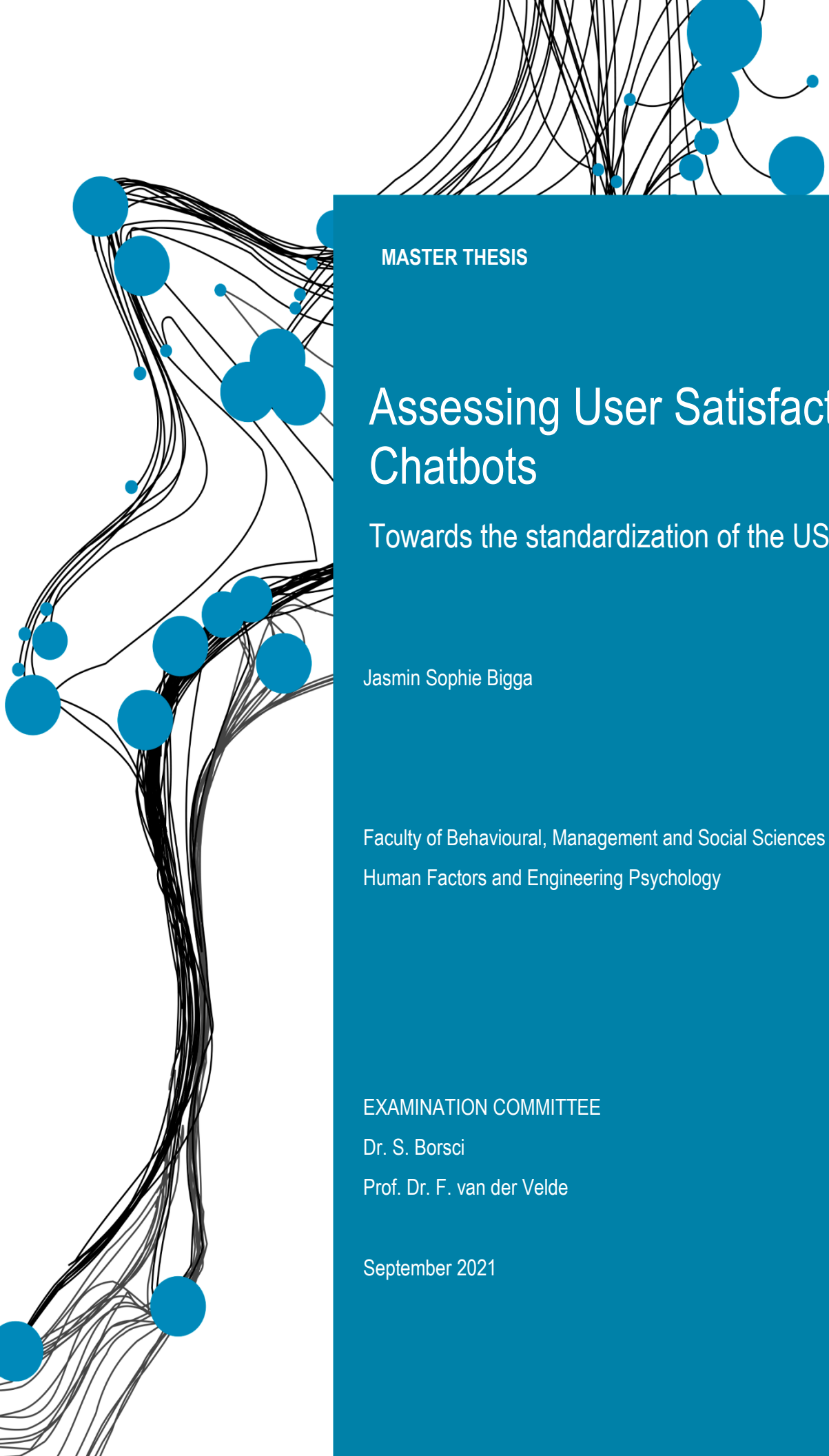


MASTER THESIS

Assessing User Satisfaction with Chatbots

Towards the standardization of the USIC scale

Jasmin Sophie Bigga

Faculty of Behavioural, Management and Social Sciences (BMS)

Human Factors and Engineering Psychology

EXAMINATION COMMITTEE

Dr. S. Borsci

Prof. Dr. F. van der Velde

September 2021

Abstract

Despite the growing demand for service chatbots, many of them fail to meet users' demands. To resolve this issue, developers need insight on aspects that influence user satisfaction with chatbots. However, user satisfaction measures for the context of chatbots are lacking. Addressing this challenge, Balaji and Borsci (2019) proposed the User Satisfaction with Information Chatbots (USIC) scale. Evidence for the reliability and validity of the USIC was gathered by several studies. However, the validity and reliability of the scale needs to be assessed repeatedly during the process of standardization, to gather evidence for the generalizability of the results. The current study replicated the usability study by Balaji and Borsci (2019). Participants interacted with five chatbots and completed the USIC and the UMUX-Lite after each interaction. Our results indicate a four-factor structure of the USIC, in line with previous work. Additionally, we examined the effect of age and affinity for technology on user satisfaction with chatbots, however, the results were non-significant. To increase the USICs applicability we reduced the scale by selecting the items with the strongest factor loadings, which resulted in a 14-item questionnaire with two latent factors. Concurrent validity of the USIC was indicated by the strong correlation with the standardized user satisfaction measure UMUX-Lite. Overall, our research provides further evidence that the USIC is a reliable tool to assess user satisfaction with chatbots and to guide developers in formulating clear design guidelines for these systems.

Keywords: Chatbots, user satisfaction, UMUX-Lite, reliability, validity

Table of contents

Introduction 3
What makes a good chatbot? The need for user satisfaction measures in human-computer interaction 4
Scale for user satisfaction with information chatbots (USIC) 6
Goal of the current study 7
Method 9
Ethical approval 9
Translation of the scales 9
Participants 9
Procedure 10
Materials 11
Data Analysis 12
Results 14
Data screening 14
Factor structure of the USIC scale 14
Scale reduction 18
Correlation USIC and UMUX-Lite 20
Effects of age on user satisfaction with chatbots 21
Effects of Affinity for Technology on satisfaction with chatbots 22
Discussion 23
Limitations and recommendations for future research 26
Conclusion 27
References 28
Appendices 32
Appendix A 32
Appendix B 38
Appendix C 39
Appendix D 44

Introduction

Chatbots are software applications that engage in some form of dialogue with a user through the use of natural language (Dale, 2016). They may either rely on text-based input or make use of speech recognition to engage in conversation with the user or to execute commands to fulfill tasks on behalf of the human user (Radziwill & Benton 2017).

Chatbots have shown to be of great use across different industries. One benefit is that chatbots can help reduce operational costs in customer services by up to 30% (Abbas, 2019). Statistics show that chatbots can handle around 80% of inquiries without the need for human intervention (Jovic, 2020). This reduces the need for manpower, as human agents are only needed for more complicated matters that go beyond the capabilities of the chatbot. Furthermore, as chatbots can address requests in real-time companies can reach more customers and avoid long waiting times which benefits customer satisfaction (LiveChat, 2021).

Aside from providing immediate solutions, chatbots can also provide a more personal experience compared to websites. Chatbots are highly interactive and therefore more flexible which makes it easy to tailor the experience to the user and provide them with exactly the information or product that they need, eliminating unnecessary information. Furthermore, users often tend to anthropomorphize and project (positive) feelings into their interaction with the chatbot (Kojouharov, 2018), creating possibilities for companies to shape the customers' perception of their brand and to create a more personal relationship with them. This might benefit the number of sales, as according to Derksen (2016), the majority of consumers (75%) is more likely to buy from retailers that offer some form of personalization.

Chatbots can also carry out predictive analyses, which allows companies to jump in with a service when a customer might need it. The American hotel chain *Roof Inn* let their chatbot software analyze flight and weather data, in order to be able to predict whether potential customers were facing flight cancellations (Kojouharov, 2018). Based on these analyses, services were then offered to mobile phone users in rough weather regions, to adjust to their newly emerged need for a hotel room. Targeted marketing through predictive analyses is therefore of a compelling competitive value as potential customers may be reached faster compared to traditional marketing methods (Kojouharov, 2018).

Research shows that chatbots are also well received by consumers. According to Press (2019) the acceptance of chatbots has doubled since 2018, with 83% of the consumers rating them as "*very helpful*". The majority of consumers (65%) feel confident in resolving issues without the involvement of a human agent (Zaboj, 2020), in fact, the use of chatbots is often preferred, as information can be

accessed quickly and immediate solutions are offered (Sweezey, 2019; Zaboj, 2020). Especially within the Millennial generation chatbots have been getting increasingly mainstream, with 60% of Millennials indicating that they already have interacted with chatbots (Press, 2019) and approximately 40% chatting with chatbots daily (Suthar, 2018).

Even though one might think of chatbots as a novel phenomenon, the first chatbot – ELIZA - was presented as early as 1966, way before the internet existed. The initial goal of the newly developed software was to mimic human conversation as well as possible, so the person on the other end would be fooled it would be talking to a real person, also known as the Turing test (Dale, 2016). However, these first chatbots appeared to be too inflexible to maintain a longer conversation, as they made use of simple keyword matching and therefore could not cope with the flexibility of human communication (Radzwill & Benton 2017).

Only recently chatbots have sparked the interest of a larger audience of major companies and their customers. Advances in fields as Artificial Intelligence have enabled chatbots to compute the vast amounts of data that are available nowadays, resulting in smoother and more flexible interactions, as the system is continuously learning (Dale, 2016). Furthermore, the changes in how we communicate today and the increased adoption of the internet and messaging platforms have facilitated the adoption of chatbots (Brandtzaeg & Folstad, 2017). Messaging apps are booming worldwide and users have become familiar and comfortable communicating via short-typed interactions. This has created an environment where chatbots can flourish, as interacting with a chatbot is not much different from what users are already familiar with in their daily interactions (Dale, 2016).

What makes a good chatbot? The need for user satisfaction measures in human-computer interaction

In order to realize this potential, chatbots have to be well adapted to the users' needs to ensure that they will form positive views about them and will continue to engage with these systems. An unsatisfactory interface could create long-term problems, for example, a decrease of trust in the quality of services/ products or the company itself (Brandtzaeg & Følstad, 2018). This is also reflected in the discontinuation of various chatbot-driven services, indicating that users' needs and expectations were not sufficiently met (Gnewuch et al., 2017). To bridge the gap between humans and machines, developers need insight into what users find important when interacting with conversational agents and how the system can satisfy these requirements.

The ISO 9241-11 (2018) describes ***user satisfaction*** as "*the extent to which the user experience that results from actual use meets the user's needs and expectations.*". Connected to this definition, ***user experience*** can be defined as the "*user's perceptions and responses that result from the use and/or anticipated use of a system, product or service*" (ISO 9241-11, 2018)

Current HCI literature offers several standardized measurement tools to capture user satisfaction and user experience. As Borsci, Federici, Bacci, Gnaldi and Bartolucci (2015) point out, short scales are favoured as they can be more easily integrated into usability testing, due to their speed and ease of administration. The ten-item System Usability Scale (short: SUS; Brooke, 1996) which is widely used, assigns a grade to the overall (perceived) usability score ranging from A+ (absolutely satisfactory) to F (absolutely unsatisfactory). Two even shorter scales are the Usability Metric for User Experience (short: UMUX), a four-item tool developed by Finstad (2010), and the UMUX-LITE which is composed of only the two positive-tone questions from the UMUX (Borsci et al, 2015; Lewis, Utesch & Maher, 2013).

Although these short scales have shown to be reliable measures of user satisfaction (Finstad, 2013) researchers frequently resort to developing their own questionnaires when evaluating chatbots. This suggests that existing user satisfaction scales are not adequate for the context of conversational agents. One possible explanation for this issue is that scales as the SUS or the UMUX were intended to measure user satisfaction with classic graphic interfaces. As Brandzaeg & Folstad (2018) argue, conversational agents, provide the possibility for a high degree of variation regarding user input, this makes the system significantly less predictable than classic interfaces with more confined paths of action. Due to the high flexibility of conversational interfaces, designers have less control over which content is going to be presented to the user, making it difficult to define interaction paths and how the chatbot should respond in these situations. The difference between these two forms of content presentation (classic vs. dynamic) suggests that natural-language interfaces might target different user needs and expectations that cannot be captured by a scale intended to evaluate more static systems.

Another explanation why current measures might be insufficient is provided by Tariverdiyeva and Borsci (2019), who concluded that while tools as the UMUX-Lite provide a good indication of the overall usability of a service or product, it does not provide diagnostic information about individual aspects of the interaction. This makes it difficult for designers to derive specific design guidelines that would benefit user satisfaction. Overall, these issues stress the need for standardized measures specific to the more dynamic context of chatbots and other conversational interfaces.

Scale for user satisfaction with information chatbots (USIC)

In 2019, Tariverdiyeva and Borsci (2019) initiated the development of a reliable measurement tool for user satisfaction to address the insufficiencies posed by the UMUX-Lite for the chatbot context. As a starting point, they conducted a qualitative systematic literature review, to identify relevant features that might influence the users' satisfaction with chatbots. From this review, 27 features relevant to usability and user satisfaction emerged, which were then presented to a panel of experts and designers as well as a group of non-expert end-users. Items or features with insufficient consensus regarding their importance were then excluded, yielding a revised list of 18 features.

Building upon the findings of Tariverdiyeva and Borsci (2019), Balaji and Borsci (2019) developed the preliminary User Satisfaction with Information Chatbots scale (short: USIC). In the first part of their study, Balaji and Borsci (2019), conducted an extended literature review to identify important features that might have been omitted earlier. The revised feature list was then used as the basis for the item generation for the questionnaire which was evaluated by several focus groups. One important limitation Tariverdiyeva and Borsci (2019) noted in their study were the significant differences between experts and end-users regarding the importance of the different features. However, as the tool is intended to measure the satisfaction of the **users** with the chatbot, Balaji and Borsci (2019) chose to only include non-experts in the focus groups.

The evaluation of the feature list and the corresponding items by the focus groups yielded a revised questionnaire comprised of 42 items, which was administered to a sample of 60 students to evaluate its reliability and underlying factor structure (Appendix A).

Based on the consistency of the data with the results from the earlier focus groups and statistical criteria Balaji and Borsci (2019) proposed a four-factor structure. The four factors were described as *communication quality*, *response quality*, *perceived privacy*, and *perceived speed*. *Communication quality* hereby refers to the ease with which users can initiate the interaction and communicate their intent, while *Response quality* places more emphasis on the output of the system. *Perceived privacy* is referring to 'the extent to which the user feels the chatbot protects one's privacy',whereas *Perceived speed* is defined as 'the (perceived) ability of the chatbot to respond timely to the user's requests' (Balaji & Borsci, 2019) Silderhuis and Borsci (2020) proposed a similar four-factor solution but reframed *Communication quality* as *Conversation start* and *Response quality* as *Communication quality.* Analyses indicated high reliability of the results suggesting a meaningful fit of the proposed structure.

Other studies suggested more factors, for example, Böcker and Borsci (2019) found five factors labelled *General usability*, *Ease of getting started*, *Perceived privacy and security*, *Response time* and

Articulateness. Neumeister and Borsci (2020) identified six factors that approached the structure proposed by Böcker and Borsci (2019). The three factors *Ease of getting started*, *Perceived privacy and security* and *Response time* were replicated, with the item distribution being almost identical to the structure proposed by Böcker and Borsci (2019). However, Neumeister and Borsci (2020) suggested the remaining factors to be divided in *Keeping track of context* and *flexibility of linguistic input* instead of *Articulateness.* Nonetheless, the authors mentioned that reliability was questionable for the factors *General satisfaction* (labelled *General usability* in Böcker and Borsci (2019)) and *keeping track of context* which suggests that these factors do not adequately capture user satisfaction with chatbots.

Goal of the current study

The current study aims to build upon the previous efforts, and to contribute to standardization of the proposed USIC scale.

During the standardization process, the reliability and validity of the scale have to be confirmed through continuous replication to assess consistency (Kyriazos & Stalikas, 2018). Another approach is the replication of the factor structure across different subject populations, to evaluate the generalizability of results (DeVellis, 2016).

To assess concurrent validity of the USIC scale, we included the UMUX-Lite as proposed by Lewis, Utesch and Maher (2013), to evaluate whether the USIC measures the same concepts as the already validated measure of user satisfaction.

Another goal of the study was to shorten the current USIC, while addressing all features without sacrificing the reliability of the scale. As the USIC is still under revision it comprises multiple redundant questions about each feature. In his paper Lewis (2014) stresses the importance of short scales to minimize user effort especially when multiple scales are integrated into a larger questionnaire. Currently, the original version of the USIC features 42 questions. Narrowing down the number of items would place less strain on the user and would enable the use of the USIC alongside other measures of user satisfaction. The four main research questions that arise are therefore as follows:

RQ1: Can the factor structure the USIC as identified in previous studies (Balaji & Borsci, 2019; Böcker & Borsci 2019; Neumeister & Borsci, 2020) be confirmed under the current population?

RQ2: Can the reliability of the USIC be confirmed?

RQ3: Can we create a shortened and reliable version of the USIC?

RQ4: Is the USIC scale correlated to the UMUX-Lite?

Furthermore, we were interested whether age has an influence on how users experience a system. According to Moore (2012) Millennials, born between 1981 and 1996, exhibit higher levels of interactive media usage (i.e. instant messaging) than the preceding cohorts Gen X (1965 – 1980) and the Baby Boomers (1946-1964). This is not surprising, as the Millennial generation is the first generation to use instant messaging, cellphones, and internet services (i.e. email) since childhood (Reisenwitz & Iyer, 2009). As Kortum and Owald (2017) point out, quantifying users' personal resources is an important factor when examining how system designs relate to user behavior and user experience. Users that frequently interact with specific systems tend to navigate new similar systems with more ease. Consequently, as Millennials are more active at integrating technology into their daily lives, they are significantly more adept at using it compared to older individuals (Moore, 2012). It is therefore possible that younger individuals will rate the interaction with the chatbot as more satisfactory compared to older individuals. The fifth research question to be answered is therefore as follows:

RQ 5: Do individuals of different ages rate their satisfaction with chatbots in a significant different way using the new scale?

Furthermore, personality styles, specifically the way users approach (new) technical systems, play an important role in the development of coping strategies (Franke, Attig & Wessel, 2019). Franke, Attig and Wessel (2019) have called this the affinity for technology interaction (short: ATI). Every new technology requires adaptation by the user who needs to have a certain set of skills and experience to cope with the challenges of the new system. Individuals that are driven to approach desirable states are more likely to actively explore new systems, broadening their problem-solving skills in the process. In contrast, individuals who display avoidance behavior often refrain from a closer interaction with new technologies to prevent experiencing problems with the system. As Franke, Attig and Wessel (2017) point out, these individual differences play an important role in explaining how users evaluate a system which leads to the final research question:

RQ6: Does affinity for technology have an influence on user satisfaction with chatbots?

Method

Ethical approval

The current study has been reviewed and approved by the ethics committee of the Faculty of Behavioral Management and Social Sciences (University of Twente). In addition, written informed consent was obtained from all participants.

Translation of the scales

The study was administered in English and German. For the English version, the original questionnaire was derived from Balaji and Borsci (2019). For the German version, the scale was translated independently by two different individuals who were fluent in both languages to ensure a high quality of the translation. Subsequently, both translations were compared to the original and inconsistencies were discussed. For a full overview of the translation scripts, the interested reader is referred to Appendix A.

Participants

Participants were selected based on the following inclusion criteria:

- Participants had to be between 18 and 70+ years of age
- All individuals had to have at least a basic understanding of either German, English and/ or Dutch language in terms of reading and writing
- All individuals had to have access to a computer with a working internet connection

Participants were recruited through a combination of convenience and snowball sampling. Potential participants were reached out to directly by the researcher as well as through advertising on social media. In both cases, participants were provided with basic information about the procedure, duration, and purpose of the study. Interested individuals were then asked to contact the researcher for more detailed information and to schedule an appointment for the experiment. In addition, participants were asked to distribute the study among their social circles, to be able to reach more potential subjects.

In total 41 subjects participated in the study (Mean $_{age}$ = 41.8 years , SD $_{age}$ = 17.4 years). All participants confirmed an at least basic understanding of the relevant language (either German or English) in reading and writing.

The English version was completed by 21.9% of the subjects. 77.8% indicated a good understanding of the English language and 22.2 % stated their comprehension level as being excellent. Since the chatbot selection for the English version included both English and Dutch chatbots, Dutch levels were assessed as well. 11.1% of the subjects indicated a basic level of Dutch, 22.2% had a good understanding and 55.5% rated their level as being excellent. To avoid confounding the results due to language barriers, only subjects that indicated a proficiency above basic level were presented with both English and Dutch chatbots.

The majority of subjects (78.1%) completed the German version, which only included German chatbots. All subjects that completed this version were native speakers.

Procedure

Due to the current COVID - 19 crisis, test sessions were conducted remotely using a video connection. At the beginning of the session, participants were asked to share their screens to enable the researcher to follow the process. During the procedure, the researcher made use of a webcam as a visual cue for her presence to facilitate communication about non-task related difficulties. Participants were free to use their webcam as well or refrain from it to minimize discomfort.

After the technical setup was completed, the researcher welcomed the subject and gave a brief overview of the study's purpose and the activities to be expected. Participants were informed that they would interact with five chatbots after which they would receive a questionnaire about their experiences with the conversational agent.

After addressing any potential questions, participants were asked to read and sign the informed consent form as displayed in Qualtrics. Participants who did not agree with the aspects and conditions mentioned in the informed consent form were thanked for their time and excluded from the study. In cases where consent was given, participants were asked to complete a short demographic questionnaire, including questions on age, level of proficiency in English/Dutch (only for the English version), education level, previous experience with chatbots and their affinity for technology interaction.

Subsequently, the researcher directed the subject to the next page with the chatbot tasks and the questionnaire. In total each participant was interacting with five chatbots that were semi-randomly assigned through the Qualtrics randomization tool. For each chatbot participants were provided with a short usage scenario, representative of the usage of the website as well as the appertaining link. One example of a usage scenario was concerned with the American railroad company Amtrak:

"*You have planned a trip to the USA. You are planning to travel by train from Boston to Washington D.C. You want to stop in New York to meet an old friend for a few hours and see the city. You want to use Amtrak's chatbot to find out how much it will cost to temporarily store your luggage at the station."*

To enhance the internal validity of the study, assignments of the chatbots were evenly distributed and the item sequence of the USIC scale was randomized. If participants needed more than one minute to locate the chatbot on the website, the researcher pointed them to the chatbot to prevent a premature abortion of the task. In scenarios in which participants were not able to complete the task despite the direction of the researcher, participants were asked to move on with filling in the USIC/UMUX-Lite questionnaires as far as possible. Any cases of assistance or premature terminations were noted by the researcher to guide the interpretation and analysis of the results.

After completing the five scenarios and the questionnaires, participants were given room for questions and were provided with the researchers' contact data for further information about the outcomes of the study. Subsequently, participants were thanked for their participation and the researcher ended the session.

Materials

The testing sessions were conducted using the video meeting platform "Whereby". One important aspect of choosing this software was that users can join meetings via a weblink without the need to create an account or download software. Therefore, the sessions were approachable at all levels of technical capabilities. For each session, audio and screen recordings were made using the *Flashback Express player*. In the few cases where no microphone was available, participants were phoned and put on a loudspeaker during the video meeting so the recording software could capture the auditory input. Furthermore, Qualtrics was used to present subjects with the written materials such as the informed consent form, the chatbot tasks, the USIC scales as well as the (translated) UMUX-Lite questionnaire.

To assess subjects' technology interaction styles, the *Affinity for technology interaction scale (ATI)* by Franke, Attig and Wessel (2019) was used. The 9-item ATI scale captures the interaction with entire technological devices (e.g. mobile phones) as well as software (e.g. apps), using a 6-point Likert scale ranging from 'completely disagree' to 'completely agree'(Appendix B).

We also included the standardized UMUX-Lite by Lewis, Utesch and Maher (2013) for comparison with the USIC, to be able to assess the USIC concurrent validity. The UMUX-Lite is a two-

item questionnaire that assesses general user satisfaction in systems. Due to its brief format, the session length was only minimally affected by this addition, which avoided placing further strain on the subjects.

For the English version, a total of 14 chatbots (7 English, 7 Dutch) were included, which were partially derived from Balaji and Borsci (2019) and Silderhuis (2020). However, one English chatbot (from the meal-kit service *Hello Fresh*) had to be excluded after a few sessions due to the discontinuation of the service.

For the German version, 7 new chatbots were selected from different areas such as Travel (*Lufthansa*) or community services (*WienBot*). The complete lists of chatbots from both versions, including the associated links, can be found in Appendix C. Furthermore, to keep the usage scenario as realistic as possible, subjects were merely equipped with a general link to the website in question, contrary to a specific URL linking directly to the chatbot (with exception of the *WienBot*). Subjects, therefore, had to locate the chatbot themselves, which was needed to capture the aspect of accessibility.

Data Analysis

After screening the dataset for missing values and the inversion of negatively worded items, the data was imported into R Studio for analysis. To examine suitability of the data for a factor analysis, the Keyser-Meyer-Olkin (KMO) measure of sampling adequacy was used, aiming for a value above the general recommended threshold of 0.6. Additionally, the Bartletts test for sphericity was performed.

To establish the number of factors to be retained, a parallel analysis was conducted using the parallel analysis (fa.parallel) function from the R package ‘psych’ (Revelle, 2017). The function uses simulated data and compares it to the actual data. The number of factors to retain is hereby indicated by where the tracings for the actual (blue line) and simulated data (red line) cross. Factors that are above this crossing point show eigenvalues above what would be attributed to chance and should be preserved. Parallel analysis is seen as an accurate factor retention predictor, however, in cases of smaller sample sizes, additional criteria are advised to be employed for factor extraction (Turner, 1998). Therefore, the scree plot inflexion point and the Kaisers criterion (Eigenvalues >1) were used to complement the results of the parallel analysis.

Based on the factor range that was suggested by the three aforementioned criteria, different factor solutions with four, five and six factors were examined using a varimax rotation. The best-fitting factor solution was determined based on the most meaningful item distribution, as well as Cronbach’s alpha for the individual factors.

For the scale reduction, all items below a cut-off value of 0.6 were excluded, yielding a

preliminary scale with 32 items. Subsequently, the items with the strongest loadings for each feature as proposed by Balaji and Borsci (2019) were selected, resulting in the 14-item version of the USIC. This procedure was repeated for the two age groups to be able to assess differences in factor distribution. Reliability analyses for the overall scale and the latent factors were conducted using the alpha function from the R package 'psych' (Revelle, 2017).

To gather evidence for the concurrent validity of the USIC, a correlational analysis was conducted for the USIC and the UMUX-Lite using Spearman's rank-order correlation. Effects of age and affinity for technology interaction on user satisfaction were assessed with a linear regression analysis using the 'rStats' package (Revelle, 2017).

Results

Data screening

The data set comprised one data line per chatbot and participant combination. As each of the 41 participants was exposed to five chatbots, this yielded a dataset of 205 observations. The data did not show extreme or missing values; therefore the complete dataset could be used for analysis.

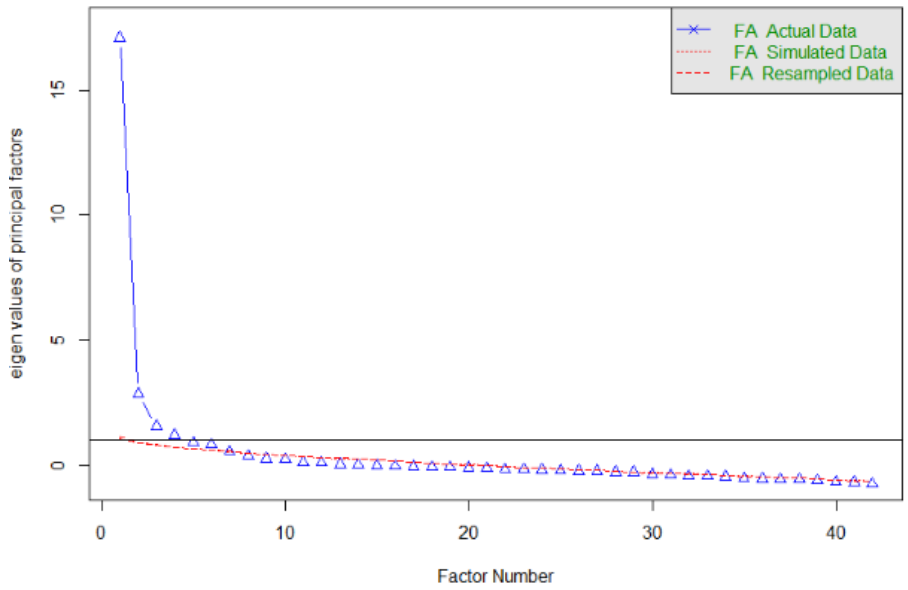
Factor structure of the USIC scale

Preceding the analysis, the factorability of the USIC was examined using several criteria. All items displayed a correlation of at 0.3 or higher with at least one other item. Furthermore, the Keyser-Meyer-Olkin measure of sampling adequacy was well over the threshold of 0.6 with an overall value of .93 and individual item values above .67 (Hair et al, 2010). The Bartletts test of sphericity was significant ($p < .001$). Based on the fulfilment of the abovementioned criteria, an exploratory factor analysis was deemed suitable for all 42 items of the scale.

A parallel analysis was conducted, as this method is seen as an accurate factor retention predictor (Hayton, Allen & Scarpello, 2004). The results suggested a solution between 4 to 6 factors based on the aforementioned criteria. As mentioned earlier, parallel analysis makes use of simulated data (red line) and compares it to the actual data (blue line). The number of factors is indicated by the crossing point of the two lines. Factors above the crossing point show eigenvalues above what would be attributed to chance and should therefore be retained. As illustrated by the screeplot below (Figure 1) six factors were above the crossing point, therefore a six-factor structure was examined initially.

Figure 1

Parallel analysis screeplot with number of factors to be retained



Analysis showed a meaningful item distribution with relatively weak cross loadings. However, Cronbach's alpha for Factor 6 was 0.59 which indicated poor reliability (DeVellis, 1991; p.85). It was therefore chosen to discard Factor 6 and to evaluate a 5 factor solution. As with the six factor structure, the latent factors could be interpreted coherently, but Cronbach's alpha was again unacceptable ($\alpha = 0.59$) for one of the factors (Factor 5).

Subsequently, four factors were extracted with alpha values of $\alpha = 0.97$ (F1), $\alpha = 0.91$ (F2), $\alpha = 0.78$ (F3) and $\alpha = 0.67$ (F4) for the individual factors. As illustrated by Table 1, the items were meaningfully distributed across the four factors, in line with previous research (Balaji and Borsci, 2019; Silderhuis and Borsci, 2020). Therefore, we opted for this four-factor solution over the others. The four factors accounted for a total variance of 56.5 % and 33.4%, 11.4%, 7.4%, 4.3% of the individual variances. A varimax rotation suggested a simple factor structure with items loading strong onto only one factor and relatively weak cross-loadings.

Table 1.
The factor structure of the 42-item USIC

Item	Description	F1 Communication quality	F2 Conversation start	F3 Perceived speed	F4 Perceived privacy
Q1	It was clear how to start a conversation with the chatbot.	0.165	**0.687**	0.165	
Q2	It was easy for me to understand how to start the interaction with the chatbot.	0.261	**0.719**		
Q3	I find it easy to start a conversation with the chatbot.	0.300	**0.699**	0.174	
Q4	The chatbot was easy to access	0.146	**0.781**	0.134	
Q5	The chatbot function was easily detectable	0.173	**0.816**		
Q6	It was easy to find the chatbot.	0.143	**0.800**		0.110
Q7	Communicating with the chatbot was clear.	**0.730**	0.387		
Q8	I was immediately made aware of what information the chatbot can give me.	**0.486**	0.379		
Q9	It is clear to me early on about what the chatbot can do	**0.518**	0.395		
Q10	I had to rephrase my input multiple times for the chatbot to be able to help me.	**0.691**			
Q11	I had to pay special attention regarding my phrasing when communicating with the chatbot.	**0.582**		-0.142	0.106
Q12	It was easy to tell the chatbot what I would like it to do.	**0.713**	0.318		
Q13	The interaction with the chatbot felt like an ongoing conversation	**0.451**	0.335		
Q14	The chatbot was able to keep track of context.	**0.773**	0.137	0.134	
Q15	The chatbot maintained relevant conversation.	**0.677**	0.135	0.137	
Q16	The chatbot guided me to the relevant service.	**0.618**	0.225	0.272	0.256
Q17	The chatbot is using hyperlinks to guide me to my goal.	0.125			**0.281**
Q18	The chatbot was able to make references to the website or service when appropriate.	**0.505**	0.176	0.160	0.251
Q19	The interaction with the chatbot felt secure in terms of privacy.	0.206	0.163	0.165	**0.695**
Q20	I believe the chatbot informs me of any possible privacy issues.	0.129			**0.550**
Q21	I believe that this chatbot maintains my privacy.	0.120		0.134	**0.687**
Q22	I felt that my intentions were understood by the chatbot.	**0.897**	0.102		
Q23	The chatbot was able to guide me to my goal.	**0.715**	0.185	0.218	0.219

Q24	I find that the chatbot understands what I want and helps me achieve my goal.	**0.843**	0.181	0.186	
Q25	The chatbot gave relevant information during the whole conversation.	**0.804**	0.192	0.127	
Q26	The chatbot is good at providing me with a helpful response at any point of the process.	**0.836**	0.237	0.142	0.102
Q27	The chatbot provided relevant information as and when I needed it.	**0.804**	0.130	0.136	0.145
Q28	The amount of received information was neither too much nor too less.	**0.738**		0.151	
Q29	The chatbot gives me the appropriate amount of information.	**0.758**		0.327	
Q30	The chatbot only gives me the information I need.	**0.796**	0.100		
Q31	The chatbot could handle situations in which the line of conversation was not clear.	**0.430**		0.291	-0.182
Q32	The chatbot explained gracefully when it could not help me.	0.196	-0.320	**0.317**	-0.105
Q33	When the chatbot encountered a problem, it responded appropriately.	0.229		**0.419**	-0.116
Q34	I found the chatbot's responses clear.	**0.779**	0.134	0.324	
Q35	The chatbot only states understandable answers.	**0.721**	0.187	0.295	
Q36	The chatbot's responses were easy to understand.	**0.655**	0.203	0.335	
Q37	I feel like the chatbot's responses were accurate.	**0.754**	0.114	0.303	0.187
Q38	I believe that the chatbot only states reliable information.	**0.592**		0.167	0.118
Q39	It appeared that the chatbot provided accurate and reliable information.	**0.762**	0.121	0.297	0.150
Q40	The time of the response was reasonable.	0.346	0.243	**0.703**	0.106
Q41	My waiting time for a response from the chatbot was short.	0.317	0.171	**0.758**	0.101
Q42	The chatbot is quick to respond.	0.343	0.198	**0.745**	0.111

Note. Item's highest factor loading in boldface.

As previously mentioned, the item distribution strongly resembled the structure that was proposed by Balaji and Borsci (2019) and Silderhuis and Borsci (2020). However, in the current study, the factors were extracted in a different order (see Table 2).

Table 2.
The factor structure of the 42-item USIC identified in the present study, compared to the factor structures by Balaji and Borsci (2019) and Silderhuis and Borsci (2020).

	Balaji and Borsci (2019)		Silderhuis and Borsci (2020)		Current study	
	Factor name	Items	Factor name	Items	Factor name	Items
F1	Communication quality	Q1, Q2, Q3 Q4, Q5, Q6, Q10, Q11	Conversation start	Q1, Q2, Q3, Q4, Q5, Q6	Communication quality	Q7,Q8,Q9, Q10,Q11, Q12, Q13,Q14,Q15, Q16,Q18 Q22,Q23,Q24, Q25,Q26,Q27, Q28,Q29,Q30, Q31,Q34,Q35, **Q36**, Q37,**Q38**,Q39
F2	Response quality	Q7, Q8, Q9, Q12, Q14, Q15, Q16, Q17, Q18, Q22, Q23, Q24, Q25, Q26, Q27, Q28, Q29, Q30, Q31, Q32, Q33, Q34, Q35, Q36, Q37, Q38, Q39	Communication quality	Q7, Q8, Q9, Q10, Q11, Q12, Q13, Q14, Q15, Q16, Q18, Q22, Q23, Q24, Q25, Q26, Q27, Q28, Q29, Q30, Q31, Q33, Q34, Q35, Q37, Q39	Conversation start	Q1, Q2, Q3, Q4, Q5, Q6
F3	Perceived privacy	Q13, Q19, Q20, Q21	Perceived privacy	Q19, Q20, Q21, Q32, Q38	Perceived speed	**Q32**,**Q33**, Q40, Q41, Q42
F4	Perceived speed	Q40, Q41, Q42	Perceived speed	Q36, Q40, Q41, Q42	Perceived privacy	Q19,Q20 Q21

The internal consistency of the USIC scale was assessed using Cronbach's alpha. The alpha values were high with $\alpha = 0.96$ for the entire scale and $\alpha = 0.97$ (F1), $\alpha = 0.91$ (F2), $\alpha = 0.78$ (F3) and $\alpha = 0.67$ (F4) for the individual factors. This indicated a high internal consistency, which allowed for reduction and refinement of the scale.

Scale reduction

As no substantial increases in the overall alpha for the 42-item USIC could have been achieved by eliminating items, a reduction of the scale based on alpha values was not feasible. Instead, the reduction was approached by excluding items based on the factor loadings. According to Floyd and Widaman (1995), to yield stable solutions for a sample with 150 observations, a more conservative cut-off value of .6 should be used. With the current sample containing 205 observations, it was therefore chosen to adhere

to this cut-off value for the exclusion of weaker items.

Based on this criterion, a total of 10 items (Q8, Q9,Q11,Q13,Q17,Q18,Q20,Q32,Q33,Q38) were excluded. Even though all items associated with *Graceful breakdown* showed factor loadings below .6 we retained Item 31 as a representation for this feature. This decision was based on the results of Balaji and Borsci (2019) who identified this feature as an important aspect of user satisfaction with chatbots.

The internal consistency of the reduced 32-item USIC remained at the same high level as the 42-item version with a value of $\alpha = 0.96$ for the entire scale and values of $\alpha = 0.97$ (F1), 0.91 (F2), 0.95 (F3) and 0.82 (F4) for the individual factors.

Even though the analysis was indicating that the 32-item USIC version is reliable, this version could still be considered quite extensive. Longer scales have the disadvantage of subjecting participants to cognitive strain, especially when the scale is used among other tools. Therefore, a further reduction of the scale was important in order to increase the applicability of the USIC for future research.

The original USIC as proposed by Balaji and Borsci (2019, Appendix A) included multiple items per chatbot feature, therefore it was chosen to only retain the items with the highest factor loading for each of the 14 features, thus those items that show the strongest relationship with the respective factors. This resulted in the 14 item version that is summarized in Table 3.

Analysis suggested a two-factor structure for the 14-item USIC, based on the Kaisers criterion, the visual inflection point of the scree plot and the parallel analysis. The two factors explained 55.4% of the total variance and 42.4% (F1) and 13.0% (F2) of the individual variances.

Cronbach's alpha decreased slightly but nonetheless remained at a high level of $\alpha = 0.92$ for the entire scale. The values for the individual factors were $\alpha = 0.93$ (F1) and $\alpha = 0.61$ (F2). Analysis indicated that, the Cronbach's alpha for Factor 2 could be improved by deleting Item 19. However, the item is representing *perceived privacy*, which was identified as an important factor for user satisfaction with chatbots. Therefore, it was chosen to retain this item.

Table 3.
Item distribution of the 14-item USIC

Factor	Feature	Item
F1 Communication quality	Expectation setting	Q7
	Communication effort	Q12
	Ability to maintain themed discussion	Q14
	Reference to service	Q16
	Recognition and facilitation of user's goal and intent	Q22
	Relevance	Q26
	Maxim of quantity	Q30
	Graceful breakdown	Q31
	Understandability	Q34
	Perceived credibility	Q39
	Perceived speed	Q42
F2 Conversation start	Ease of starting a conversation	Q2
	Accessibility	Q5
	Perceived privacy	Q19

Correlation USIC and UMUX-Lite

To evaluate the USIC scale's concurrent validity, the correlation between the 14-item USIC and the UMUX-Lite was examined using Spearman's rank-order correlation. Before the analysis, row means were computed for all items of the scales.

The proposed 14-item USIC displayed a strong correlation with the UMUX-Lite, suggesting a high concurrent validity (Table 4). Factor 1 (*Communication quality*) displayed the strongest relationship of the individual factors, while Factor 2 *(Conversation start)* was only moderately correlated to the UMUX-Lite. All correlations were significant.

Table 4.
Correlations between UMUX-Lite and the 14-item USIC

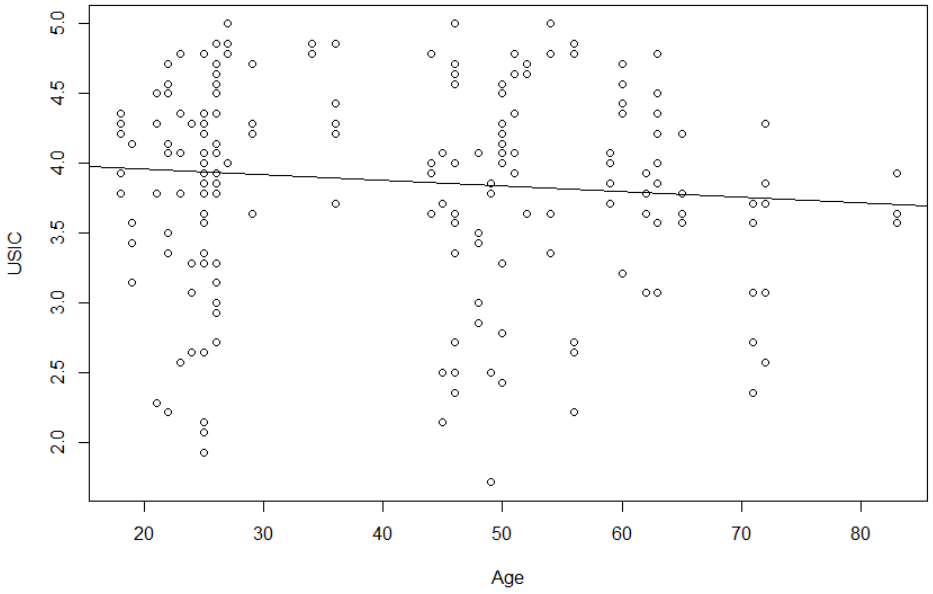
	UMUX-Lite
14-item USIC	.841
(F1) Communication quality	.819
(F2) Conversation start	.610

Effects of age on user satisfaction with chatbots

To investigate whether subjects of different ages differ in their ratings of user satisfaction with the chatbots, a simple linear regression was employed. Analysis indicated a slight negative trend, with ratings of overall user satisfaction decreasing for older ages (Figure 2). However, this effect was non-significant ($p = .168$)

Figure 2

Effects of age on ratings on the 14-item USIC

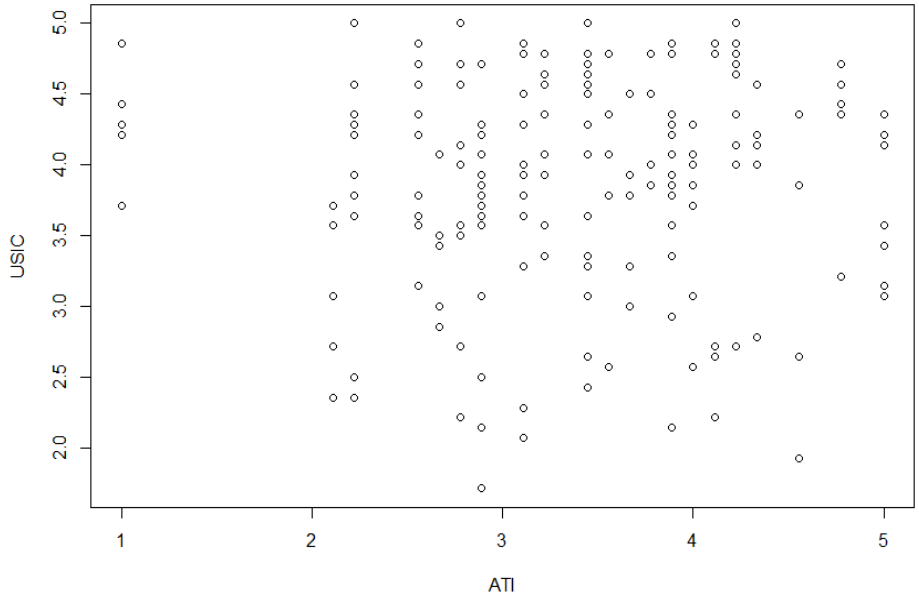


Effects of Affinity for Technology on satisfaction with chatbots

Another objective was to examine the possible effects of affinity for technology interaction on user satisfaction with chatbots. Results of the linear regression indicated no significant effect of affinity for technology ($p = .848$; Figure 3).

Figure 3

Effects of affinity for technology interaction on ratings on the 14-item USIC



Discussion

The current study aimed to contribute to the psychometric evaluation of the USIC questionnaire's reliability and validity across different age groups. The data suggested a meaningful fit of the four-factor structure in line with previous work (Balaji & Borsci, 2019; Silderhuis & Borsci, 2020). Furthermore, we gathered evidence for the concurrent validity of the USIC, which was indicated by the strong correlation with the validated UMUX-Lite for the scale and the factor *communication quality* (F1).

The first research question was whether the factor structure that was suggested in previous studies (Balaji & Borsci, 2019; Böcker & Borsci, 2019; Silderhuis & Borsci, 2020 and Neumeister & Borsci, 2020) can be replicated.

The first inspection of the data based on the Kaisers criterion and the screeplot inflection point suggested two to six factors which is in line with the number of factors that were suggested by previous works on this scale. A four-factor solution showed the best fit for our data, in line with the findings of Balaji and Borsci (2019) and Silderhuis and Borsci (2020). Furthermore, the item distribution under the current population closely resembled the structure of previous studies, indicating generalizability.

However, there are some differences in the item distribution that should be discussed. In fact, while the overall structure of the USIC as proposed by Balaji and Borsci (2019) and Silderhuis and Borsci (2020) could be confirmed, five items loaded onto different factors in the current study, as follows:

- **Q17** refers to the chatbot providing hyperlinks during the interaction to guide users to their goal. In the study of Balaji and Borsci (2019), this item was included in the factor *Response quality* while it did not load on any of the factors in Silderhuis and Borsci (2020). The results of the current study indicated that item 17 is part of the factor *perceived privacy*. We argue that if a chatbot uses hyperlinks to guide the user to a different website, users perception of their privacy may change, as privacy policies vary across different sites. This might be a reason why this item loaded onto the *perceived privacy* factor.
- **Q32** and **Q33** are associated with how gracefully handles problems that arise during the interaction. Balaji and Borsci (2019) proposed that this item is related to *response quality*, while Silderhuis and Borsci (2020) associated this feature with *perceived privacy*. However, our results suggested that this feature is related to *perceived speed.* A possible explanation for this finding is, that graceful breakdown also encompasses that the chatbot provides immediate feedback when issues are encountered, avoiding pauses that might confuse the (unexperienced) user.
- **Q36** captures how easy the answers of the chatbot are to understand. Our analysis suggested that this item is related to *Communication quality.* This supports the findings of Balaji and Borsci

(2019) who also found an association with response quality. But Silderhuis and Borsci (2020) proposed that this item is related to *perceived speed.* However, we argue that our proposed categorization provides a more meaningful fit, as the understandability of the chatbots' answers is unlikely to be associated with the response rate of the chatbot.

- **Q38** evaluates how users rate the reliability of the information that the chatbot provides. Our results suggest that this item belongs to the factor *Communication quality* which, again, is in line with Balaji and Borsci (2019). Yet, Silderhuis and Borsci (2020) linked this item to the factor *perceived privacy*. However, providing information that is accurate and reliable can also be seen as an aspect of the quality of the interaction. Therefore, our categorization is a viable alternative explanation.

The second research question of the present work was whether the reliability of the USIC that was indicated by previous studies (Balaji & Borsci, 2019; Böcker & Borsci, 2019; Neumeister & Borsci, 2020, Silderhuis & Borsci, 2020) could be confirmed under the current population. The analysis showed high alpha values for the preliminary 42- Item version as proposed by Balaji and Borsci (2019) and Silderhuis and Borsci (2020) as well as for our suggested refined 32-item scale. Furthermore, alpha values were high for the individual factors for both versions. This indicates a high internal consistency of the scale which provides evidence that the USIC is a reliable tool to assess user satisfaction with chatbots.

Moreover, our third research question was whether it was possible to propose a shorter but still reliable version of the USIC. To shorten the scale, items below the established cut-off value of 0.6 were excluded, which yielded a preliminary version with 32 items. From this scale, the items with the strongest factor loading per feature were retained, to capture all relevant aspects of user satisfaction with chatbots. This resulted in the 14- item USIC with a high level Cronbach's alpha (0.92) for the entire scale, divided in two factors: *Communication quality* (F1) composed by 11 items (Cronbach's alpha $\alpha = 0.93$) and *Conversation start* (F2) composed by 3 items (Cronbach's alpha .61)

Furthermore, the results showed a strong correlation between the UMUX-Lite and the refined 14-Item USIC (in line with the fourth research question). The relationship was the strongest for the factor *Communication quality* (F1), while *Conversation start* (F2) was only moderately correlated with the UMUX-Lite. This suggests that the factor *Communication quality* captures the same aspects of user satisfaction that are measured by the UMUX-Lite. These findings are directly in line with Tariverdiyeva and Borsci (2019) who argued that user satisfaction with chatbots is multifaceted. The authors found, that the UMUX- Lite only captured perceived ease of use. This was also affirmed by Waldera and Borsci (2019) and Silderhuis and Borsci (2020). Waldera and Borsci (2019) identified a strong correlation of UMUX-Lite with the features, *Reference to service, Recognition of user's intent and goal, Perceived*

credibility, and *Ability to maintain themed discussion*. In the current study as well as in Silderhuis and Borsci (2020), all of the mentioned features loaded onto the *communication quality* factor, which strongly correlated with the UMUX-Lite.

The moderate to low correlation of the UMUX-Lite with the remaining factor *Conversation start* (F2), provide further evidence for the added value of the USIC. While the UMUX-Lite is a broad assessment of user satisfaction (Lewis, 2013) the USIC provides information on additional aspects of the interaction (Balaji & Borsci, 2019). This contributes to the diagnostic character of the USIC that other user satisfaction tools, i.e. the SUS (Brooke, 1996) or the UMUX- Lite (Lewis, Utesch & Maher, 2013) are lacking.

In line with our fifth research question we also investigated whether age has an influence on the user satisfaction ratings with the 14-items scale. Research shows that individuals from the Millennial generation (25 – 40 years old) and Baby Boomers (56 – 75 years old) have vastly different levels of interactive media usage. We therefore expected that these differences would be reflected in the user satisfaction ratings.

Even though the analysis indicated a slight negative trend, thus slightly lower ratings of user satisfaction for older subjects, this effect was non-significant. A possible explanation is that the sample was not diverse enough. The sample was relatively young with majority of the participants being Millennials or individuals from Gen X. The two mentioned generational cohorts are often described as homogenous in regard to their use of interactive media such as chatbots. This was also reflected in our data, as the ratings of the subjects between 18 and 55 were highly similar. Older individuals in contrast were underrepresented in this study, with only ten participants above the age of 56 years. It is therefore likely, that the results are not a realistic reflection of the differences between the age groups, due to this underrepresentation.

Finally we investigated whether affinity for technology interaction has an effect on user satisfaction ratings. The results do not indicate a significant effect of affinity for technology interaction on user satisfaction with chatbots. Our rationale for this research question was based on the work of Franke, Attig and Wessel (2019), who point out that users differ in their interaction styles and therefore in their evaluation of (new) systems. Individuals with a high affinity for technology interaction actively seek to explore new systems, thereby broadening their skillset in coping with a variety of systems. We, therefore, expected that subjects with a high affinity for technology interaction would show higher USIC ratings, compared to subjects with more limited coping skills.

One possible explanation for the lack of effect of affinity for technology interaction is that the ATI scale might not be appropriate for the context of chatbots. Franke, Attig and Wessel (2018) included

a variety of technologies in their definition of technical systems which includes not only software applications but also entire digital devices such as computers or navigation systems. We argue that users might use different strategies when interacting with these devices that are usually more limited in their paths of action, compared to interacting with a chatbot that is highly dynamic.

Another potential reason why affinity for technology interaction did not predict satisfaction with chatbots is that subjects might have quickly formed a cognitive schema on how the chatbot works. This assumption is supported by statements of participants during the sessions, who indicated that solving the tasks became easier after the first chatbots. As the session progressed, participants had clearly developed a strategy and knew where to look for the chatbot and how to formulate their request. It is therefore likely, that this compensated for the limited coping skills of subjects with low affinity for technology interaction, leading them to evaluate the interaction with the chatbots more positively.

Limitations and recommendations for future research

Our research outcomes were generally in line with previous research; however the results should be treated with caution due to several limitations of the current study. One important issue that might have influenced the representativeness of the results is the lack of diversity in our sample in regard to age. Younger age groups were vastly overrepresented in our study, as the majority of the subjects were younger than 56 years. Due to the current COVID-19 crisis, we were forced to conduct the sessions remotely, which made it difficult to reach older participants. Numerous older individuals we reached out to did not have access to the required hardware or expressed that they did not feel confident to setup the connection by themselves. This reduced the number of potential subjects in this age group, which contributed the imbalance of the sample. To avoid exclusion of subjects because of these circumstances we advise to repeat the study in a laboratory, where the necessary equipment can be provided and subjects can be better supervised.

Another point of consideration is the use of the ATI scale. As previously mentioned, we argued that the ATI scale, as it was used here, might have been too broad and therefore not appropriate for the context of chatbots. The ATI scale was developed to assess *general* interaction styles with a wide range of different technologies. However, Franke, Attig and Wessel (2017) point out that the instruction text that introduces the scale can be adjusted to fit more specific technologies. Therefore, we recommend to specifically address chatbots in the instructions in future studies and to re-evaluate effects of affinity for technology interaction on user satisfaction with chatbots.

Future research should also consider examining the influence of prior experience on user satisfaction with chatbots. Borsci et al. (2015) found that prior experience with a system or product was

associated with user satisfaction. Subjects that were already familiar with the tested system were likely to rate it as more satisfactory, compared to subjects that had never interacted with the system before. However, it should be noted, that Borsci et al. (2015) assessed user satisfaction with an online platform, which is not comparable to the interaction with a highly dynamic chatbot. Therefore, future studies should include subjects with different levels of experience with conversational agents to evaluate whether this effect is observable in the specific context of information chatbots.

Conclusion

The current study contributed to the standardization of the newly developed USIC questionnaire, by replicating the four-factor structure that was proposed by previous research. The comparable item distribution provided a strong indication for the reliability and validity of the scale suggesting that the USIC is a promising tool for the assessment of user satisfaction. Additionally, the USICs value as a diagnostical measure was supported by the strong correlation with the UMUX- Lite for the factor *Communication quality* and the comparably low correlation with the remaining factor *Conversation start*, which indicates that the USIC captures additional aspects of user satisfaction.

The compact 14-item version allows researchers to administer the scale alongside other user satisfaction measures. This contributes to a deeper understanding of the relevant aspects of user satisfaction with chatbots and the development of clear design guidelines, which is necessary to realize the full potential of conversational agents.

References

Abbas, A. (2019, February 12). *Chatbot 2019: Trends and Stats with Insider Reports*. Retrieved from https://chatbotslife.com/chatbot-2019-trends-and-stats-with-insider-reports-fb71697deee4

Balaji, D., & Borsci, S. (2019*). Assessing user satisfaction with information chatbots: A preliminary investigation* [Master's thesis, University of Twente]. Retrieved from http://purl.utwente.nl/essays/79785

Borsci, S., Federici, S., Bacci, S., Gnaldi, M., & Bartolucci, F. (2015). Assessing user satisfaction in the era of user experience: Comparison of the SUS, UMUX, and UMUX- LITE as a function of product experience. *International Journal of Human-Computer Interaction, 31*(8), 484–495. https://doi.org/10.1080/10447318.2015.1064648

Böcker, N., & Borsci, S. (2019). *Usability of information-retrieval chatbots and the effects of avatars on trust* [Bachelor Thesis, University of Twente]. Retrieved from http://essay.utwente.nl/78097/

Brandtzaeg, P. B., & Følstad, A. (2017). Why people use chatbots. *Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 10673 LNCS*, 377–392. http://doi.org/10.1007/978-3-319-70284-1_30

Brandtzaeg, P. B., & Følstad, A. (2018). Chatbots: Changing user needs and motivations. *Interactions, 25*(5), 38–43. https://doi.org/10.1145/3236669

Brooke, J. (1996). SUS: A "quick and dirty" usability scale. *Usability Evaluation in Industry, 189.* https://doi.org/10.1002/hbm.20701

Dale, R. (2016). The return of chatbots. *Natural Language Engineering. 22*(5): 811–817. doi:10.1017/S1351324916000243

Derksen, J. R. (2016, October 13). *Consumers Welcome Personalized Offerings but Businesses Are Struggling to Deliver, Finds Accenture Interactive Personalization Research.* Retrieved from https://newsroom.accenture.com/news/consumers-welcome-personalized-offerings-but-businesses-are-struggling-to-deliver-finds-accenture-interactive-personalization-research.htm

DeVellis, R. F. (1991). *Scale development: Theory and applications.* SAGE Publications.

DeVellis, R. F. (2016). *Scale development*: *Theory and applications* (4th ed.). SAGE Publications.

Finstad, K. (2010). The Usability Metric for User Experience. *Interacting with Computers, 22*(5), 323-327. https://doi.org/https://doi.org/10.1016/j.intcom.2010.04.004

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7*(3), 286–299. https://doi.org/10.1037/1040-3590.7.3.286

Fogg, B. J., & Tseng, H. (1999). The elements of computer credibility. *Conference on Human Factors in Computing Systems - Proceedings*, 80–87. http://doi.org/10.1145/302979.303001

Franke, T., Attig, C., & Wessel, D. (2017). *Assessing Affinity for Technology Interaction – The Affinity for Technology Interaction (ATI) Scale*. Unpublished manuscript.

Franke, T., Attig, C., & Wessel, D. (2019). A Personal Resource for Technology Interaction: Development and Validation of the Affinity for Technology Interaction (ATI) Scale. *International Journal of Human-Computer Interaction, 35*(6), 456–467. http://doi.org/10.1080/10447318.2018.1456150

Press, G. (2019, September 25). *AI Stats News: Chatbots Lead To 80% Sales Decline, Satisfied Customers And Fewer Employees*. Retrieved from https://www.forbes.com/sites/gilpress/2019/09/25/ai-stats-news-chatbots-lead-to-80-sales-decline-satisfied-customers-and-fewer-employees/?sh=79f8304448e0

Gnewuch, U., Morana, S., & Maedche, A. (2017). Towards designing cooperative and social conversational agents for customer service. *Proceedings of the International Conference on Information Systems (ICIS)*.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2010). *Multivariate data analysis: A global perspective* (7th ed.). Pearson Education.

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational research methods, 7*(2), 191- 205.

International Organization for Standardization. (2018). *Ergonomic of human system interaction - Part 11: Usability: Definitions and concepts* (ISO Standard No. 9241-11). Retrieved from https://www.iso.org/standard/63500.html

Jovic, D. (2020, November 20). *The Future is Now - 37 Fascinating Chatbot Statistics*. Retrieved from https://www.smallbizgenius.net/by-the-numbers/chatbot-statistics/#gref

Kojouharov, S. (2018, January 24). *How Businesses are Winning with Chatbots & Ai*. Retrieved from https://chatbotslife.com/how-businesses-are-winning-with- chatbots-ai-5df2f6304f81

Kortum, P., & Owald, F. L. (2017). The impact of personality on the subjective assessment of usability. *International Journal of Human- Computer Interaction, 34*,177–186. doi:10.1080/10447318.2017.1336317

Kyriazos, T. A., & Stalikas, A. (2018). Applied psychometrics: The steps of scale development and standardization process. *Psychology, 9*(11), 2531–2560. https://doi.org/10.4236/psych.2018.911145

Lewis, J. R., Utesch, B. S., & Maher, D. E. (2013). UMUX-LITE - When there's no time for the SUS. *Proceedings of CHI 2013*, 2099–2102. https://doi.org/10.1145/2470654.2481287

Lewis, J. R. (2014). Usability: Lessons learned ... and yet to be learned. *International Journal of Human-Computer Interaction, 30*, 663–684. doi:10.1080/10447318.2014.930311

LiveChat. (2021). *Interactive 2021 Customer Service Report*. Retrieved from https://www.livechat.com/customer-service-report/

Moore, M. (2012). Interactive media usage among millennial consumers. *Journal of Consumer Marketing, 29*(6), 436–444. https://doi.org/10.1108/07363761211259241

Neumeister, S., & Borsci, S. (2020*). Testing of a usability assessment tool for chatbots: Investigating the effect of believing that a chatbot might be a human* [Bachelor thesis, University of Twente]. Retrieved from http://essay.utwente.nl/81714/

Radziwill, N., & Benton, M. (2017). Evaluating quality of chatbots and intelligent conversational agents. *ArXiv Preprint ArXiv:1704.04579.*

Reisenwitz, T. & Iyer, R. (2009). Differences in Generation X and Generation Y: Implications for the Organization and Marketers. *Marketing Management Journal, 19*(2), 91–103.

Revelle, W. R. (2017). *psych: Procedures for personality and psychological research*.

Silderhuis, I., & Borsci, S. (2020). *Validity and Reliability of the User Satisfaction with Information Chatbots Scale (USIC)* [Master's thesis, University of Twente]. Retrieved from http://essay.utwente.nl/83495/

Suthar, S. (2018, February 5). Review of *11 Chatbot Trends that Help Grow Your Business*. Retrieved from https://acquire.io/blog/chatbots-trends/

Sweezey, M. (2019, August 4). *Key Chatbot Statistics to Know in 2019*. Retrieved from https://www.salesforce.com/blog/chatbot-statistics/

Tariverdiyeva, G., & Borsci, S. (2019). *Chatbots' perceived usability in information retrieval tasks: An exploratory analysis* [Master's thesis, University of Twente]. Retrieved from http://essay.utwente.nl/77182/

Turner, N. E. (1998). The effect of common variance and structure pattern on random data eigenvalues: Implications for the accuracy of parallel analysis. *Educational and Psychological Measurement, 58*(4), 541-568.

Waldera, L., & Borsci, S. (2019). *Development of a preliminary measurement tool of user satisfaction for information-retrieval chatbots* [Bacherlor's thesis, University of Twente]. Retrieved from http://purl.utwente.nl/essays/79258

Zaboj, D. (2020, May 6). *Key Chatbot Statistics You Should Follow in 2021*. Retrieved from https://www.chatbot.com/blog/chatbot-statistics/

Appendices

Appendix A

Table A1.
The 42-item USIC with corresponding features

	Chatbot feature	Description		Questionnaire item
1	Ease of starting a conversation	How easy it is to start interacting with the chatbot	Q1	It was clear how to start a conversation with the chatbot.
			Q2	It was easy for me to understand how to start the interaction with the chatbot.
			Q3	I find it easy to start a conversation with the chatbot.
2	Accessibility	The ease with which the user can access the chatbot	Q4	The chatbot was easy to access.
			Q5	The chatbot function was easily detectable.
			Q6	It was easy to find the chatbot.
3	Expectation setting	The extent to which the chatbot sets expectations for the interaction with an emphasis on what it can and cannot do	Q7	Communicating with the chatbot was clear.
			Q8	I was immediately made aware of what information the chatbot can give me
			Q9	It is clear to me early on about what the chatbot can do.
4	Communication effort	The ease with which the chatbot understands a range of user input	Q10	I had to rephrase my input multiple times for the chatbot to be able to help me.
			Q11	I had to pay special attention regarding my phrasing when communicating with the chatbot
			Q12	It was easy to tell the chatbot what I would like it to do
5	Ability to maintain themed discussion	The ability of the chatbot to maintain a conversational theme once introduced and keep track of context	Q13	The interaction with the chatbot felt like an ongoing conversation.
			Q14	The chatbot was able to keep track of context.
			Q15	The chatbot maintained relevant conversation.
6	Reference to service	The ability of the chatbot to make references to the relevant service	Q16	The chatbot guided me to the relevant service.
			Q17	The chatbot is using hyperlinks to guide me to my goal.
			Q18	The chatbot was able to make references to the website or service when appropriate
7	Perceived privacy	The extent to which the user feels the chatbot protects one's privacy	Q19	The interaction with the chatbot felt secure in terms of privacy.
			Q20	I believe the chatbot informs me of any possible privacy issues.
			Q21	I believe that this chatbot maintains my privacy.

8	Recognition and facilitation of the user's goal and intent	The ability of the chatbot to understand the user's intention and help them accomplish their goal	Q22	I felt that my intentions were understood by the chatbot.
			Q23	The chatbot was able to guide me to my goal
			Q24	I find that the chatbot understands what I want and helps me achieve my goal.
9	Relevance	The ability of the chatbot to provide information that is relevant and appropriate to the user's reques	Q25	The chatbot gave relevant information during the whole conversation
			Q26	The chatbot is good at providing me with a helpful response at any point of the process
			Q27	The chatbot provided relevant information as and when I needed it.
10	Maxim of quantity	The ability of the chatbot to respond in an informative way without adding too much information	Q28	The amount of received information was neither too much nor too less
			Q29	The chatbot gives me the appropriate amount of information
			Q30	The chatbot only gives me the information I need.
11	Graceful breakdown	The ability of the chatbot to respond appropriately when it encounters a situation it cannot handle	Q31	The chatbot could handle situations in which the line of conversation was not clear.
			Q32	The chatbot explained gracefully when it could not help me.
			Q33	When the chatbot encountered a problem, it responded appropriately
12	Understandability	The ability of the chatbot to communicate clearly and in an easily understandable manner	Q34	I found the chatbot's responses clear.
			Q35	The chatbot only states understandable answers
			Q36	The chatbot's responses were easy to understand.
13	Perceived credibility	The extent to which the user believes the chatbot's responses to be correct and reliable	Q37	I feel like the chatbot's responses were accurate.
			Q38	I believe that the chatbot only states reliable information.
			Q39	It appeared that the chatbot provided accurate and reliable information.
14	Perceived speed	The ability of the chatbot to respond timely to user's requests	Q40	The time of the response was reasonable.
			Q41	My waiting time for a response from the chatbot was short
			Q42	The chatbot is quick to respond.

Note. Adapted from "Assessing User Satisfaction with Information Chatbots: A Preliminary Investigation" by D. Balaji and S. Borsci, 2019, Master's Thesis, University of Twente.

Table A2.
The USIC's original wording, its initial and final translation to German and back translations to English

			Back translation		
	Original	Initial translation	Translator 1	Translator 2	Final translation
Q1	It was clear how to start a conversation with the chatbot	Es war deutlich, wie man ein Gespräch mit dem Chatbot beginnt	It was clear how to start a conversation with the chatbot	It was clear how to start a conversation with the chatbot	Es war deutlich, wie man ein Gespräch mit dem Chatbot beginnt
Q2	It was easy for me to understand how to start the interaction with the chatbot	Es war leicht für mich zu verstehen, wie man eine Interaktion mit dem Chatbot beginnt	It was easy for me to understand how to start an interaction with the chatbot	It was easy for me to understand how to start the interaction with the chatbot	Es war leicht für mich zu verstehen, wie man eine Interaktion mit dem Chatbot beginnt
Q3	I find it easy to start a conversation with the chatbot.	Ich finde es leicht ein Gespräch mit dem Chatbot zu beginnen.	I find it easy to start a conversation with the chatbot	I find it easy to start a conversation with the chatbot	Ich fand es leicht ein Gespräch mit dem Chatbot zu beginnen
Q4	The chatbot was easy to access.	Der Chatbot war leicht zu erreichen	The chatbot was easy to access	The chatbot was easy to reach	Der Chatbot war leicht zu erreichen
Q5	The chatbot function was easily detectable.	Die Chatbot-Funktion war leicht zu erkennen	The chatbot was easy to recognize	The chatbot function was easy to detect.	Die Chatbot-Funktion war leicht zu erkennen
Q6	It was easy to find the chatbot.	Es war leicht den Chatbot zu finden	It was easy to find the chatbot	It was easy to find the chatbot	Es war leicht den Chatbot zu finden
Q7	Communicating with the chatbot was clear	Die Kommunikation mit dem Chatbot war deutlich	The communication with the chatbot was clear	The communication with the chatbot was clear	Die Kommunikation mit dem Chatbot war deutlich
Q8	I was immediately made aware of what information the chatbot can give me.	Ich wurde sofort darauf aufmerksam gemacht, welche Informationen der Chatbot mir geben kann	I was immediately made aware what information the chatbot can give me	I was immediately made aware what information the chatbot can give me	Ich wurde sofort darauf aufmerksam gemacht, welche Informationen der Chatbot mir geben kann
Q9	It is clear to me early on about what the chatbot can do	Es war früh klar für mich was der Chatbot kann	It was clear early on what the chatbot can do	It was clear for me early on what the chatbot can do	Es war früh klar für mich was der Chatbot kann
Q10	I had to rephrase my input multiple times for the chatbot to be able to help me	Ich musste meine Anfrage mehrfach neu formulieren, damit der Chatbot mir helfen konnte	I had to rephrase my request multiple times before the chatbot could help me	I had to rephrase my request multiple times so the chatbot could help me	Ich musste meine Anfrage mehrfach neu formulieren, damit der Chatbot mir helfen konnte
Q11	I had to pay special attention regarding my phrasing when communicating with the chatbot.	Ich musste bezüglich meiner Formulierungen besonders aufmerksam sein, als ich mit dem Chatbot kommuniziert habe	I had to pay close attention regarding my phrasing when I was communicating with the chatbot	I had to pay special attention to my phrasing when I was communicating with the chatbot	**Ich musste mit meinen Formulierungen besonders aufmerksam sein**, als ich mit dem Chatbot kommuniziert habe

Q12	It was easy to tell the chatbot what I would like it to do.	Es war leicht dem Chatbot zu sagen was ich von ihm möchte	It was easy to tell the chatbot what I would like it to do	It was easy to tell the chatbot what I would like it to do	Es war leicht dem Chatbot zu sagen, was ich von ihm möchte
Q13	The interaction with the chatbot felt like an ongoing conversation.	Die Interaktion mit dem Chatbot hat sich angefühlt wie ein fliessendes Gespräch.	The interaction with the chatbot felt like an ongoing conversation	The interaction with the chatbot felt like an ongoing conversation	Die Interaktion mit dem Chatbot hat sich angefühlt wie ein fliessendes Gespräch
Q14	The chatbot was able to keep track of context	Der Chatbot war in der Lage den Kontext im Auge zu behalten	The chatbot was able to keep track of context	The chatbot was able to keep track of context	Der Chatbot war in der Lage den Kontext im Auge zu behalten
Q15	The chatbot maintained relevant conversation	Der Chatbot hielt ein relevantes Gespräch aufrecht	The chatbot maintained a relevant conversation	The chatbot maintained a relevant conversation	Der Chatbot hielt ein relevantes Gespräch aufrecht
Q16	The chatbot guided me to the relevant service	Der Chatbot führte mich zum relevanten Service	The chatbot guided me to the relevant service	The chatbot guided me to the relevant service	Der Chatbot führte mich zum relevanten Service
Q17	The chatbot was using hyperlinks to guide me to my goal	Der Chatbot benutzte Hyperlinks um mich zu meinem Ziel zu führen	The chatbot used hyperlinks to guide me to my goal	The chatbot used hyperlinks to guide me to my goal	Der Chatbot benutzte Hyperlinks um mich zu meinem Ziel zu führen
Q18	The chatbot was able to make references to the website or service when appropriate.	Der Chatbot war in der Lage einen Bezug zu der Webseite oder dem Service herzustellen falls nötig	The chatbot was able to make reference to the website or service when needed	The chatbot was able to make references to the website or service if needed	Der Chatbot war in der Lage einen Bezug zu der Webseite oder dem Service herzustellen **wenn** nötig
Q19	The interaction with the chatbot felt secure in terms of privacy	In Bezug auf Privatsphäre, hat sich die Interaktion mit dem Chatbot sicher angefühlt.	Regarding privacy, the interaction with the chatbot felt secure	The chatbot felt secure in terms of privacy	In Bezug auf Privatsphäre, hat sich die Interaktion mit dem Chatbot sicher angefühlt
Q20	I believe the chatbot informs me of any possible privacy issues.	Ich glaube, dass der Chatbot mich über potientielle Datenschutzprobleme informiert.	I believe that the chatbot informs me about potential privacy issues	I believe that the chatbot informs me about potential privacy issues	Ich glaube, dass der Chatbot mich über potentielle Datenschutzprobleme informiert
Q21	I believe that this chatbot maintains my privacy	Ich glaube das dieser Chatbot meine Privatsphäre wahrt.	I believe that the chatbot maintains my privacy	I believe that this chatbot maintains my privacy	Ich glaube, dass der Chatbot meine Privatsphäre wahrt
Q22	I felt that my intentions were understood by the chatbot.	Ich hatte das Gefühl, dass der Chatbot meine Intention verstanden hat.	I felt that the chatbot understood my intention	I had the feeling that the chatbot understood my intention	Ich hatte das Gefühl, dass der Chatbot meine Intention verstanden hat
Q23	The chatbot was able to guide me to my goal.	Der Chatbot war in der Lage mich zu meinem Ziel zu bringen	The chatbot was able to guide me to my goal	The chatbot was able to guide me to by goal	Der Chatbot war in der Lage mich zu meinen Ziel zu **führen.**
Q24	I find that the chatbot understands what I want and helps me achieve my goal	Ich finde, dass der Chatbot versteht was ich will und mir hilft mein Ziel zu erreichen.	I find that the chatbot understands what I want and helps me to reach my goal	I find that the chatbot understands what I want and helps me to reach my goal	Ich finde, dass der Chatbot versteht was ich will und mir hilft mein Ziel zu erreichen

Q25	The chatbot gave relevant information during the whole conversation	Der Chatbot gab mir das ganze Gespräch über relevante Informationen	The chatbot gave relevant information during the whole conversation	The chatbot provided me with relevant information during the whole conversation	Der Chatbot gab mir das ganze Gespräch über relevante Informationen
Q26	The chatbot is good at providing me with a helpful response at any point of the process.	Der Chatbot ist gut darin, mir in jedem Moment der Interaktion hilfreiche Antworten zu geben	The chatbot is capable to give me helpful answers at any point in the interaction	The chatbot is good to provide me with relevant information during any point in the conversation	Der Chatbot ist gut darin mir in jedem Moment der Interaktion hilfreiche Antworten zu geben
Q27	The chatbot provided relevant information as and when I needed it.	Der Chatbot gab mir relevante Informationen als ich sie brauchte.	The chatbot gave me relevant information when I needed it	The chatbot gave me relevant information when I needed it	Der Chatbot gab mir relevante Informationen als ich sie brauchte
Q28	The amount of received information was neither too much nor too less	Die Menge an Informationen war weder zu viel noch zu wenig	The amount of information was neither too much nor too less	The amount of information was neither to much nor too less	Die Menge an Informationen war weder zu viel noch zu wenig
Q29	The chatbot gives me the appropriate amount of information	Der Chatbot gibt mir eine angemessene Menge an Informationen	The chatbot gives an appropriate amount of information	The chatbot gives me an appropriate amount of information	Der Chatbot **gab** mir eine angemessene Menge an Informationen
Q30	The chatbot only gives me the information I need	Der Chatbot gibt mir nur die Informationen die ich brauche	The chatbot only gives me the information I need	The chatbot only gives me the information I need	Der Chatbot gab mir nur Informationen die ich brauchte
Q31	The chatbot could handle situations in which the line of conversation was not clear	Der Chatbot konnte mit Situationen umgehen in denen das Gespräch nicht deutlich war	The chatbot could handle situations where the conversation was not clear	The chatbot could handle situations in which the conversation was not clear	Der Chatbot konnte mit Situationen umgehen, in denen das Gespräch nicht deutlich war
Q32	The chatbot explained gracefully when it could not help me	Der Chatbot erklärte mir in kontrollierter Weise, wenn er mir nicht helfen konnte	The chatbot explained gracefully when it could not help me	The chatbot explained in a controlled manner when it could not help me	Der Chatbot **erklärte mir freundlich** wenn er mir nicht helfen konnte
Q33	When the chatbot encountered a problem, it responded appropriately	Wenn der Chatbot auf ein Problem stiess, antwortete er angemessen.	When the chatbot encountered a problem, it answered appropriately	When the chatbot encountered a problem, it responded appropriately	Wenn der Chatbot auf ein Problem stiess, **reagierte** er angemessen
Q34	I found the chatbot's responses clear	Ich fand die Antworten des Chatbots deutlich.	I find the answers the chatbot gave were clear	I found the answers of the chatbot were clear	Ich fand die Antworten des Chatbots deutlich.
Q35	The chatbot only states understandable answers.	Der Chatbot gibt nur verständliche Antworten.	The chatbot gives only understandable answers	The chatbot only gives understandable answers	Der Chatbot gab nur verständliche Antworten
Q36	The chatbot's responses were easy to understand.	Die Antworten des Chatbots waren leicht zu verstehen	The answers from the chatbot were easy to understand	The answers of the chatbot were easy to understand	Die Antworten des Chatbots waren leicht zu verstehen
Q37	I feel like the chatbot's responses were accurate	Ich glaube**,** dass die Antworten des Chatbots korrekt waren	I believe that the chatbots answers were accurate	I believe that the answers of the chatbot were correct	**Ich hatte das Gefühl**, dass die Antworten des Chatbots korrekt waren

Q38	I believe that the chatbot only states reliable information.	Ich glaube, dass der Chatbot nur verlässliche Informationen gibt	I believe the chatbot gives only reliable information	I believe that chatbot only gives reliable information	Ich glaube der Chatbot gibt nur verlässliche Informationen
Q39	It appeared that the chatbot provided accurate and reliable information	Es schien, dass der Chatbot genaue und verlässliche Informationen gab	It seemed that the chatbot only gave precise and reliable information	It appeared that the chatbot only gave precise and reliable information	Es schien, dass der Chatbot genaue und verlässliche Informationen gab
Q40	The time of the response was reasonable	Die Reaktionszeit war angemessen.	The reaction time was appropriate	The reaction time was appropriate	Die Reaktionszeit des Chatbots war angemessen
Q41	My waiting time for a response from the chatbot was short.	Meine Wartezeit auf eine Antwort des Chatbots war kurz.	My waiting time for the chatbot was short	My waiting time for an answer of the chatbot was short	Meine Wartezeit auf eine Antwort des Chatbots war kurz
Q42	The chatbot is quick to respond.	Der Chatbot antwortet schnell.	The chatbot answers quickly	The chatbot is fast to respond	Der Chatbot **reagierte** schnell

Appendix B

Session Script (English)

<Participant and researcher will set up a connection using Whereby>

"Hi. My name is Jasmin. Welcome and thank you for taking the time to participate in today's study. Are you ready to start and have me explain what we are going to do?

<Check if participant is ready to start>

"Great. Before we start, can you silence or switch-off your phone for the duration of the session? Also, please let me know if my microphone might encounter any issues.

For this research, we are evaluating a questionnaire to capture the user's satisfaction for chatbots. Today, I will ask you to work with five chatbots. For each chatbot, we have a brief scenario and one or two tasks. After every task, I will ask you to fill out a questionnaire about your experience with the chatbot. The questionnaire has 42 questions.

Please don't feel nervous or under any kind of pressure. It is not a test of how well you interact with the chatbots. Rather, we are interested in your honest feedback on the chatbots, which you can give by filling out the questionnaire. The session is scheduled to last an hour. Do you have any questions at this point?"."

"First, I will send you a link to start the research. I would like you to open the link in a new browser tab."

<Participant opens Qualtrics>

"I have an informed consent form for you. I would like you to read the form. Please let me know if you have any questions. I like to point out that we will make a recording of the audio and screen- today for data-analysis purposes. I will let you know once I will start it. Please let me know if you are not comfortable with this.

If you are ok with everything the form notes, please tick the boxes below the form."

<Have participants read the informed consent form and tick the boxes>

"Before we start off with the chatbot tasks, I have a couple of questions for you regarding your background. Could you please fill these out?"

<Participant fills out background questionnaire.>

"Ok, I would like to ask you to share your screen with me. Please close any other windows on your computer, if you don't want me to see those."

<Participant shares screen>

"Then we will start now with the first task. I will start the screen- and audio-recording now."

<Start audio and screen recording>

"Today, you will work on a task with five chatbots. While you are working on the chatbot task, I will like to share your thoughts with me and tell me what you do and see. The chatbot can send you a link, you can click on these if you like, but I will like to ask to not go much further into the website than that particular page.

The last two out of five chatbots are English. I would like to ask you to talk English with these chatbots. If you have any difficulty with the language, you can ask me for help.

If anything is unclear, you can ask me. However, I may not be able to answer all questions to not influence the research.

Let me know once you achieved the task or if you feel the task is not achievable. We can then move on to the questionnaire.

<Participant works on chatbot tasks and fills out questionnaires>

That completes all of the planned activities for today. Do you have any questions or comments? If you know someone who will participate in today's study, I will like to ask you to not discuss the study. Thank you very much for participating in this study.

Appendix C

Table C1.
List of the chatbots used for the English/Dutch version with task prompts

Chatbot	English task prompt	Dutch task prompt
	English chatbots	
Absolut	You want to buy a bottle of Absolut vodka to share with your friends for the evening. One of your friends cannot consume gluten. You want to use Absolut's chatbot to find out if Absolut Lime contains gluten or not.	Je wilt een fles Absolut wodka kopen om te delen met je vrienden 's avonds. Een van je vrienden mag geen gluten innemen. Je wilt de Absolut chatbot gebruiken om te weten te komen of Absolut Lime wodka gluten bevat of niet. *Let op: dit is een Engelse chatbot. Schrijf je vraag in het Engels.*
ATO	You moved to Australia from the Netherlands recently. You want to know when the deadline is to lodge/submit your tax return using ATO's chatbot to find out.	Je bent recentelijk vanuit Nederland naar Australië verhuisd. Je wilt weten wanneer de deadline is om je belastingsaangifte te doen en gebruikt de ATO chatbot om meer te weten te komen. *Let op: dit is een Engelse chatbot. Schrijf je vraag in het Engels.*

HSBC UK	You live in the Netherlands but are travelling to Turkey for 2 weeks. During your travel, you would like to be able to use your HSBC credit card overseas at payment terminals and ATMS. You want to use HSBC's chatbot to find out the relevant procedure.	Je woont in Nederland en reist voor twee weken naar Turkije. Tijdens je reis wil je graag je HSBC credit card kunnen gebruiken bij betaal- en geldautomaten. Je wilt de HSBC chatbot gebruiken om de relevante procedure te weten te komen. *Let op: dit is een Engelse chatbot. Schrijf je vraag in het Engels.*
United States Citizenship and Immigration Services (USCIS)	You are a US citizen living abroad and want to vote in the upcoming federal elections. You want to use the USCIS chatbot to find out how.	Je bent een Amerikaanse staatsburger die in het buitenland woont. Je wilt stemmen bij de komende federale verkiezingen. Je wilt de USCIS chatbot gebruiken om uit te vinden hoe je dat kunt doen. *Let op: dit is een Engelse chatbot. Schrijf je vraag in het Engels.*
Amtrak	You have planned a trip to the USA. You are planning to travel by train from Boston to Washington D.C. You want to stop at New York to meet an old friend for a few hours and see the city. You want to use Amtrak's chatbot to find out how much it will cost to temporarily store your luggage at the station.	Je hebt een trip naar de VS gepland en bent van plan om de trein te nemen van Boston naar Washington D.C.. In New York City wil je je reis voor een paar uur onderbreken, omdat je met een oude vriend hebt afgesproken om samen de stad te bekijken. Door gebruik te maken van de Amtrak chatbot wil je erachter komen hoe veel het gaat kosten om je baggage tijdelijk op te geven bij het station. *Let op: dit is een Engelse chatbot. Schrijf je vraag in het Engels.*
Inbenta	You have an interview with Inbenta in a few days and you want to use Inbenta's chatbot to find out the address of Inbenta's Mexico office.	Je hebt een interview met Inbenta en je wil met behulp van de Inbenta chatbot het adres van de vestiging in Mexico opzoeken. *Let op: dit is een Engelse chatbot. Schrijf je vraag in het Engels*.
HelloFresh	You are thinking about joining HelloFresh and want to know how it works and what the average price per meal is based on two persons.	Je denkt erover naar een abonnement van HelloFresh te nemen. Je wil te weten komen hoe het werkt en wat de gemiddelde prijs per gerecht is op basis van twee personen. *Let op: dit is een Engelse chatbot. Schrijf je vraag in het Engels.*
	Dutch chatbots	
Amsterdam Medisch Centrum	You need to get your blood tested at the Amsterdam Medical Center (AMC). You want to use the chatbot to find out where in the hospital you need to be and what the procedure is for blood sampling.	Je moet bloed laten prikken in het Amsterdam Medisch Centrum (AMC) voor een onderzoek. Je wilt de chatbot gebruiken om erachter te komen waar je in het ziekenhuis moet zijn, en wat de procedure is bij bloedprikken.
A.S.R.	Your motorbike has been hit while you were parked at a gas station. You can't continue driving. You are insured with ASR and visit the website to report the damage and see if you can get a replacement vehicle.	Je motor is aangereden terwijl je geparkeerd stond bij een benzinestation. Je kunt niet meer verder rijden. Je bent verzekerd bij ASR en gebruikt de chatbot om de schade te melden en om te kijken of je vervangend vervoer kunt krijgen.
Bol.com	You forgot to buy a present for a friend who is celebrating her birthday tonight, and you want to buy a 10 euro Bol.com gift card. You want to	Je bent vergeten een cadeau te kopen voor een vriendin die vanavond haar verjaardag viert en je wilt nog snel een Bol.com cadeaukaart van 10

	use the Bol.com chatbot to find out in which shop you can buy the gift card and what is the lowest amount you can put on a gift card.	euro kopen. Je wilt de Bol.com chatbot gebruiken om erachter te komen in welke winkel je de cadeaukaart kunt kopen en wat het laagste bedrag is wat je op een cadeaukaart kunt zetten.
KPN	You are a KPN customer and have a prepaid SIM card for your mobile phone. You need new prepaid credit and you want to use the chatbot to find out how long prepaid credit is valid after purchase. Additionally, you have a new account number that you want to pass on to KPN. You want to use the KPN chatbot to find out how you can change your account number.	Je bent klant bij KPN en hebt een prepaid simkaart voor je mobiele telefoon. Je hebt nieuw prepaid tegoed nodig. Je wilt met behulp van de KPN chatbot te weten komen hoe lang prepaid tegoed geldig is na aankoop. Daarnaast heb je een nieuw rekeningnummer dat je door wilt geven aan KPN. Je wilt door middel van de KPN chatbot te weten komen hoe je jouw rekeningnummer kunt wijzigen.
Oxxio	You're considering switching to the Oxxio's green energy. However, the contract with your current energy supplier has not yet ended, and your energy supplier will impose a cancellation penalty if you switch suppliers before the end date. You want to use the chatbot to find out whether Oxxio will pay this fine for you if you switch to Oxxio.	Je overweegt om over te stappen naar de duurzame stroom van Oxxio. Het contract bij je huidige energieleverancier is echter nog niet afgelopen, en je energieleverancier rekent een opzegboete als je voor de einddatum overstapt. Je wilt er met behulp van de chatbot achter komen of Oxxio deze boete betaalt voor jou als naar Oxxio overstapt.
Vattenfall	You are a Vattenfall customer and receive monthly 'exclusive points', which you can donate to charity, among other things. You want to ask the chatbot which charities you can donate these 'exclusive points' to.	Je bent klant bij Vattenfall en krijgt maandelijks 'exclusief punten', die je o.a. kunt doneren aan het goede doel. Je wilt de chatbot vragen aan welke goede doelen je deze 'exclusief punten' kunt doneren.

Table C2.
List of the chatbots used for the German version with task prompts

Chatbot	German task prompt	English translation
Elster	Sie wollen Ihre Steuererklärung machen und nutzen dafür das Online-Portal 'ELSTER'. Sie sind bereits auf MeinElster registriert und haben ein sogenanntes 'Elster-Zertifikat" das Ihre Identität bestätigt. Da Sie vor kurzem umgezogen sind, möchten Sie nun wissen, wie Sie Ihre Adresse bei ELSTER ändern können und ob Sie noch weitere Schritte unternehmen müssen.	You want to file your tax return and use the 'ELSTER' online portal to do so. You are already registered on MeinElster and have a so-called 'Elster certificate' that confirms your identity. Since you have recently moved, you would now like to know how you can change your address at ELSTER and whether you still need to take further steps.
Lufthansa	Sie haben im März bei Lufthansa einen Flug von Frankfurt nach Los Angeles für den 15. April gebucht. Aufgrund der aktuellen Lage, wurde der Flug jedoch annuliert. Sie wollen sich noch nicht auf einen neuen Abflugtermin festlegen, deshalb wollen Sie mit Hilfe des Lufhansa Chatbots herausfinden, ob Sie Ihr Ticket auch in einen Gutschein umtauschen können und wie Sie diesen beantragen.	In March you booked a flight from Frankfurt to Los Angeles for April 15 with Lufthansa. Due to the current situation, however, the flight was canceled. You don't want to commit to a new departure date yet, so you want to use the Lufhansa chatbot to find out whether you can exchange your ticket for a voucher and how to apply for it.

Congstar	Sie sind Kunde beim Mobilfunkunternehmen Congstar und haben eine Prepaidkarte. Da Sie diesen Monat mehr mit Ihrem Handy im Internet gesurft haben, ist ihr Datenvolumen bereits aufgebraucht. Sie wollen nun zusätzliches Datenvolumen zubuchen und möchten mit Hilfe des Congstar-Chatbots herausfinden wie es funktioniert und wieviel Sie 1GB Daten kosten.	You are a customer of the cell phone company Congstar and have a prepaid card. Since you have surfed the Internet a lot on your mobile phone this month, your data volume has already been used up. You now want to book additional data and would like to find out with the help of the Congstar chatbot how it works and how much 1GB of data costs you.
OTTO	Sie wollen eine neue Waschmaschine bei OTTO kaufen und möchten Ihre alte Waschmaschine fachgerecht entsorgen. Da Sie jedoch kein Auto haben, können Sie Ihr altes Gerät nicht selbst zum Schrotthandel bringen. Sie wollen deshalb mit Hilfe des OTTO-Chatbots herausfinden, ob OTTO eine Altgerätemitnahme anbietet und wieviel diese kostet.	You want to buy a new washing machine from OTTO and want to dispose of your old washing machine properly. However, since you don't have a car, you can't take your old device to the scrap dealer yourself. You therefore want to find out with the help of the OTTO chatbot whether OTTO offers to take old devices and how much it costs.
Deutsche Bahn	Sie haben Ihre BahnCard25 verloren und möchten diese nun ersetzen. Sie wollen mit Hilfe des Bahn-Chatbots in Erfahrung bringen, wie Sie eine Ersatzkarte beantragen können und wieviel diese kostet.	You have lost your BahnCard25 and now want to replace it. With the help of the Bahn chatbot, you want to find out how you can apply for a replacement card and how much it costs.
WienBot	Sie wohnen in Wien und Ihr Personalausweis läuft in Kürze ab. Sie wollen mit Hilfe des Chatbots herausfinden, welche Dokumente sie für die Beantragung des neuen Personalausweises mitbringen müssen und wieviel dieser kostet.	You live in Vienna and your identity card will expire soon. You want to use the chatbot to find out which documents you need to bring with you to apply for a new ID card and how much it costs.
Troisdorfer Stadtwerke	Sie sind Stromkunde bei den Troisdorfer Stadtwerken. Da sich Ihre Bankverbindung vor Kurzem geändert hat, wollen Sie mit Hilfe des Chatbots herausfinden, wie sie Ihre hinterlegten Daten ändern können.	You are an electricity customer at Troisdorf public utilities. Since your bank details have recently changed, you want to use the chatbot to find out how you can change your stored data.

Table C3.
ATI scale as proposed by Franke, Attig & Wessel (2019)

In the following questionnaire, we will ask you about your interaction with technical systems. The term "technical systems" refers to apps and other software applications, as well as entire digital devices (e.g., mobile phone, computer, TV, car navigation).

	Please indicate the degree to which you agree/disagree with the following statements.	completely disagree	largely disagree	slightly disagree	slightly agree	largely agree	completely agree
01	I like to occupy myself in greater detail with technical systems.	☐	☐	☐	☐	☐	☐
02	I like testing the functions of new technical systems.	☐	☐	☐	☐	☐	☐
03	I predominantly deal with technical systems because I have to.	☐	☐	☐	☐	☐	☐
04	When I have a new technical system in front of me, I try it out intensively.	☐	☐	☐	☐	☐	☐
05	I enjoy spending time becoming acquainted with a new technical system.	☐	☐	☐	☐	☐	☐
06	It is enough for me that a technical system works; I don't care how or why.	☐	☐	☐	☐	☐	☐
07	I try to understand how a technical system exactly works.	☐	☐	☐	☐	☐	☐
08	It is enough for me to know the basic functions of a technical system.	☐	☐	☐	☐	☐	☐
09	I try to make full use of the capabilities of a technical system.	☐	☐	☐	☐	☐	☐

Appendix D

Table D1

Preliminary 32- item version of the USIC

Item	Description	Feature	F1 Communication quality	F2 Conversation start	F3 Perceived speed	F4 Perceived privacy
Q1	It was clear how to start a conversation with the chatbot.	Ease of starting a conversation	0.149	**0.688**	0.156	
Q2	It was easy for me to understand how to start the interaction with the chatbot.		0.242	**0.709**	0.143	
Q3	I find it easy to start a conversation with the chatbot.		0.285	**0.697**	0.209	
Q4	The chatbot was easy to access	Accessibility	0.131	**0.795**	0.150	0.106
Q5	The chatbot function was easily detectable		0.161	**0.824**	0.107	
Q6	It was easy to find the chatbot.		0.137	**0.810**		0.102
Q7	Communicating with the chatbot was clear.	Expectation setting	**0.706**	0.372	0.164	0.102
Q10	I had to rephrase my input multiple times for the chatbot to be able to help me.	Communication effort	**0.645**		0.104	
Q12	It was easy to tell the chatbot what I would like it to do.		**0.688**	0.307	0.102	
Q14	The chatbot was able to keep track of context.	Ability to maintain themed discussion	**0.757**	0.139	0.164	
Q15	The chatbot maintained relevant conversation.		**0.677**	0.145	0.130	
Q16	The chatbot guided me to the relevant service.	Reference to service	**0.603**	0.239	0.275	0.176
Q19	The interaction with the chatbot felt secure in terms of privacy.	Perceived privacy	0.188	0.158	0.140	**0.819**
Q21	I believe that this chatbot maintains my privacy.		0.106		0.103	**0.767**
Q22	I felt that my intentions were understood by the chatbot.	Recognition and facilitation of the users' goal and intent	**0.891**	0.105	0.145	
Q23	The chatbot was able to guide me to my goal.		**0.711**	0.193	0.248	0.174
Q24	I find that the chatbot understands what I want and helps me achieve my goal.		**0.835**	0.192	0.212	0.100
Q25	The chatbot gave relevant information during the whole conversation.	Relevance	**0.789**	0.182	0.213	
Q26	The chatbot is good at providing me with a helpful response at any point of the process.		**0.813**	0.236	0.216	

Q27	The chatbot provided relevant information as and when I needed it.		**0.795**	0.133	0.186	0.129
Q28	The amount of received information was neither too much nor too less.	Maxim of quantity	**0.745**		0.158	
Q29	The chatbot gives me the appropriate amount of information.		**0.773**		0.267	
Q30	The chatbot only gives me the information I need.		**<u>0.798</u>**	0.100	0.104	
Q31	The chatbot could handle situations in which the line of conversation was not clear.	Graceful breakdown	**<u>0.438</u>**		0.175	-0.122
Q34	I found the chatbot's responses clear.	Understandability	**<u>0.777</u>**	0.152	0.280	0.113
Q35	The chatbot only states understandable answers.		**0.728**	0.205	0.241	
Q36	The chatbot's responses were easy to understand.		**0.659**	0.220	0.293	0.128
Q37	I feel like the chatbot's responses were accurate.	Perceived credibility	**0.732**	0.121	0.325	0.177
Q39	It appeared that the chatbot provided accurate and reliable information.		**<u>0.754</u>**	0.133	0.296	0.140
Q40	The time of the response was reasonable.		0.294	0.197	**0.821**	
Q41	My waiting time for a response from the chatbot was short.	Perceived speed	0.268	0.121	**0.880**	
Q42	The chatbot is quick to respond.		0.286	0.139	**<u>0.893</u>**	

Note. Item's highest factor loading in boldface and feature's highest factor loading underlined