27/01/2023

Artificial Intelligence Conversational Agents: Using Card Sorting To Re-Evaluate The Chatbot Usability Scale

Bachelor Thesis

Sander Verschoor

S2322226

First supervisor: MSc Jule Landwehr

Second supervisor: Dr. Borsci

University of Twente

BMS Faculty

Department of Psychology

Abstract

Research previously published on chatbot quality of interaction by Borsci et al. (2021, 2022) resulted in a Bot Usability Scale consisting of 11 items (BUS-11). This scale aims to support the development and assessment of chatbots, and with that increase user satisfaction in the utilization of such systems. With their latest research showing the potential of models that vary in their number of factors (Borsci & Schmettow, 2023) the present research aimed to explore the mental model of chatbot users to determine what model would be the most accurate to use. To uncover the natural mental model of chatbot users this study made use of open card sorting. During card sorting the participants were tasked to organize the items of the BUS-11 in groups. A total of 58 participants took part in the study, resulting in 47 usable results. The results were presented in a heatmap to visualize the found mental model. The heatmap revealed a mental model that was nearly identical to both the 4 and 5-factor models of the BUS-11. Along with these models, the heatmap also presented outlying points of interest. Based on this data the results suggested a model that exists of 4 factors, with one of these factors including 2 sub-factors.

Keywords: Chatbots, Usability, Chatbot Usability Scale, User Satisfaction, Card Sort, Heatmap

Acknowledgments

To start, I would like to thank my supervisor MSc Jule Landwehr for all the patience, care, advice, and even emotional support that they have given me during this thesis. Because of this, I managed to write something that I feel proud of.

Furthermore, I would like to thank Chiara Wermter for collaborating with me and helping me when I felt lost. Working together has been a true pleasure.

Finally, I would like to thank my friends, family, and especially my girlfriend for always supporting me, and providing me with a comfortable environment to be in.

Contents

Introduction

With fast advances in technology and its growing integration into society, the utilization of conversational agents has become increasingly prominent in daily life activities such as listening to music, noting down the grocery list, and customer support (Li & Wang, 2023). Conversational agents, or chatbots, are applications of artificial intelligence that are designed for user interaction by making use of their natural language represented in text or speech (Borsci et al., 2021; Jiang et al., 2022; Rese et al., 2020). An example of such a conversational agent is Google Assistant, an application that listens to the requests of its user and responds accordingly by performing tasks or providing information. Although early chatbots existed in the 1960s (Rese et al., 2020) their shape and purpose have shifted over the years. Due to the rising interest in the use of technology and automation in daily life, the concept of a conversational agent has been adopted in a variety of settings such as smart home devices and customer service (Borsci et al., 2021).

This possibility to automate interaction is not just limited to the use in people's homes, nowadays businesses often make use of chatbots instead of human personnel (Xu et al., 2022). A company, for example, could utilize conversational agents to provide clients with guidance on their website for a longer time, without the need for extra employment. When associated with company service, Borsci et al. (2021) refer to these tools as customer relationship management (CRM) chatbots. They state that the utilization of these CRM chatbots does not only provide 24/7 interaction with customers but may also be of financial interest as costs in staff are reduced. As these chatbots are often designed to correspond to a specific service their features can vary massively (Borsci et al., 2021; Borsci et al., 2022; Xu et al., 2022). A chatbot could for example be used to redirect a customer to the correct information or contact person but could also provide them with assistance while shopping for a new outfit.

Regardless of these benefits there still exists a reluctance for customers to make use of these automated chatbot systems (Rese et al., 2020; Xu et al., 2022). This reluctance is likely caused by unfamiliarity or negative prior experiences. According to Rizomyliotis et al. (2022), failure to understand input and even fear of using chatbots makes cause of interactions that do not satisfy the needs of the user. This dissatisfaction is likely to be caused by the appearance of inhuman characteristics such as impersonal communication and the feeling of being observed by

the conversational agent (Rese et al., 2020). Xu et al. (2022) also express that further humanization of chatbots will improve the customers' willingness and satisfaction of needs. According to the International Organization for Standardization, (ISO, 2018) satisfaction is a part of usability. More explicitly they note usability as the level to which a system is usable by its target audience in order to accomplish pre-determined goals with effectiveness, efficiency, and satisfaction. Therefore, decreasing the levels of dissatisfaction is of interest to increase the usability of CRM chatbots. Borsci et al. (2021) state that satisfaction is a difficult measure as it includes end-user reaction and reasoning. This is in the context of systems and relates to efficiency, effectiveness, accuracy, and reliable modalities of assessment.

To improve the development and customization of these CRM chatbots Borsci et al. (2021) proposed a usability scale to assess user satisfaction with conversational agents. This scale is set up to be a standardized evaluation tool based on the already available chatbot attribute list that was proposed by Radziwill and Benton (2017). The aim of this list originally was to provide designers with a set of guidelines. Borsci et al. (2021), however, used this list to construct an inventory with the purpose of measuring the satisfaction of chatbot usage. They state that previous tools used for measuring usability lack in evaluating the interaction between user and chatbot. This interaction is a characterising element of conversational agents and is therefore important to include in a chatbot usability scale, especially in terms of satisfaction.

Borsci et al. (2021) set their goal to support chatbot designers by developing appropriate tools to evaluate the interactive quality of conversational agents. This is a concept that they could not find any previous studies of at the time. To accomplish this goal, they performed a sequence of four studies. During the first study the possible attributes that may be used by end-users in their assessment of their recent interaction with an information retrieval chatbot were identified by making use of a systematic review. In the second study they developed an online survey with chatbot designers as well as end-users to reach a consensus on this list of attributes. To further expand on the attribute list and start developing the evaluation tool the third study consisted of focus group sessions. With this an initial version of their scale was developed, the Bot Usability Scale (BUS). In the fourth and final study, the initial BUS was piloted, and its psychometric properties were explored to develop a final version of the BUS (Borsci et al., 2021).

The version of the BUS that was developed by Borsci et al. (2021) is referred to as the BUS-15 as the final product concluded with 15 items (derived from the initial 42 items associated with chatbot interaction). These items were divided into 5 factors: perceived accessibility to chatbot functions, perceived quality of chatbot functions, perceived quality of conversation and information provided, perceived privacy and security, and time response. However, in their confirmatory factorial analysis, Borsci et al. (2022) tested the psychometric properties of the BUS-15 as well as potential alternative factorial models. For this, an analysis with the designometric perspective was also conducted, which highlights the difference between the purpose of a user experience self-report scale and psychometrics (Schmettow, 2021). This difference is in the focus of the metrics. While psychometrics is focussing on people, a user experience self-report sets its purpose on comparing designs. Being aware of this difference is important, as leaning to either of these extremes could lead to either lesser functionality or dissatisfaction from end-users. From this confirmatory factorial analysis, Borsci et al. (2022) concluded that another reliable version of the BUS is more suitable, the BUS-11. This new scale still consists of 5 factors, however, as can be seen in Table 1, the number of items that are divided among these factors has decreased to a total of 11. They state that the BUS-11 is a shorter and more reliable solution than the BUS-15.

Table 1

BUS-11 5-Factor Model Items per Factor

Factor	Item
1 – Perceived Accessibility to Chatbot Functions	1. The chatbot function was easily detectable
	2. It was easy to find the chatbot
2 – Perceived Quality of Chatbot Functions	3. Communicating with the chatbot was clear
	4. The chatbot was able to keep track of context
	5. The chatbot's responses were easy to understand
3 – Perceived Quality of Conversation and Information Provided	6. I find that the chatbot understands what I want and helps me achieve my goal

	7. The chatbot gives me the appropriate amount of information
	8. The chatbot only gives me the information I need
	9. I feel like the chatbot's responses were accurate
4 – Perceived Privacy and Security	10. I believe the chatbot informs me of my possible privacy issues
5 – Time Response	11. My waiting time for a response from the chatbot was short

In currently unpublished research Borsci and Schmettow (2023) have discovered another reasonable option of the Bot Usability Scale. Their results showed that two factors (Factors 2 and 3) of the original BUS-11 could be merged to become a sole factor called Functional Interactive Conversation. With both structures of the BUS-11 (four and five factors) being acceptable and reliable possibilities, it becomes a point of interest whether the 4-factor scale is more effective than the 5-factor scale.

The first Bot Usability Scale by Borsci et al. (2021) was designed with the consideration of chatbot designers and end-users. However, the newer versions that have been developed in further research have been modified and evaluated based on mainly statistical analyses. To ensure that the usability scale is still compatible with its users it needs to once again be compared with their natural mental model, as this might diverge from what is presented by a statistical analysis. A mental model is the categorization of items, made in a person's mind (Schmettow & Sommer, 2016; Ntouvaleti & Katsanos, 2022). These models, or knowledge structures, can vary between groups of people as they have different perspectives. As it is difficult to predict such mental models, it is required to make use of empirical methods to acquire data on the subject. For this card sorting is often used. During card sorting research participants are asked to categorize a given set of items. This can either be pre-made categories (closed card sorting) or the participant is free to create their own categories (open card sorting). Ntouvaleti and Katsanos (2022) also explain a hybrid version of card sorting in which participants are asked to create sub-

categories under pre-defined ones. In terms of results there are no differences between manual (real world) card sorting and digital card sorting (using software).

In this paper open card sorting is used as an explorative analysis to unfold the natural mental model of participants in context of the Bot Usability Scale. By Comparing this mental model with both the 4 and 5-factor models of the pre-existing BUS-11 the aim is to determine what its most optimal model would be. To accomplish this first it will be established for each of these models to what extend they correspond to the participants' natural mental model. Additionally, the ambiguities results will be investigated and their imporantce for the development of the BUS-11 will be evaluated.

Methods

The data in this study was collected using a card sort, a qualitative research method, with the items of the pre-existing BUS-11 by Borsci et al. (2022). Before starting data collection, ethical approval was obtained from the ethical committee of the University of Twente. Before starting the final study, a total of two pilot studies were performed, each containing 2-4 participants. This was done to improve the fluency and unambiguousness of the study.

Participants

A total of 58 participants took part in the study. Of these participants 39 were female and 16 were male, with an age range of 18-59 and a mean age of 23 (SD: 6.38). The card sort was completed by 54 participants. The study was available in three different languages, and the participants completed the card sort in the ratios of: 8 Dutch, 16 German, and 30 English. All entries that had not been properly completed were eliminated from the study, resulting in 47 usable results. Proper compltetion refers to reaching the end of the study while following the instructions as intended.

Convenience sampling was used for participant gathering alongside the use of the Sona System. The Sona System is a platform of the University of Twente on which a study can be published to reach out to students. These students are incentivised with course credits in return for their participation.

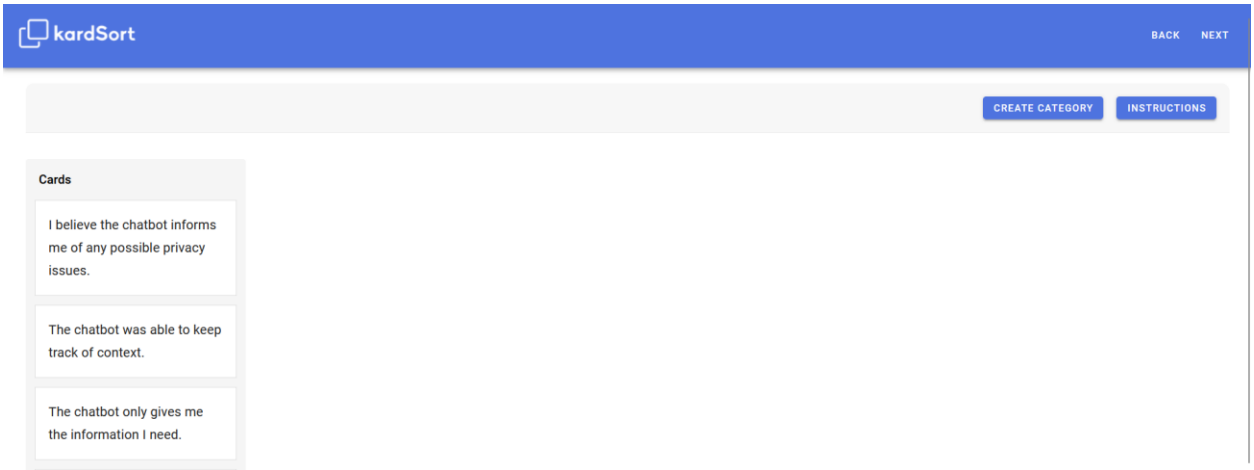
Materials

For the collection of demographical data, Qualtrics was used. Qualtrics is an online platform designed for the collection and storage of data. The Qualtrics part of the study contained descriptive explanations of the study, informed consent, a survey containing questions about the participants' demographics, and a task using a chatbot that was followed by the pre-existing BUS-11 by Borsci et al. (2022) (see Appendix A). This task was used to familiarise the participants with the BUS-11 and decrease misinterpretation of the items. In the final step the participants were provided with a unique ID and sent to a second platform, KardSort.

The KardSort platform is specialised in collecting data using the card sorting method. At the time of this study, the options available on this platform were either a closed or an open card sort. Additionally, KardSort provides the option for survey questions, although this functionality is limited. The KardSort part of the study included a page requesting the participant-ID gained from the Qualtrics platform, instructions on the card sort, and the card sort itself. As can be seen in Figure 1, during the card sort the participant was presented with the items of the BUS-11 on the left side of the screen. On the right side of the screen buttons were available to create a new category or be presented once again with the instructions.

Figure 1

Card Sort on KartSort Platform

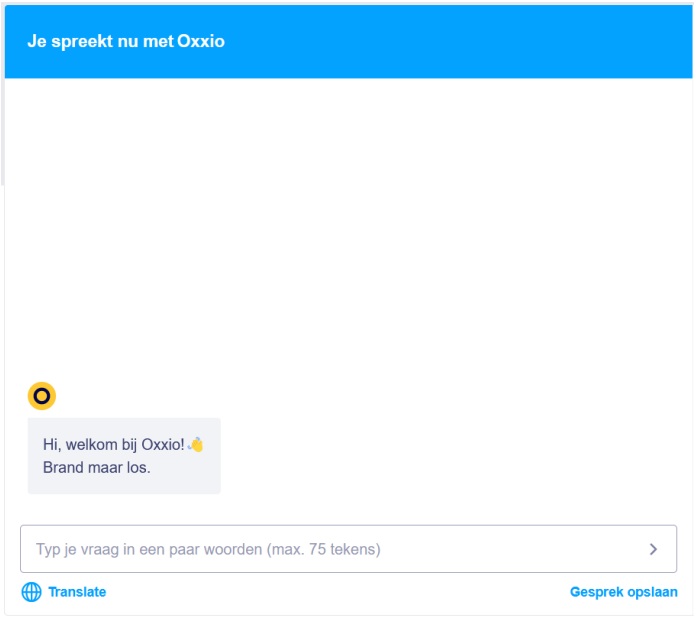


Both the Qualtrics and the KardSort surveys were available in Dutch, German, and English, making use of the respective versions of the BUS-11 provided by Borsci et al. (2022). The participants could choose for themselves in what language they preferred to complete the

study and were asked about their language proficiency in this chosen language. A chatbot from the website oxxio.nl (see Figure 2) was used to provide the participant with experience in chatbot interaction. This chatbot is designed with the aim to support the customer and inform them about the services of the company and using the “translate” button the language can be changed.

Figure 2

Oxxio Chatbot



Procedure

After reading the introduction to the study, the participants were asked to give their informed consent. After consent was given, they were asked to continue the study by filling in the demographics questionnaire. Once completed the participants were asked to interact with a given chatbot to perform a task. After the chatbot interaction, the participant was asked questions related to the task, as well as the questions of the BUS-11 (Borsci et al., 2022). To continue with the study, the participants were now provided with a participant-ID and sent to a different platform on which they had to fill in the given ID and perform a task of open card sorting. Once the card sorting was completed, the participant got presented with a final screen, telling them they have successfully completed the study.

Data analysis

The collected data on Qualtrics and KardSort was exported and then analysed using SynCaps Version 3 and R Studio Version 2022.07.2. Before the actual process of analysis,

however, the answers of each participant had to be checked for correctness. The data from Qualtrics was attainable with all languages put together easily through their translation services, creating a file that is ready for use while KardSort provided a data file for each language, meaning extra preparation was required before analysation could start.

Data Exclusion

The data from KardSort was exported as SynCaps data in a .txt file. Using the SynCaps software, the data was presented in a comprehensible manner. For each language file, first the results for each participant were reviewed and, when necessary, eliminated from the study. The data was eliminated when the participant had not completed the card sort as intended. An example of this is sorting the items between groups called “TRUE” and “FALSE”, which indicates that the participant had failed to understand the assignment, providing an incorrect representation of the mental model. Once all unusable data was filtered from each file the data sets were merged into a singular document.

Heatmap

A heatmap was made to visualize the clustering of items. For the values used in the heatmap, the data collected was analysed by making use of the Jaccard coefficient. With this coefficient, the similarity of the items was assessed. This was done by dividing the number of groups that two items are both a part of by the number of groups that at least one of these items is a part of (Schmettow & Sommer, 2016). As this study had only one grouping level, the values were either 0 or 1. Using SynCaps this process automatically occurred for each participant, creating an Item-to-Item matrix per participant. SynCaps then merged all participant matrices into a final matrix (see Appendix B). This matrix was then imported into RStudio and generated as a heatmap using the package ComplexHeatmap (see Appendix C).

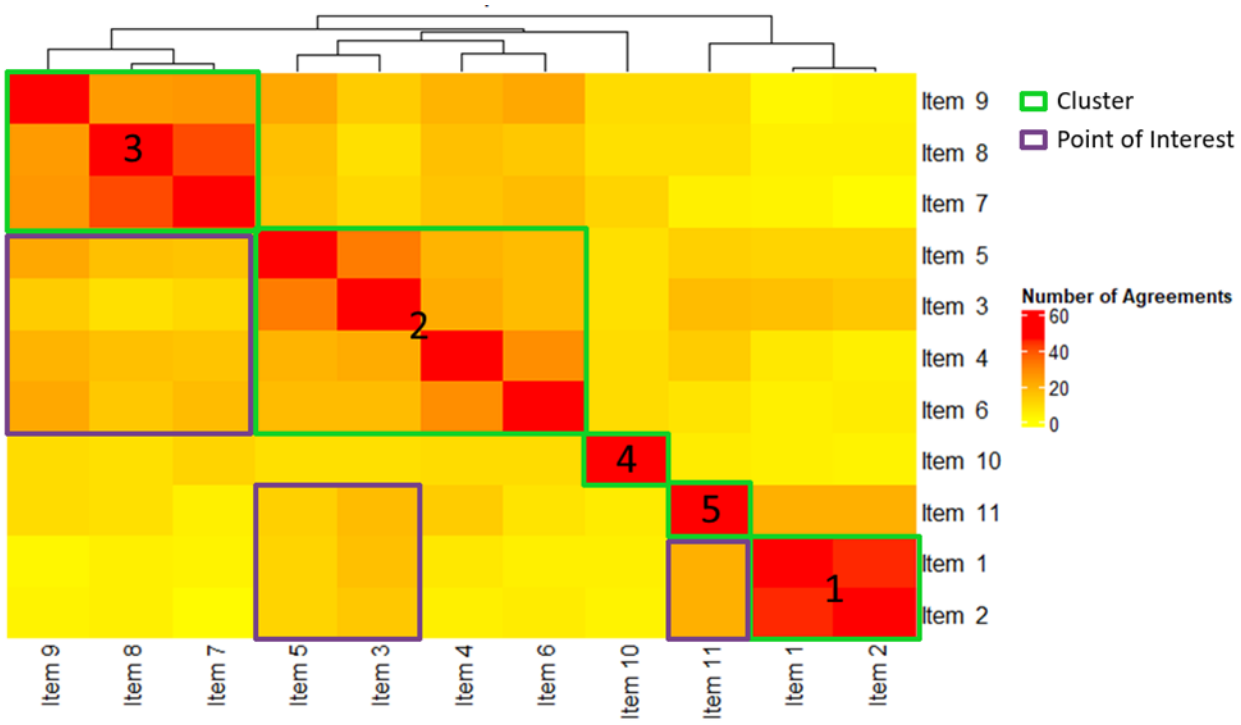
As the aim is to evaluate the BUS-11 model it is important to ensure that this study is actually measuring what it is intended to measure. This is the level of face validity (Johnson, 2021). The face validity was measured by comparing the heatmap of the found mental model with the BUS-11 by Borsci et al. (2022). The level of similarity between the models determined the face validity of the scale.

Results

The heatmap in Figure 3 displays the grouping of items by the participants on an item-item basis. Each coloured square in the heatmap represents the number of times a participant has put the 2 connected items in the same group. The lowest value (0) appears as yellow, and the highest value (47) appears as red. With these values, the mental model of the participants is displayed. The heatmap shows 5 clear clusters among the items, these clusters are outlined with a green border. The dendrograms on top of the heatmap also show the strength of the clusters. The lower the connecting points go, the stronger the connection between the items is. In addition to these clusters, some outlying points of interest will also be discussed. These points are outlined on the heatmap with a purple border. The clusters on this heatmap represent the participants' mental model. This mental model will be compared with the pre-existing BUS-11 model by Borsci et al. (2022). Clusters were formed by first prioritizing the areas with a value of 50% and higher. After that the surrounding cells of these areas were evaluated, where the focus lied on medium range values (40 – 50%) and particularly the connection with surrounding values. As a final determinator of the clusters each of the statistically derived clusters was looked at to ensure they consisted of items that are relevant to each other in terms of their context. When looking at the dendrograms some further connections could be found, creating the possibility of different models. The following paragraphs will describe two potential models that are visible in the heatmap, the before mentioned 5-factor model, and a fused 4-factor model.

Figure 3

Clustered Heatmap of Card Sort with 5 Factors



Note: Clustered heatmap of performed card sort with presenting 5 factors. Each cluster is numbered (1 – 5) and is outlined in green. Outlined in purple are points of interest, or ambiguities, that are not necessarily part of a cluster. Dendrograms on top of the figure display de strength of the connection between al items/clusters, the higher the dendrogram, the weaker the connection.

5-Factor Model

A total of 5 clusters can be seen in the heatmap of Figure 3 based on the grouping of items by the participants (see Appendix D). Each of these clusters makes up a factor, together forming a 5-factor model. By analysing the input from the participants these factors were named: *Accessibility*, *Interaction Quality*, *Accuracy*, *Privacy*, and *Responsiveness*.

Accessibility

The first cluster contains items 1 and 2. Both items are by definition related to locating the chatbot. The participant often grouped these items together under categories such as "*Design of Chatbot"*, "*Accessibility*", and "*Ease of Finding*", therefore, this cluster was dubbed as the

factor *Accessibility*. This factor aims at how the accessibility to the functions of the chatbot is perceived.

Interaction Quality

Included in this cluster are the items 3, 4, 5, and 6. Communication and context of conversation are topics that are discussed using these items. Some common categories these items were placed under were "*Response Quality*", "*Communication*", and "*Content*". As these were all related to the interaction between chatbot and user, especially in the context of quality, this factor was called *Interaction Quality*.

Accuracy

The items 7, 8, and 9 form this cluster. The subject of these items is focused on the information that is received from the chatbot. Categories appointed to these items were "*Content of Message*", "*Information Provided*", and "*Accuracy*". Hence, this cluster became the factor *Accuracy*, aiming at the correctness of the information, as well as the quantity of information that is given.

Privacy

Item 10 is about privacy issues, and more specifically being informed about them. This is also visible in the participant responses, as they put this item under category names such as "*Trust*", "*Privacy*", and "*Transparency*". Therefore, this became a factor called *Privacy*. This factor measures how privacy and safety are perceived by the user.

Responsiveness

Although item 11 is often grouped together with other items such as those in the first cluster (item 1 and 2), more often is it put under its own category. The participants gave this category a name related to time and duration. Therefore, item 11 has its own factor called *Responsiveness*.

Remaining Points of Interest

Besides the clusters mentioned above, there are more possible connections between items. The items 1, 2, and 11, for example, show a shared connection with items 3 and 5. Notable here, as mentioned above, is that in the heatmap item 11 is highly related to item 1 and 2. Another point of interest is the connection between the items 4, 5, and 6 and the *Accuracy* cluster

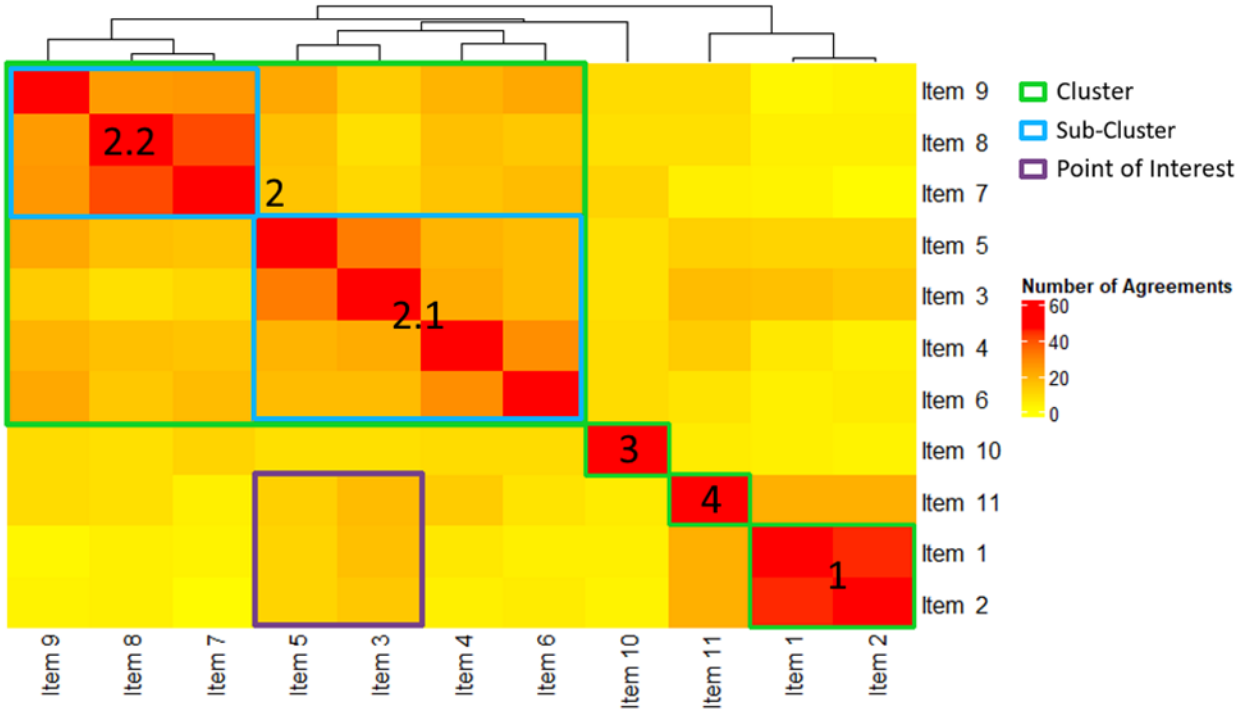
(items 7, 8, and 9). This connection is visible with all items in this cluster, but namely with item 9. Item 9 also shows a higher value of the agreement with item 3, although this is not as high as with items 4, 5, and 6.

4-Factor Model

Another model that can be seen consists of only 4 clusters, or factors (see Figure 4). This model is similar to the 5-factor model. However, the factors *Interaction Quality* and *Accuracy* are fused together as one factor. This factor in its entirety would be focused on the functionality of a chatbot, especially concerning the quality of its functions.

Figure 4

Clustered Heatmap of Card Sort with 4 Factors



Note: Clustered heatmap of performed card sort with presenting 4 factors. Each cluster is numbered (1 – 4) and is outlined in green. 2 sub-clusters are noted as part of cluster 2 (2.1 and 2.2) and displayed with a blue border. Outlined in purple are points of interest, or ambiguities, that are not necessarily part of a cluster. Dendrograms on top of the figure display de strength of the connection between al items/clusters, the higher the dendrogram, the weaker the connection.

Comparison BUS-11 Model

Both the mental model of the participants and the BUS-11 model show possibilities for either a 4 or a 5-factor model. Comparing these models with each other, respectively, shows both similarities as differences. Table 2 and 3 show a comparison of the mental model with the 4 and 5-factor models of the BUS-11 respectively. For each item, the tables show whether there exists an agreement (Y), partial agreement (P), or disagreement (N) between the models. Agreement exists when items of the mental model are grouped with the same items as the BUS-11 model, with the exception margin of 1 item. Partial agreement exists when an item is in a different factor compared to the BUS-11 model, but still shows high values in the heatmap towards the items of its factor in the BUS-11 model.

Table 2

Comparison BUS-11 4-Factor Model with Participants' Mental Model; Agreement (Y), Partial Agreement (P), Disagreement (N)

Item	Factor BUS-11 (4-factor)	Factor Mental Model	Match
1 The chatbot function was easily detectable	Accessibility	Accessibility	Y
2 It was easy to find the chatbot	Accessibility	Accessibility	Y
3 Communicating with the chatbot was clear	Functional interactive conversation	Interaction Quality	P
4 The chatbot was able to keep track of context	Functional interactive conversation	Interaction Quality	P
5 The chatbot's responses were easy to understand	Functional interactive conversation	Interaction Quality	P
6 I find that the chatbot understands what I want and helps me achieve my goal	Functional interactive conversation	Interaction Quality	P

7 The chatbot gives me the appropriate amount of information	Functional interactive conversation	Accuracy	P
8 The chatbot only gives me the information I need	Functional interactive conversation	Accuracy	P
9 I feel like the chatbot's responses were accurate	Functional interactive conversation	Accuracy	P
10 I believe the chatbot informs me of any possible privacy issues	Privacy	Privacy	Y
11 My waiting time for a response from the chatbot was short	Responsiveness	Responsiveness	Y

Table 3

Comparison BUS-11 5-Factor Model with Participants' Mental Model; Agreement (Y), Partial Agreement (P), Disagreement (N)

Item	Factor BUS-11 (5-factor)	Factor Mental Model	Match
1 The chatbot function was easily detectable	Accessibility	Accessibility	Y
2 It was easy to find the chatbot	Accessibility	Accessibility	Y
3 Communicating with the chatbot was clear	Quality of Functions	Interaction Quality	Y
4 The chatbot was able to keep track of context	Quality of Functions	Interaction Quality	Y
5 The chatbot's responses were easy to understand	Quality of Functions	Interaction Quality	Y

6 I find that the chatbot understands what I want and helps me achieve my goal	Quality of Conversation and Information Provided	Interaction Quality	P
7 The chatbot gives me the appropriate amount of information	Quality of Conversation and Information Provided	Accuracy	Y
8 The chatbot only gives me the information I need	Quality of Conversation and Information Provided	Accuracy	Y
9 I feel like the chatbot's responses were accurate	Quality of Conversation and Information Provided	Accuracy	Y
10 I believe the chatbot informs me of any possible privacy issues	Privacy and Security	Privacy	Y
11 My waiting time for a response from the chatbot was short	Time Response	Responsiveness	Y

Differences

As can be seen in Table 3, when comparing the mental model with the BUS-11 5-factor model, the only difference is found in item 6. Borsci et al. (2022) grouped item 6 along with items 7, 8, and 9, while the mental model shows item 6 to be stronger related to items 3, 4, and 5.

Regarding the BUS-11 4-factor model, the fundamental difference in comparison to the mental model is the separation of the factor *Functional Interaction Conversation* into the factors *Interaction Quality* and *Accuracy* (see Table 2).

Similarities

Overall, the mental model and pre-existing models of the BUS-11 appear to be similar. The factors *Accessibility* (items 1 and 2), *Privacy* (item 10), and *Responsiveness* (item 11) are identical in all models. The high amount of similarity to the pre-existing models shows that the study is measuring the intended context meaning a high degree of face validity.

When comparing the 5-factor models with each other they are very similar. Although there is a difference in the allocation of item 6, the heatmap shows that it still has a high agreement score with items 7, 8, and 9, just as in the Borsci et al. (2022) model.

As mentioned before, an overlap between the factors *Interaction Quality* and *Accuracy* is visible in the heatmap. With these factors fused together, the 4-factor mental model would be identical to the 4-factor model of the BUS-11 as together they form its factor *Functional Interactive Conversation*.

Discussion

The aim of this study was to use card sorting as an explorative analysis and with that unfold the natural mental model of participants in the context of the Bot Usability Scale, developed by Borsci et al. (2022). This was done to make a comparison between the models and determine whether the 4 and 5-factor structures of the BUS-11 correspond to the natural mental model of the participants and what would be the most accurate model to use. The results show a total of 5 clear factors that are mostly in line with the 5-factor model of the BUS-11. However, ambiguities in the heatmap, in addition to a strong similarity with the 4-factor model of the BUS-11, show a potential for a fusion of factors, creating a 4-factor mental model that is identical to the 4-factor model of the BUS-11.

The Factor Model of the Bot usability scale

The research question relates to determining the correspondence between the 4 and 5-factor structures of the BUS-11 and the participants' mental model. As the results of this study show a 5-factor mental model, it is indicated that there is at least a resemblance of the pre-existing model in the mental model. When comparing these models with each other the results show a high level of similarity as there is only a difference in one of the items. It should also be noted that even this difference is minimal, as this item still shows high agreement values with the items it is grouped with in the pre-existing model. These results show that the statistically built BUS-11 model with 5 factors has a very high coherence with the mental model of the participants.

One outstanding difference between the models, however, is the relation between the factors of *Accessibility* and *Responsiveness*. When looking at the items from these factors in a

more subjective way this connection seems peculiar. The items 1 (*The chatbot function was easily detectable*) and 2 (*It was easy to find the chatbot*) are both related to locating the chatbot, while item 11 (*My waiting time for a response from the chatbot*) is related to the spend time on waiting for a response from the chatbot. In terms of context, these two factors do not seem relevant to each other, showing a possible misunderstanding by participants. This misunderstanding could either be about the items themselves, or a misunderstanding of the card sorting study.

The results also indicate a potential 4-factor model where the factors *Interaction Quality* and *Accuracy* are fused together as one factor. With this combined factor the mental model would have a total of 4 factors, and in terms of factors it would be identical to the 4-factor model of the BUS-11. This shows that indeed the 4-factor model by Borsci et al. (2022) is also an accurate representation of the mental model of the participants.

Although the 4-factor mental model perfectly matches the 4-factor model of the BUS-11, the outlying values of agreement in the heatmap should also be looked at. The heatmap does show higher levels of agreement between the items from the factors of *Interaction Quality* and *Accuracy*, however, these levels are notably lower and inconsistent compared to the smaller clusters that make up these separated factors. This indicates that the participants at least felt like there was a difference between the items, therefore showing the importance of some sort of division of this factor, while keeping these items connected.

The difference between the 5-factor mental model and the 5-factor model of the BUS-11 concerning item 6 is also highly relevant. Since this item is part of different factors between the two 5-factor models, and these factors are fused together in the 4-factor model, this shows that most of the connection between these factors could be relying on item 6. This is also enforced by the heatmap as the connection with items 7, 8, and 9 is the strongest for item 6. With this being the case, simply fusing these factors to form the 4-factor model would be a less accurate representation of the mental model of the participants than the 5-factor model.

A more advanced model can be found when exploring the option of sub-factors. This would lead to a 4-factor model as concluded by prior research (Borsci & Schmettow, 2023). However, the second factor, *Quality of Conversation and Information Provided* (*Interaction Quality* and *Accuracy* Fused, would still be divided to an extent. These sub-factors are

highlighted in blue in Figure 4. A division in sub-factors would increase accuracy as both the overlap and differences between the items of the factor are included, leading to a better representation of the user's mental model.

Additional Ambiguities

In addition to the above-mentioned outlying agreement values something that should not be overlooked is the connection between items 1, 2, and 11 and items 3 and 5. Notable is that this occurrence is not visible in the confirmatory factor analysis of Borsci et al. (2022), as factor loadings between these factors are below 0.40, a value that falls far below the threshold of 0.6. Although the values on the heatmap between these items are not necessarily on the high end, it still shows a pattern in reasoning among participants. This could either be because the items are somewhat related to the mental model, or some other factor related to either the card sorting study or the items from the BUS-11. In terms of the final mental model, however, these values are too low to be incorporated into the factors.

Limitations

Since the participants for this study were collected through convenience sampling, and both researchers were students at the same university, there is a potential limitation in the representation of the target group. As the social network was limited to family and peers/ fellow students, most participants ended up being in the same group, 18-30 years old and often being a student. In addition to the social network the SONA system of the University of Twente was used, which led to a lot of the participants. The participants gathered from this system are all students, enforcing the limits even further. Chatbots are used by all sorts of users, and therefore having a sample with mostly students might not bring an accurate representation of a general mental model, but instead, a mental model that is more focussed on younger people. Future research should expand their study to a broader and more varying pool of participants such as people in a shopping centre.

Another potential limitation is in the quality of instructions to the participants while performing the card sort. Some of the collected data was eliminated because the participants had not correctly understood what their task was. For example, on occasion, they sorted the cards under "true" and "false", showing that they understood the card sort to just be a different format from the actual BUS-11 they had done right before. To prevent this from happening in future

research a gap or break could be created between tasks and the difference between tasks could be mentioned more explicitly. In addition to misunderstanding the instructions the outlying values in the heatmap also show that potentially the participants have interpreted item 11 of the BUS-11 differently from its intended definition. Although the participants had been exposed to the actual BUS-11 moments before, this could perhaps be more thoroughly avoided by providing the participants with a short description of each of the items prior to the card sorting task. While performing the card sort, the participants had no context to relate the items to, therefore providing them with the freedom to interpret their own subject, changing their reasoning behind their grouping.

Implications

Even with these limitations, the results of this study further advance the progression of the BUS-11. As chatbots are being used more often for both company and personal use it is important for chatbot developers to be aware of the effectiveness of their products, especially user satisfaction as a part of usability. With the aim of the BUS-11 to enhance the process of both evaluation and development of chatbots, the insights into the natural mental model gained from this study can be used to make the scale more accurate, and potentially bring interest in further research on the topic.

Conclusion

By analysing the items that were sorted in groups by the participants, this thesis has provided a 4-factor mental model which included 2 sub-factors. Based on a comparative analysis between both models it can be concluded that both the 4 and 5-factor models of the BUS-11 highly correspond to the mental model of the partcipants. However, since the data still reveals a notable connection between factor *Interaction Quality* and *Accuracy*, this combined factor model appears most accurate. With this new model of the BUS-11, the scale can be used to accurately evaluate and design (CRM) chatbots. To further improve and validate the BUS-11, future research could be done at the item level, ensuring that each item is interpreted as intended. Additionally, the implementation of the scale in both chatbot development and chatbot assessment should be evaluated to gain a more definitive overview of its effectiveness.

References

Borsci, S., Malizia, A., Schmettow, M., van der Velde, F., Tariverdiyeva, G., Balaji, D., & Chamberlain, A. (2021, July 21). The Chatbot Usability Scale: the Design and Pilot of a Usability Scale for Interaction with AI-Based Conversational Agents. *Personal and Ubiquitous Computing*, *26*(1), 95–119. https://doi.org/10.1007/s00779-021-01582-9

Borsci, S., Schmettow, M., Malizia, A., Chamberlain, A., & van der Velde, F. (2022, August 4). A confirmatory factorial analysis of the Chatbot Usability Scale: a multilanguage validation. *Personal and Ubiquitous Computing*. https://doi.org/10.1007/s00779-022-01690-0

Borsci, S., & Schmettow, M. (2023). The Universality of the chatbot Usability Scale (BUS-11): from a psychometric to a designometric perspective. Manuscript in preparation.

International Organization for Standardization. (2018). *Ergonomics of Human-System Interaction – Part 11: Usability: Definitions and concepts* (ISO/DIS Standard No. 9241-11). Retrieved from https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en

Jiang, H., Cheng, Y., Yang, J., & Gao, S. (2022, September). AI-powered chatbot communication with customers: Dialogic interactions, satisfaction, engagement, and customer behavior. *Computers in Human Behavior*, *134*, 107329. https://doi.org/10.1016/j.chb.2022.107329

Johnson, E. (2021). Face Validity. *Encyclopedia of Autism Spectrum Disorders*, 1957–1957. https://doi.org/10.1007/978-3-319-91280-6_308

Li, M., & Wang, R. (2023). Chatbots in e-commerce: The effect of chatbot language style on customers' continuance usage intention and attitude toward brand. *Journal of Retailing and Consumer Services*, *71*, 103209. https://doi.org/10.1016/j.jretconser.2022.103209

Ntouvaleti, M., & Katsanos, C. (2022, April 27). Validity of the Open Card Sorting Method for Producing Website Information Structures. *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. https://doi.org/10.1145/3491101.3519734

Radziwill, N. M., & Benton, M. C. (2017). Evaluating quality of chatbots and intelligent conversational agents. arXiv preprint arXiv:1704.04579.

Rese, A., Ganster, L. & Baier, D. (2020, September). Chatbots in retailers' customer communication: How to measure their acceptance? *Journal of Retailing and Consumer Services*, *56*, 102176. https://doi.org/10.1016/j.jretconser.2020.102176

Rizomyliotis, I., Kastanakis, M. N., Giovanis, A., Konstantoulaki, K., & Kostopoulos, I. (2022, December). “How mAy I help you today?” The use of AI chatbots in small family businesses and the moderating role of customer affective commitment. *Journal of Business Research*, *153*, 329–340. https://doi.org/10.1016/j.jbusres.2022.08.035

Schmettow M (2021). New statistics for design researchers. Springer International Publishing, Cham

Schmettow, M., & Sommer, J. (2016, April 12). Linking card sorting to browsing performance – are congruent municipal websites more efficient to use? *Behaviour &Amp; Information Technology*, *35*(6), 452–470. https://doi.org/10.1080/0144929x.2016.1157207

Xu, Y., Zhang, J., & Deng, G. (2022, July 22). Enhancing customer satisfaction with chatbots: The influence of communication styles and consumer attachment anxiety. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.902782

Appendix

Appendix A: Qualtrics Pages

Introduction

Dear participant,

Thank you for participating in this research! **Please use a PC or laptop to participate in this study!**

Before we begin, you will receive information about the research and your rights. Taking part in this research is voluntary, and you can withdraw at any moment. Withdrawing will not have negative consequences for you.

Purpose of the research
This research is about investigating how humans experience interaction with different chatbots. Chatbots are programmes that communicate with you like a human but do not require a human to operate.

Content of the research
Taking part in this research consists of different components. After reading this introduction, we will ask you to give informed consent and inquire about your demographics. Further, we will ask you to interact with a chatbot and evaluate your experience. Lastly, you will be redirected to a different website to sort cards.

Data processing
The data of this research will be used to gather the mental models of our participants. The information will be used for a Bachelor Thesis. Your data will be anonymised and cannot be traced back to you. Your data will not be shared with third parties. The anonymised information is stored in a secure environment and kept for use in future studies. This research is approved by the Ethics Committee of the University of Twente.

You can navigate this survey by clicking the arrows at the bottom on the page. To proceed, please click on the arrow on the bottom right.

Consent Form

Consent Form

Taking part in the study
I have read and understood the study information. I voluntarily take part in this research and understand that I can refuse to answer questions. I know that I can withdraw from this study at any time, without having to give a reason. I understand that I have to interact with a chatbot and that participating does not involve any risks. I am at least 18 years old.

Use of the information in the study
I understand that providing demographic data, interacting with a chatbot, and filling in a questionnaire after is also part of the study. Further, I will be asked to sort cards in the final part of the study

Future use and reuse of the information by others
I understand that the information that I provide will be used for a bachelor thesis. I know that all information will be anonymised and stored in a secure environment. I consent that the anonymised information provided by me is kept for use in future studies.

Contact information for questions about Your rights as a participant
If you have any questions regarding your participation in this study, you can email c.e.wermter@student.utwente.nl. You can also reach the supervisor by emailing j.landwehr@utwente.nl. If you have any questions about your rights as a participant, the use of your data, or other questions and concerns about this research, you can contact the secretariat of the Ethics Committee of the Faculty of Behavioural, Management, and Social Sciences of the University of Twente: ethicscommittee-bms@utwente.nl.

Do you consent to participating in this research?

- ○ Yes, I consent
- ○ No, I want to stop

Demographic Data

Demographic Data 1/2

You will now receive questions about your demographics.

How old are you in years?

What is your current gender identity? (check all that apply)

- ☐ Man
- ☐ Woman
- ☐ Female-to-Male (FTM)/Transgender Male/Trans Man
- ☐ Male-to-Female (MtF)/Transgender Female/Trans Woman
- ☐ Genderqueer, neither exclusively male or female;
- ☐ Additional Gender Category/(or Other), please specify
- ☐ Decline to answer

What is your sex (as assigned at birth)?

- ○ Male
- ○ Female
- ○ Intersex

What is your nationality?

- ○ Dutch
- ○ German
- ○ Other:

What is your level of English proficiency?

- ○ No proficiency- Knowing few to no words; unable to form full sentences.
- ○ Elementary proficiency-Able to form basic sentences and answer simple questions.
- ○ Limited proficiency- Able to use social phrases and carry limited casual conversations.
- ○ Basic proficiency- Having fairly extensive vocabulary and being able to hold conversations.
- ○ Full proficiency- Able to have advanced discussions on a wide range of topics.
- ○ Native/bilingual proficiency- Able to speak completely fluent.

Are you diagnosed with ADHD or ADD?

- ○ Yes
- ○ No

Are you medicated for ADHD/ADD?

- ○ Yes
- ○ No

Demographic Data 2/2

You will now receive a number of statements about your experience with chatbots. Please indicate for each statement how much you agree.

Familiarity

	Fully Disgree	Disagree	Neutral	Agree	Fully Agree
I am familiar with chatbots and/or other conversational interfaces.	○	○	○	○	○
I know how chatbots work.	○	○	○	○	○
I feel confident with using chatbots.	○	○	○	○	○

Before we begin, please indicate how often you use chatbots.

	Never	Seldom	1 time per month	2-3 times a month	4-6 times a month	Daily
How many times do you use a chatbot per months?	○	○	○	○	○	○

Chatbot Task

Explanation Task

We will now ask you to interact with a chatbot. You will receive a link to a website. Please open this website in a second tab. We will then give you a task to solve with the help of the chatbot.

Do not worry if you cannot solve the task. You can simply continue to participate and complete the questionnaire about the chatbot. If you solve the task, please also continue to answer the questionnaire about the chatbot.

Note: You never have to provide any personal information.

Chatbot

Please open the link in a second tab (leave the survey tab open) and find the chatbot. In the next step, you will be asked to solve a task with the chatbot.

To change the chatbot to English, click on "Translate" and select English.

https://www.oxxio.nl/klantenservice

Perform the following task using the chatbot:

What are the advantages of a Smart meter? (two answers)

- [ ] You can view your energy consumption anytime.
- [ ] You can control your home remotely.
- [ ] You do not have to report your meter readings manually.
- [ ] You can automatise processes such as temperature regulation.

Was it possible to complete the task?

- ○ Yes
- ○ No (Why not, please specify)
- ○ I am not sure

Please answer the following questions on the basis of your experience interacting with the chatbot.

	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
The chatbot gives me the appropriate amount of information.	○	○	○	○	○
Communicating with the chatbot was clear.	○	○	○	○	○
The chatbot function was easily detectable.	○	○	○	○	○
The chatbot was able to keep track of context.	○	○	○	○	○
I find that the chatbot understands what I want and helps me achieve my goal.	○	○	○	○	○
It was easy to find the chatbot.	○	○	○	○	○
The chatbot's responses were easy to understand.	○	○	○	○	○
My waiting time for a response from the chatbot was short.	○	○	○	○	○
The chatbot gives me the information I need.	○	○	○	○	○
I believe the chatbot informs me of any possible privacy issues.	○	○	○	○	○
I feel like the chatbot's response was accurate.	○	○	○	○	○

Card Sorting

Card Sorting

Next, we will ask you to sort a number of items into categories.

You will be given a participant-ID. You will be asked to provide this number on the next page. Please copy the ID shown below!

This is your ID: ${e://Field/ID}
Please **copy** this number, you will need to fill it in before starting the card sorting.

Now, please open this link and continue the study there: https://study.kardsort.com/copy-of-chatbot-usability-english

You can now end this part of the survey by clicking on the right arrow.

Appendix B: Item-Item Matrix Card Sort

ItemNo		The chatb	It was eas	My waiting	The chatb	The chatb	I feel the c	Communic	The chatb	The chatb	I find that	I believe th
1	The chatb		45.0	20.0	3.0	4.0	2.0	16.0	11.0	6.0	4.0	4.0
2	It was eas	45.0		20.0	1.0	4.0	3.0	14.0	11.0	4.0	5.0	3.0
11	My waiting	20.0	20.0		4.0	8.0	9.0	17.0	12.0	13.0	7.0	5.0
7	The chatb	3.0	1.0	4.0		41.0	26.0	10.0	15.0	15.0	17.0	11.0
8	The chatb	4.0	4.0	8.0	41.0		25.0	8.0	16.0	16.0	14.0	8.0
9	I feel the c	2.0	3.0	9.0	26.0	25.0		13.0	22.0	19.0	22.0	9.0
3	Communic	16.0	14.0	17.0	10.0	8.0	13.0		32.0	21.0	17.0	8.0
5	The chatb	11.0	11.0	12.0	15.0	16.0	22.0	32.0		19.0	17.0	8.0
4	The chatb	6.0	4.0	13.0	15.0	16.0	19.0	21.0	19.0		28.0	9.0
6	I find that	4.0	5.0	7.0	17.0	14.0	22.0	17.0	17.0	28.0		9.0
10	I believe th	4.0	3.0	5.0	11.0	8.0	9.0	8.0	8.0	9.0	9.0	

Appendix C: R Script Heatmap Generation

```
#install and load libraries
if (!require("tibble")) {
  install.packages("tibble", dependencies = TRUE)
  library(tibble)
}
if (!require("dplyr")) {
  install.packages("dplyr", dependencies = TRUE)
  library(dplyr)
}
```

```
if (!requireNamespace("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
#BiocManager::install("ComplexHeatmap")
library(ComplexHeatmap)
library(circlize)

library(readxl)
#load results matrix
cc_1M  <- read_excel("Combined Results.xlsx")

#define variable for sample size
participant_count <- 47
N = as.character(participant_count)

#rename items to Item #
colnames(cc_1M)[colnames(cc_1M) == "The chatbot function was easily detectable."] = "Item 1"
colnames(cc_1M)[colnames(cc_1M) == "It was easy to find the chatbot."] = "Item 2"
colnames(cc_1M)[colnames(cc_1M) == "Communicating with the chatbot was clear."] = "Item 3"
colnames(cc_1M)[colnames(cc_1M) == "The chatbot was able to keep track of context."] ="Item 4"
colnames(cc_1M)[colnames(cc_1M) == "The chatbot's responses were easy to understand."] ="Item 5"
colnames(cc_1M)[colnames(cc_1M) == "I find that the chatbot undertands what I want and helps me achieve my goal."] ="Item 6"
colnames(cc_1M)[colnames(cc_1M) == "The chatbot gives me the appropriate amount of information."] ="Item 7"
colnames(cc_1M)[colnames(cc_1M) == "The chatbot only gives me the information I need."] ="Item 8"
colnames(cc_1M)[colnames(cc_1M) == "I feel the chatbots reponses were accurate."] ="Item 9"
colnames(cc_1M)[colnames(cc_1M) == "I believe the chatbot informs me of any possible privacy issues."] ="Item 10"
colnames(cc_1M)[colnames(cc_1M) == "My waiting time for a response from the chatbot was short."] ="Item 11"



#deleting 2nd column (Item description)
cc_1M$...2 <- NULL

#replace NA with participant number
```

```
cc_1M[is.na(cc_1M)]<-N

#changing the rownames to the itemNo
reference <- data.frame(cc_1M)
cc_1M$ItemNo <- NULL
row.names(reference) <- reference$ItemNo
rownames(cc_1M) = rownames(reference)

#convert data into numerical matrix
data <- as.matrix(cc_1M)

dims <- dim(data)
data <- as.numeric(data)
dim(data) <-dims

#name rows and columns in prep for heatmap
rownames(data) = paste("Item ", rownames(cc_1M))
colnames(data) = colnames(cc_1M)

#configure and generate heatmap
color_conf = colorRamp2(c(0, participant_count), c("yellow", "red"))
Heatmap(data,
    name = "Number of Agreements",
    show_row_dend = FALSE,
    col = color_conf,
    column_title = "Clustered Heatmap of Card Sort")
```

Appendix D: List of Categories by Participants

Participant	Category	Items
1	How to find	1, 2
	User gratification	6, 10, 11
	Level of response	3, 4, 5, 7, 8, 9
2	Detection	1, 2
	Communication	3, 4, 6, 11

	Answers	5, 7, 8, 9, 10
3	Functions	4, 6, 9, 11
	Accessibility	1, 2, 3, 5
	Information	7, 8, 10
4	Communication	3, 4, 5, 6, 10
	Usability	1, 2, 11
	Information	7, 8, 9
5	Improvement	4, 10
	Smart chatbot	3, 6, 11
	Information given	5, 7, 8, 9
	Accessibility	1, 2
6	Interface	1, 2, 11
	Content of message	7, 8, 10
	Usefulness	3, 4, 5, 6, 9
7	Functionaliy	1, 2, 4, 6
	Information provided	7, 8
	Communication	3, 5, 9, 11
	Privacy	10
8	Privacy issues	10
	Usability	1, 2, 3, 5, 7, 8, 11
	Chatbot's understanding of user request/question	4, 6, 9
9	Privacy and safety	10
	Communication	3, 4, 6, 9, 11
	Information	5, 7, 8
	Ease of finding	1, 2
10	Data privacy	10
	Time	11
	Simplicity and accuracy	7, 8, 9
	Comprehension	3, 4, 5, 6
	Accessibility	1, 2
11	Waiting time	11
	Communication	3, 5

	Privacy	10
	Findability	1, 2
	Chatbot understand context	4, 6
	Given information	7, 8, 9
12	Chatbox	1, 2, 3, 4
	Time	11
	Information	5, 6, 7, 8, 9, 10
13	Subjective Opinion about the experience with the Chatbot	3, 6, 7, 10
	Experience with the chatbot	4, 5, 8, 9, 11
	Finding the Chatbot	1, 2
14	Chatbox's role	6, 10
	Chatbox's answers	5, 7, 8, 9
	Communication with chatbox	3, 4, 11
	Finding chatbot	1, 2
15	Impressions	2, 5, 6, 8, 9
	Efficiency	1, 3, 4, 7, 10, 11
16	Privacy	10
	User interaction	1, 2
	Performance	11
	Quality of the answers of the chatbot	3, 4, 5, 6, 7, 8, 9
17	Helpfulness	6, 7, 8
	Privacy	10
	Communication	3, 4, 5, 9, 11
	Accessibility	1, 2
18	Intelligence of the chatbot	4, 10
	Quality of the answers	5, 6, 7, 8, 9
	Ease of use	1, 2, 3, 11
19	Waiting time	11
	Chat box findability	1, 2
	Chat box information	7, 8, 10
	Chat box communication	3, 4, 5, 6, 9

20	Others	10, 11
	Usability	4, 6
	Accuracy of chatbot	3, 7, 8
	Feeling after using chatbot	1, 2, 5, 9
21	Information Not Given	10
	Waiting time	11
	Information Given	5, 7, 9, 8
	Understanding me	3, 4, 6
	Locating Ease	1, 2
22	Privacy	10
	Finding chatbot function	1,2
	Waiting time	11
	Information provided	3, 4, 5, 6, 7, 8
23	Funcionality	7, 8, 9, 11
	Response quality	3, 4, 5, 6, 10
	Visualization	1, 2
24	Personal bond	6
	Clear	3, 4, 5, 9
	Concise	7
	Transparency	10
	Convenience	1, 2, 11, 8
25	Usability	1, 2, 3, 5, 10
	Functionality	4, 6, 7, 8, 9, 11
26	Waiting time	11
	Quality of information in answer	7, 8, 9
	Contextual understanding	4, 6
	Privacy	10
	Quality of Chatbots expression	3, 5
	Location of Chatbot-icon	1, 2
27	Menge	7, 8
	Ergebnis auf Antwort	3, 4, 5, 10
	Funktion	1, 2, 6, 9, 11
28	Individuelles Gefühl	3, 9, 10

	Funktion	1, 2, 4, 6
	Inhalt	7, 8
	Bequemlichkeit	5, 11
29	Informationsmenge	7, 8
	Vertrauen	9, 10
	Sichtbarkeit	1, 2, 11
	Verständlichkeit	3, 4, 5, 6
30	Bequemlichkeit	1, 2, 3, 5, 10, 11
	Informationen Nutzen	4, 6, 7, 8, 9
31	Information	7, 8, 9
	Datenschutz	10
	User Experience	1, 2, 11
	Inhalt / Verständnis	3, 4, 5, 6
32	Ergebnis	5, 6, 7, 9, 10
	Funktion Chatbot	4, 8, 11
	Bedienung	1, 2, 3
33	Qualität	4, 6, 7, 8, 9
	Sicherheit	10
	Benutzerfreundlichkeit	1, 2, 3, 5, 11
34	3	2, 11
	2	6, 9, 10
	1	1, 3, 4, 5, 7, 8
35	Qualität / skills	4, 6, 7, 8
	Unsicherheit	9, 10
	Chatbot Nutzung	1, 2, 3, 5, 11
36	Benutzerfreundlichkeit des Chatbots	1, 2, 3 ,11
	Informationen des Chatbots	7, 8, 10
	Antworten des Chatbots	4, 5, 6, 9
37	Marketing einstellungen	6, 7, 8
	Technische Leistungen	3, 4, 5, 9, 11
	Designer Leistungen	1, 2, 10
38	Vorteil	11

	Vertrautheit	9, 10
	Compexity	1, 2, 3 ,5
	Funktion des Chatbots	4, 6, 7, 8
39	Regulierung des Chattbott/Bewertung	4, 10, 11
	Design des Chattbott	1, 2, 6
	Bewertung der Chattbott antworten	3, 5, 7, 8, 9
40	Functionalität des Cb	1, 2, 4, 11
	Verständlichleit der Antworten	3, 5
	Datenschutz	10
	Benötigte Informationen	6, 7, 8, 9
41	Context	4, 6
	Privacy	10
	Gebruiksvriendelijkehid	1, 2, 3, 5, 11
	Inhoud	7, 8, 9
42	Vertrouwen	6, 9
	Communicatie	3, 5
	Informatie	4, 7, 8, 10
	Gebruiksvriendelijkheid	1, 2, 11
43	Communicatie	3, 5, 6
	Juistheid	9
	Privacy	10
	Hoeveelheid	4, 7, 8
	Gebruiksvriendelijkheid	1, 2, 11
44	Privacy	10
	Antwoordkwaliteit	3, 6, 7, 9
	Gebruikgemak	4, 5, 8, 11
	Vindbaarheid	1, 2
45	Privacy	10
	Communicatie	3, 5, 6
	Chatbot ontdekken	1, 2, 11
	Juistheid	4, 7, 8, 9

46	Zorgen	10
	Doel	4, 5, 6, 7, 9
	Gebruiks Ervaring	1, 2, 3, 8, 11
47	Gemak van chatbot	1, 2, 3, 5, 11
	Informatie	4, 6, 7, 8, 9, 10