer service - which is the aim of any chatbot for customer service. Nevertheless, it would be beneficial for future research to address trust in chatbots for customer service also for other markets. In particular, it may be relevant to conduct such studies in markets with other cultural characteristics, as trust and culture is seen as closely related (Bensabat, Gefen, & Pavlou, 2008)

6.5   Conclusion

This study has provided a contribution to address the current gap in the research literature about users' perceptions of factors related to trust in chatbots for customer service. On the

basis of our study, we have proposed an initial model of users' trust in chatbots for customer service. Given the limitations of the presented study, as well as the foreseen changes in the development and uptake of chatbots, we consider our model only a starting point for the research and development needed to establish chatbots as a trusted and preferred channel for customer service. This development will be highly valuable, as natural language interaction with computer systems may represent benefits to both users and service providers in terms of access and efficiency in customer service.

References

Abu Shawar, B., & Atwell, E. (2007). Chatbots: Are they really useful? *Journal for Language Technology and Computational Linguistics*, *22*(1), 29–49.

Accenture. (2016). *Chatbots in customer service*. Retrieved from https://www.accenture.com/t00010101T000000__w__/br-pt/_acnmedia/PDF-45/Accenture-Chatbots-Customer-Service.pdf

Benbasat, I., Gefen, D., & Pavlou, P. A. (2008). Trust in online environments. *Journal of Management Information Systems, 24*(4), 5-11.

Brandtzaeg, P. B., & Følstad, A. (2017). Why people use chatbots. In I. Kompatsiaris, J. Cave, A. Satsiou, G. Carle, A. Passani, E. Kontopoulos, S. Diplaris, & D. McMillan (Eds.), *Internet Science: 4th international conference, INSCI 2017* (pp. 377–392). Cham, Switzerland: Springer.

Butler, J. K. (1991). Toward understanding and measuring conditions of trust: Evolution of a conditions of trust inventory. *Journal of Management*, *17*(3), 643–663.

Candello, H., Pinhanez, C., & Figueiredo, F. (2017). Typefaces and the perception of humanness in natural language chatbots. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 3476–3487). New York, NY: ACM.

Cheung, C. M., & Lee, M. K. (2001). Trust in Internet shopping: Instrument development and validation through classical and modern approaches. *Journal of Global Information Management*, *9*(3), 23–35.

Coniam, D. (2014). The linguistic accuracy of chatbots: Usability from an ESL perspective. *Text & Talk*, *34*(5), 545–567.

Corritore, C. L., Kracher, B., & Wiedenbeck, S. (2003). On-line trust: Concepts, evolving themes, a model. *International Journal of Human-Computer Studies*, *58*(6), 737–758.

Corritore, C. L., Marble, R., Wiedenbeck, S., Kracher, B., & Chandran, A. (2005). *Measuring on-line trust of websites: Credibility, perceived ease of use, and risk*. In *Proceedings of the Americas Conference on Information Systems - AMCIS 2005*, article 370. Retrieved from https://aisel.aisnett.org/amics2005/370

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

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319–340.

De Matos, C. A., Henrique, J. L., & Rossi, C. A. V. (2007). Service recovery paradox: A meta-analysis. *Journal of Service Research, 10*(1), 60-77.

De Visser, E. J., Monfort, S. S., McKendrick, R., Smith, M. A., McKnight, P. E., Krueger, F., & Parasuraman, R. (2016). Almost human: Anthropomorphism increases trust

resilience in cognitive agents. *Journal of Experimental Psychology: Applied*, *22*(3), 331.

Detweiler, C., & Broekens, J. (2009). Trust in online technology: Towards practical guidelines based on experimentally verified theory. In J. A. Jacko (Ed.), *Human-computer interaction. Ambient, ubiquitous and intelligent interaction* (pp. 605–614). Berlin, Germany: Springer.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, *144*(1), 114–126.

Dixon, M., Freeman, K., & Toman, N. (2010). Stop trying to delight your customers. *Harvard Business Review*, *88*(7/8), 116–122.

Ezzy, D. (2002). *Qualitative Analysis: Practice and Innovation*. London, UK: Routledge.

Fogg, B. J., & Tseng, H. (1999). The elements of computer credibility. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 80–87). New York, NY: ACM.

Følstad, A., Kvale, K., & Haugstveit, I. M. (2014). Customer support as a source of usability insight: Why users call support after visiting self-service websites. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction* (pp. 167–170). New York, NY: ACM.

Følstad, A., Nordheim, C. B., & Bjørkli, C. (2018). What makes users trust a chatbot for customer service? An exploratory interview study. In *Proceedings of the 5th International Conference on Internet Science, INSCI 2018*. Cham, Switzerland: Springer.

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, *114*, 254–280.

Grand View Research. (2017). *Chatbot market size to reach $1.25 billion by 2025*. Retrieved from https://www.grandviewresearch.com/press-release/global-chatbot-market

Griffith, E., & Simonite, T. (2018). *Facebook's virtual assistant M is dead. So are chatbots*. Retrieved from https://www.wired.com/story/facebooks-virtual-assistant-m-is-dead-so-are-chatbots/

Ha, H. Y. (2004). Factors influencing consumer perceptions of brand trust online. *Journal of Product & Brand Management*, *13*(5), 329–342.

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., de Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors*, *53*(5), 517–527.

Hardin, J. (2003). Gaming trust. In E. Ostrom & J. Walker (Eds.), *Trust and reciprocity: Interdisciplinary lessons from experimental research* (pp. 84–106). New York, NY: Russell Sage Foundation.

Hill, J., Ford, W. R., & Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. *Computers in Human Behavior*, *49*, 245–250.

Ho, C. C., & MacDorman, K. F. (2010). Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices. *Computers in Human Behavior*, *26*(6), 1508–1518.

Holtgraves, T. M., Ross, S. J., Weywadt, C. R., & Han, T. L. (2007). Perceiving artificial social agents. *Computers in Human Behavior*, *23*(5), 2163–2174.

Jarvenpaa, S. L., Tractinsky, N., & Saarinen, L. (1999). Consumer trust in an Internet store: A cross-cultural validation. *Journal of Computer-Mediated Communication*, *5*(2), 45–71.

Jian, J.-Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, *4*(1), 53–71.

Leviathan, Y., & Matias, Y. (2018, May 8). Google Duplex: An AI system for accomplishing real-world tasks over the phone [Blog post]. Google AI Blog. Retrieved from https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html

Luger, E., & Sellen, A. (2016). 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* (pp. 5286–5297). New York, NY: ACM.

Madsen, M., & Gregor, S. (2000). *Measuring human-computer trust.* Paper presented at the 11th Australasian Conference on Information Systems. Retrieved from https://pdfs.semanticscholar.org/b8ed/a9593fbcb63b7ced1866853d9622737533a2.pdf

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *The Academy of Management Review*, *20*(3), 709–734.

McKnight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems*, *2*(2), article no. 12.

Merritt, S. M., & Ilgen, D. R. (2008). Not all trust is created equal: Dispositional and history-based trust in human-automation interactions. *Human Factors, 50*(2), 194-210.

Meuter, M. L., Bitner, M. J., Ostrom, A. L., & Brown, S. W. (2005). Choosing among

    alternative service delivery modes: An investigation of customer trial of self-service

    technologies. *Journal of Marketing*, *69*(2), 61–83.

Meuter, M. L., Ostrom, A. L., Roundtree, R. I., & Bitner, M. J. (2000). Self-service

    technologies: Understanding customer satisfaction with technology-based service

    encounters. *Journal of Marketing*, *64*(3), 50–64.

Mone, G. (2016). The edge of the uncanny. *Communications of the ACM*, 59(9), 17–19.

Muir, B. M., & Moray, N. (1996). Trust in automation. Part II. Experimental studies of trust

    and human intervention in a process control simulation. *Ergonomics*, *39*(3), 429–460.

Murgia, A., Janssens, D., Demeyer, S., & Vasilescu, B. (2016). Among the machines:

    Human-bot interaction on social Q&A websites. In *Proceedings of the 2016 CHI*

    Conference Extended Abstracts on Human Factors in Computing Systems (pp. 1272–

    1279). New York, NY: ACM.

Nass, C., Moon, Y., & Carney, P. (1999). Are people polite to computers? Responses to

    computer-based interviewing systems. *Journal of Applied Social Psychology*, *29*,

    1093–1110.

Ostrom, A. L., Parasuraman, A. Bowen, D. E., Patrício, L., & Voss, C. A. (2015). Research

    priorities in a rapidly changing context. *Journal of Service Research*, *18*(2), 127–159.

Rogers, E. M. (2010). *Diffusion of Innovations* (5th ed.). New York, NY: Free Press.

Rousseau, D., Sitkin, S., Burt, R., & Camerer, C. (1998). Not so different after all: A cross-

    discipline view of trust. *Academy of Management Review*, *23*(3), 393–404.

Servion. (2017). *AI will power 95% of customer interactions by 2025*. Retrieved from

    http://servion.com/blog/ai-will-power-95-customer-interactions-2025/

Shah, H., Warwick, K., Vallverdú, J., & Wu, D. (2016). Can machines talk? Comparison of Eliza with modern dialogue systems. *Computers in Human Behavior*, *58*, 278–295.

Statistics Norway (2018). ICT usage in households. Use of smartphone and security related to this (per cent), by sex, age, contents and year. Retrieved from https://www.ssb.no/en/statbank/table/12344/tableViewLayout1/

Tezcan, T., & Zhang, J. (2014). Routing and staffing in customer service chat systems with impatient customers. *Operations Research*, *62*(4), 943–956.

Van Lange, P. A. M. (2015). Generalized trust: Four lessons from genetics and culture. *Current Directions in Psychological Science*, *24*(1), 71–76.

Weizenbaum, J. (1966). ELIZA - A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *9*(1), 36–45.

Xu, A., Liu, Z., Guo, Y., Sinha, V., & Akkiraju, R. (2017). A new chatbot for customer service on social media. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 3506–3510). New York, NY: ACM.

Zumstein, D., & Hundertmark, S. (2017). Chatbots - An interactive technology for personalized communication, transactions and services. *IADIS International Journal on WWW/Internet*, *15*(1), 96–109.

Appendix 1

Questionnaire items used in the explanatory part of the study. Three items were removed from the final set; these are presented in italics and square brackets.

Variable	Original items and their source	Items applied in the study
Trust	I believe this website is trustworthy (Corritore et al., 2005)	I experience this chatbot as trustworthy
	I believe this website will not act in a way that harms me (Corritore et al., 2005)	I do not think this chatbot will act in a way that is disadvantageous for me
	I am suspicious of the system's intent, action, or outputs (Jian et al., 2000)	I'm suspicious of this chatbot
	The system is deceptive (Jian et al., 2000)	The chatbot appears deceptive
	I trust this website (Corritore et al., 2005)	I trust this chatbot
Expertise	Self-composed	I experienced to get my question answered
	The website content reflects mastery of knowledge (Corritore et al., 2005)	The chatbot appears knowledgeable
	The website content reflects expertise (Corritore et al., 2005)	The content of the chatbot reflects expertise
	I feel very confident about top management's skills (Mayer et al., 1995)	I feel very confident about the chatbot's competence
	Top management is very capable of performing its job (Mayer et al., 1995)	The chatbot is well equipped for the task it is set to do
Predictability	The website content is what I expected (Corritore et al., 2005)	The chatbot behaves predictably
	There were no surprises in how the website responded to my actions (Corritore et al., 2005)	There were no surprises in how the chatbot answered me
	The website is what I anticipated (Corritore et al., 2005)	The chatbot behaved as predicted
	I find it predictable that the website has the type of content it does (Corritore et al., 2005)	I think it is predictable that the chatbot has the type of content is does
	The website content is predictable (Corritore et al., 2005)	The content of the chatbot was as expected
Human-likeness	Fake – Natural (Ho & MacDorman, 2010)	The chatbot is natural
	Machinelike – Humanlike (Ho & MacDorman, 2010)	The chatbot is humanlike
	Artificial – Lifelike (Ho & MacDorman, 2010)	The chatbot is realistic
	Unconscious – Conscious (Ho & MacDorman, 2010)	The chatbot is present
	Artificial – Lifelike (Ho & MacDorman, 2010)	The chatbot is authentic
Ease of use	Learning to operate this website was easy for me (Corritore et al., 2005)	It was easy for me to learn how to use this chatbot

	[I found it easy to get this website to do what I wanted it to do (Corritore et al., 2005)]	*[I find it easy to get the chatbot to do what I want it to do.]*
	I found the website easy to use (Corritore et al., 2005)	The chatbot is easy to use
	My interaction with chart-master would be clear and understandable (Davis, 1989)	My dialogue with the chatbot was clear and understandable
	[I would find chart-master to be flexible to interact with (Davis, 1989)]	*[This chatbot is flexible to interact with]*
Risk	I feel vulnerable when I interact with this website (Corritore et al., 2005)	I feel vulnerable when I interact with this chatbot
	I believe that there could be negative consequences from using this website (Corritore et al., 2005)	I think there could be negative consequences when using this chatbot
	I feel it is unsafe to interact with this website (Corritore et al., 2005)	I feel it is unsafe to talk to this chatbot
	I feel I must be cautious when using this website (Corritore et al., 2005)	I feel I must be cautious when I use this chatbot
	It is risky to interact with this website (Corritore et al., 2005)	I feel there is risk involved in talking to this chatbot
Reputation	This store is well known (Jarvenpaa et al., 1999)	The chatbot is well known by others
	The website is highly regarded (Corritore et al., 2005)	The chatbot has a good reputation
	The website is respected (Corritore et al., 2005)	The chatbot is respected by others
	The website has a good reputation (Corritore et al., 2005)	I have heard others talking positively about this chatbot
	[This store has a bad reputation in the market (Jarvenpaa et al., 1999)]	*[I have heard others talking positively about this chatbot]*
Propensity to trust technology	My typical approach is to trust new technologies until they prove to me that I shouldn't trust them (McKnight et al., 2011)	My typical approach is to trust new technology
	I usually trust a technology until it gives me a reason not to trust it (McKnight et al., 2011)	I generally trust new technology until it gives me a reason not to
	I generally give a technology the benefit of the doubt when I first use it (McKnight et al., 2011)	Even under doubt, I will choose to trust new technology
	It is easy for me to trust a person/thing (Cheung & Lee, 2001)	It is easy for me to trust new technology
	My tendency to trust a person/thing is high (Cheung & Lee, 2001)	My tendency to trust new technology is high

Note: All items applied in the survey were translated from Norwegian by the first author.

Appendix 2

Principal component analysis for the questionnaire items included in the explanatory analysis

Factor	1	2	3	4	5	6	7	8
Trust1	.48							.32
Trust2		-.30						.34
Trust3_RE			-.27				-.34	.45
Trust4_RE								.70
Trust5	.37							.26
Expertise1	.94							
Expertise2	.77							
Expertise3	.79					.27		
Expertise4	.69							
Expertise5	.81							
Predictability1				.75				
Predictability2				.79				
Predictability3	.27			.72			-.28	
Predictability4	.26			.58				
Predictability5	.59			.40				
Human-likeness1						.72		
Human-likeness 2						.84		
Human-likeness3						.80		
Human-likeness4						.77		
Human-likeness5						.82		
Ease of use1							.64	
Ease of use3			-.26				.65	
Ease of use4	.45						.52	
Risk1			.70					
Risk2			.78					
Risk3			.87					
Risk4			.90					
Risk5			.88					
Reputation1					.92			
Reputation2					.87			
Reputation3					.69			
Reputation4					.57	.29		-.31
Propensity to trust technology1		-.88						
Propensity to trust technology2		-.84						
Propensity to trust technology3		-.71						
Propensity to trust technology4		-.91						
Propensity to trust technology5		-.93						

Note. Rotation method: Oblimin with Kaiser normalization. Rotation converged in 11 iterations. Missing values excluded pairwise. Factor loadings < .25 suppressed. _RE signifies a reversed item.