Privacy Strategies for Conversational AI and their Influence on Users' Perceptions and Decision-Making

Anna Leschanowsky
anna.leschanowsky@iis.fraunhofer.de
Fraunhofer IIS
Erlangen, Germany

Birgit Popp
birgit.popp@iis.fraunhofer.de
Fraunhofer IIS
Erlangen, Germany

Nils Peters*
nils.peters@fau.de
International Audio Laboratories Erlangen
Erlangen, Germany

ABSTRACT

Conversational AI (CAI) systems are on the rise and have been widely adopted in homes, cars and public spaces. Yet, people report privacy concerns and mistrust in these systems. Current data protection regulations ask providers to communicate data practices transparently and provide users with options to control their data. However, even if users are given control, their decisions can be subject to heuristics and biases leaving people frustrated and regretful. Based on the idea of conversational privacy and debiasing, we design three privacy strategies for CAI that allow people to have their data deleted while at the same time promoting rational decision-making. We conduct a user study to test our strategies in two widespread scenarios using a text-based CAI system and evaluate their impact on peoples' privacy perception, usability and attitude-behaviour alignment. We find that our strategies can significantly change people's behaviour, but do not influence peoples' privacy perception. Finally, we discuss evaluation metrics and future research directions to investigate privacy controls in Conversational AI systems.

KEYWORDS

Conversational AI, privacy, debiasing, chatbot

ACM Reference Format:
Anna Leschanowsky, Birgit Popp, and Nils Peters. 2023. Privacy Strategies for Conversational AI and their Influence on Users' Perceptions and Decision-Making. In *The 2023 European Symposium on Usable Security (EuroUSEC 2023), October 16–17, 2023, Copenhagen, Denmark.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3617072.3617106

1 INTRODUCTION

One of the key goals of the European General Data Protection Regulation (GDPR) and several other international data protection laws is to strengthen the control individuals have over their personal data [20]. While usable privacy controls have been long researched for traditional user interfaces, similar controls are missing for Conversational AI systems. Yet, with the rise of voice assistants and text-based chatbots and the public discourse around privacy concerns in CAI, research on suitable privacy controls for CAI is timely and much needed. Current privacy controls in CAI often rely on traditional graphical user interfaces which were found insufficient and cumbersome to use [29]. They require users to switch modalities to adjust privacy settings which can increase cognitive load and make users less likely to engage with these settings. Some CAI systems may not have graphical user interfaces at all and therefore, have to rely on purely voice or text-based privacy controls. In addition, CAI systems are missing standards for privacy controls. For instance, at the time of writing this article, Apple's Siri only allows deleting voice recordings by using a traditional interface, while Google and Amazon's voice assistants allow for deleting voice recordings via voice commands [51]. Text-based systems which are frequently deployed on websites to assist users are often missing privacy controls altogether and make it laborious and complicated for users to exert their rights as they would need to contact the data controllers manually. Therefore, an increasing stream of research argues for conversational privacy and expressing privacy-related information in dialogue form [14, 22, 44].

When designing for privacy, it is important to consider that privacy decision-making can be subject to heuristics and biases [2]. Thereby, individuals' choices may not be aligned with their attitudes – a discrepancy that can lead to frustration and regret [2, 31]. One possible explanation for why people do not act according to their attitudes and values is based on the dual-process theory of cognition [26]. The theory describes humans' decision-making process by two systems – fast and intuitive thinking (System 1 thinking) and slow, effortful and analytical thinking (System 2 thinking). It states that System 2 is triggered only occasionally as it usually accepts impressions generated by System 1 and turns them into actions. Thus, most of people's actions originate from fast and intuitive thinking. While this allows us to effectively accomplish daily life decisions, decisions based on fast and intuitive thinking may be more easily biased and not aligned with peoples' attitudes. System 2 - and thus, analytical thinking - is triggered in states of cognitive strain, surprise or doubt [26]. Debiasing strategies that promote transitioning towards System 2 can support people in making better privacy decisions. While debiasing strategies have been investigated in previous privacy-related studies on mobile applications and social media [6, 57], they are yet to be studied for Conversational AI.

Given previous research on debiasing and conversational privacy, our study was guided by the following questions:

- How can we design privacy controls for CAI that aim for transparency and at debiasing peoples' decision-making?

*The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg and Fraunhofer IIS.

EuroUSEC 2023, October 16–17, 2023, Copenhagen, Denmark

ACM ISBN 979-8-4007-0814-5/23/10.
https://doi.org/10.1145/3617072.3617106

- What are these privacy controls' influences on peoples' behaviour and perception?

Based on the idea of conversational privacy and debiasing, we designed privacy strategies for CAI to support users in their privacy decision-making. In particular, we focus on privacy controls that allow users to delete their data. While deletion covers only one aspect of privacy controls in CAI, our design considerations and evaluation metrics can inform the design of privacy controls in CAI more generally. Moreover, peoples' right to erasure has gained importance and has been emphasized by jurisdictions worldwide [20, 38]. We present our design considerations for privacy strategies for CAI in Section 2. Based on those, we implemented and tested our strategies in a crowdsourced chatbot experiment as described in Section 3. Finally, we present and discuss our results in Section 4 and Section 5.

2 DESIGN CONSIDERATIONS FOR PRIVACY STRATEGIES IN CAI

2.1 Conversational Privacy

Current CAI systems require users to share personal information but lack easily available and accessible protective mechanisms [7, 31]. The data minimization principle of the GDPR suggests that data collectors should only collect, process and store information that is necessary for a service. Likewise, other jurisdictions have adopted the data minimization principles, e.g., the California Privacy Rights Act (CPRA) – an amendment of the California Consumer Privacy Act (CCPA) [38, 39]. But even if only necessary data is collected, processed and stored, the process might not be transparent to the users. This is why international privacy laws and guidelines, require transparency. The GDPR as well as the CCPA incorporate a principle of transparency by requiring businesses to share information with consumers about their data collection and sharing practices [20, 38]. Similarly, the Personal Information Protection Law (PIPL) in China emphasizes transparency to promote accountability and responsibility of data handling by data controllers and give individuals greater control over their personal information [40]. Moreover, global non-profit organizations such as the Open Voice Network inherit transparency as a key principle for future voice assistance and Conversational AI [36].

Conversational privacy can be a means of providing transparency in interactions with CAI. Conversational privacy in the context of CAI refers to expressing privacy-related information in dialogue form [22]. Thereby, conversational privacy can deliver notice by expressing privacy policies as well as choice by allowing users to change their privacy settings in natural language [22]. In this study, we will focus on the choice principle, i.e., providing users with an option to exhibit control over their previously shared data. Given previous research on conversational privacy, we have reason to believe that small changes in dialogue can have significant effects on user behaviour and perceptions [14]. Thus, dialogue strategies should not be neglected for privacy design as they can enhance the transparency of CAI systems and lead to increased user trust.

2.2 Timing

We make use of conversational privacy to allow users to control their personal information. Privacy settings offer a range of choices for managing personal information, e.g., controlling the data shared with applications or the usage of one's voice recordings for improving voice assistance. Yet, in this study, we focus on the privacy setting that enables the deletion of data. Regulations worldwide have emphasized the importance of giving users the right to have their data deleted at any time. One of the key user rights outlined in the GDPR is the right of individuals to have their information erased [20]. In the US, the California Consumer Privacy Act (CCPA) and the Virginia Consumer Data Protection Act (CDPA) give users the right to have their data deleted [38, 41]. Similarly, the Personal Information Protection Law (PIPL) in China gives data subjects the right to erasure under certain conditions [40]. While companies like Amazon or Google recently started to let users delete their recordings via voice commands [51], they require users to actively engage in privacy-preserving behaviour. However, given previous research on privacy decision-making and privacy controls in CAI, users are unlikely to act according to their values and use the offered privacy self-management controls [2, 29]. Instead of providing reactive privacy controls, users' decision-making can be supported by proactive control options. Thereby, the timing of privacy controls is a key factor to ensure usability and alignment with user's needs [47]. Various points in time can be used to present privacy controls, e.g., "at setup", "just-in-time", "context-dependent", "periodic", "persistent" and "on demand" [47]. In our experimental setup, we provide two different services to users, e.g., the possibility to order pizza and to check the credit card balance. We assume that users interact and engage in data sharing with the service for the first time in our experiment. Moreover, the service is designed to be publicly accessible and not permanently installed on users' devices as this would ask for a privacy setup interface on initial use. Instead, we focus on just-in-time strategies that allow users to delete their data after finishing transactions with the service.

2.3 Debiasing

Even if users are presented with proactive privacy controls, their choices might be subject to heuristics and biases. Their decisions might be based on intuitive thinking rather than on an analytical benefit-risk assessment. As situations of privacy decision-making are often characterized by uncertainty due to information asymmetry or difficulties in predicting future outcomes, people are likely to rely on mental shortcuts and make biased choices that are not aligned with their intentions [3]. Some of the heuristics and biases that have been identified in privacy decision-making include the availability heuristic, representativeness heuristic and optimism bias and overconfidence [2]. Depending on their implementation, proactive control options utilizing natural language can either reinforce or mitigate the effects of biases, e.g. deceptive patterns on "trick wording" [11]. By incorporating the principle of debiasing, we aim at designing privacy strategies that can reduce the impact of biases on peoples' decision-making.

Debiasing strategies have been successfully applied in various research fields, e.g., medical field, economics, nutrition and health.

Based on the dual-process theory, these strategies can aim at triggering a more rational thinking process by producing competing intuitions or making people reflect on their choices [26]. For example, in the medical field, it has been demonstrated that diagnostic errors could be significantly reduced when practitioners were prompted to reflect alternative diagnoses or reconsider their initial diagnoses [28]. In addition, diagnostic time-outs can stimulate the reflection on previous choices. Similarly, a study on overreliance on AI introduced a waiting time before participants were presented with an AI answer [15]. They showed that when slowing down decision-making, performance improved in cases where the AI prediction was incorrect. In the privacy research field, debiasing strategies have been applied for traditional user interfaces, app development, social networks and mobile applications [2, 6, 8, 18, 56]. In mobile applications, privacy-relevant information can be displayed, e.g., how often a phone's location was shared with different apps, to nudge users into changing their apps' privacy settings [6]. The authors found their strategies to be effective with 95% of participants reviewing their permissions and 58% changing the corresponding settings [6]. Regarding social media, three different privacy strategies have been investigated to encourage real-time adjustments when posting content on Facebook [57]. Again, one of the strategies would delay the post and allow users to reflect and possibly cancel their actions. The study showed that delaying the post was perceived positively as it provides the chance to correct typos, post better quality content or cancel unnecessary posts [57]. Given previous research on debiasing strategies, we believe that privacy strategies for CAI can apply similar techniques and effectively support users in making better judgements.

3 EXPERIMENT

3.1 CAI System and Scenarios

We use a text-based CAI system to test our strategies. We will refer to the implementation as chatbots which use natural language to interact with a human via text [45]. In particular, we use Chatbot Language (CBL) [45] to implement and test our strategies on Amazon Mechanical Turk (Mturk).

We investigate two chatbot scenarios, a banking chatbot asking for permission to access users' credit card information and a location chatbot asking for permission to access users' location. Because information sensitivity can impact people's perception of privacy, the two scenarios were chosen to differ in their sensitivity. When analyzing information sensitivity across nations, it was found that financial account numbers and credit card numbers are perceived as highly sensitive in Germany and the US while GPS location data and home address were perceived as medium sensitive with only a little difference between the German and the US population [48]. A similar trend regarding information sensitivity was perceived in a study on smart home personal assistants [1]. In a previous study on conversational privacy, a similar banking chatbot has been tested [14]. In this study, participants were presented with an artificial credit card number and they were asked to check the corresponding credit card balance [14]. In contrast, we want users to believe that we can access their real personal data as this is closer to a real-life scenario and might significantly impact behaviour and perceptions. Therefore, in the banking scenario, we ask users to allow access to their credit card information assumed to be stored in a cache. In the location scenario, we request access to their current location. Moreover, the two scenarios were designed to ask only for information that is required to fulfil the task and thus respect the legal principle of data minimization and follow current best practices of privacy design [20, 31].

3.2 Experimental Conditions

After granting or denying access to their data, participants were exposed to the control condition or to one of the three privacy strategies (the conditions are displayed together with their response options in Table 1). By designing our conditions, we focused on providing the users with the choice of deleting their data as emphasized by jurisdictions worldwide. The corresponding dialogue trees for the banking and location chatbot followed by the four conditions are provided in Appendix B. While our original study included a control condition (asking "Is there anything else I can help you with?") unrelated to data-sharing practices and meant to serve as an additional baseline. However, it does not significantly contribute to the analysis of this study as it showed a 50-50 "yes-no" decision ratio and participant's perceptions did not vary significantly. Therefore, we only report on the control condition that is specific to the behaviour we want to investigate, i.e., storage or deletion of data. By asking users "I will save your data for future interactions now, okay?", we give them the opportunity to actively control their privacy while at the same time nudging them into disclosing behaviour. In interface design, "dark patterns" describe similar strategies [9]. For example, cookie banners are often designed such that individuals make decisions that are beneficial for the data collectors rather than for themselves [9]. We expect users who respond with "yes" to anticipate their data to be saved and users who respond "no" to assume data deletion. Thus, we intend the control condition, i.e., dark pattern, to serve as a baseline for storage and deletion requests of data.

We implement three different privacy strategies based on the concept of debiasing in order to disrupt heuristic reasoning. They are designed to make users engage in System 2 thinking and support the process of rational cost-benefit analysis. Based on the control condition, we implement a slow-down condition. Drawing from previous studies, people are given 20 seconds time to think about their response before it is sent to the system [57]. Importantly, the interaction can not be terminated earlier but participants have to wait until their answer is processed. Similarly to delay conditions in previous studies, our slow-down condition is designed to trigger slow thinking and give users the time to reflect and possibly reconsider their decision [15, 57].

Our second strategy is based on debiasing strategies applied in the medical context where participants are asked to consider alternatives instead of deciding intuitively [28]. In our CAI privacy context, we present users with two alternatives, "delete data from the interaction" and "save data for future interactions". Importantly, users need to explicitly state their decision as they are only allowed to proceed once the words "delete" or "save" were recognized by the system. Implementing active choice in complex situations, e.g., when choosing from hundred of retirement funds, has been subject to criticism [53]. However, choices in our scenario are simple and

Condition	Question	User Response Options	Meaning of User Responses
Control	I will save your data for future interactions now, okay?	Yes/No	Save/Delete
Slow Down	I will save your data for future interactions now, okay? I'll give you 20 seconds to think about it.	Yes/No	Save/Delete
Alternative	Do you want me to delete your data from this interaction or have it saved for future interactions?	Save/Delete	Save/Delete
Deletion	Do you want me to delete your data from this interaction now?	Yes/No	Delete/Save

Table 1: Questions, user response options and their meaning for four conditions, including a control condition and three privacy strategies. The corresponding dialogue trees can be found in Appendix B

thus, an active choice can be a suitable strategy. Moreover, by requiring active choice and displaying both options to people, we aim at stimulating competing intuitions, making people think about the risks and benefits of deletion and storage and motivating decisions that better reflect people's attitudes.

Our last privacy strategy provides the option to delete data and allows people to reconsider their disclosure to the chatbot. Similarly to the other privacy strategies, this option aims at making people rethink their data sharing with the chatbot. Moreover, it might come as a surprise to participants as it is not frequently applied in real-life scenarios. Based on the dual-process theory, surprise is likely to trigger System 2 activation, i.e., analytical thinking, and can thus lead to better privacy judgements. In the medical field, similar strategies that ask practitioners to reconsider previously made diagnoses have shown improved accuracy on erroneous decisions [28]. Moreover, the option to delete data was successfully applied in a previous study on conversational privacy and led to an increase in perceived privacy [14].

For our experiment, we adopted a between-subject design and assigned conditions randomly to participants, while excluding those who had participated in previous chatbot experiments to eliminate potential confounding factors. We utilized CBL to inform participants about data protection regulations, provide a task description, and administer a post-experiment survey on chatbot interaction.

3.3 Measurements and Survey Design

3.3.1 Perceived Realism. The results of this study are only generalizable to real-life scenarios if people believed that the chatbot had access to their personal information. While we designed the dialogue with that in mind, we additionally assessed the level of perceived realism in the questionnaire. This allows us to validate whether users perceived the chatbot scenario to be real. Therefore, we adapted Cho et al.'s scale on perceived realism to fit the context of CAI [17].

3.3.2 Privacy Perceptions and Usability. To measure privacy perceptions, we rely on scales that have been used before in privacy research and more specifically in a study on conversational privacy where they showed satisfying reliability and validity results [14]. We assess usability via the System Usability Scale (SUS) which is a quick, reliable and valid scale frequently used in HCI research [12]. While SUS has been criticized for assessing usability aspects within Conversational AI, e.g., chatbots or voice assistants [10, 60], other studies found SUS to be reliable and valid in the CAI context [13, 19, 27]. Moreover, usability scales specifically designed for CAI such as the BOT Usability Scale (BUS) [10] or Voice Usability Scale (VUS) [60] include constructs, e.g., usability in noisy environments or perceived privacy and security, that are irrelevant for our study or already assessed by a dedicated privacy scale. Therefore, in this study, we will rely on SUS to assess the usability of our CAI system. SUS takes on scores between 0 and 100 with higher scores indicating better usability.

3.3.3 Control Variables. Lastly, we measure several control variables which could have an impact on our results. All control variables are assessed at the end of the survey before asking for demographic information to ensure that no priming of participants for privacy occurs. While all control variables were originally measured on a 7-point Likert Scale, we use a 5-point Likert Scale to ensure that constructs included for control are easily comparable to the constructs we are mainly interested in, i.e., realism, privacy and usability. Moreover, as we conduct crowdsourcing tests we are keen on keeping the questionnaire clear and simple. It has been shown that a 5-point Likert Scale is less confusing to be interpreted and easy to use for respondents while at the same time being sufficient for participants to express their views and perceptions [23, 34, 37]. Additionally, we tested the respective scales in a pilot study and found that they are reliable and valid.

First, we investigate trust in the chatbot. Trust can influence users' willingness to disclose personal information to a chatbot [5, 43]. We adopt a previously used trust scale measuring trust in online companies to fit the chatbot context [33].

Second, we include a measure of privacy concerns. In contrast to the control variable trust, we aim to measure privacy concerns as a trait-like characteristic rather than specific to the chatbot interaction. Privacy concerns are known to influence users' willingness to disclose personal information in varying contexts such as e-commerce, mobile applications and voice assistants [5, 25, 33]. Privacy concerns are frequently measured using the Internet Users' information Privacy Concerns (IUIPC) scale which has been recently reevaluated using a confirmatory factor analysis [21, 33]. The results suggest reducing the original 10-item scale to an 8-item scale as it showed improved construct validity and reliability [21]. Therefore, we use the newly validated 8-item IUIPC scale to measure privacy concerns as a control variable.

Third, privacy literacy could provide knowledge and skills to understand chatbot behaviour, e.g., the impact of having data stored. Privacy literate users may be less influenced by debiasing as they

are already well aware of possible privacy violations and do not have to rely on privacy strategies to activate rational assessment. Privacy literacy can be measured objectively or subjectively by relying on participants' self-assessment [35]. Even though a self-assessment might be less accurate, we measure privacy literacy subjectively as it has the advantage of not being too exhaustive or prolonging surveys unnecessarily [35].

Lastly, participants' behaviour and perceptions might be affected by cultural differences such as uncertainty avoidance. It was found that for people from cultures with high uncertainty avoidance, privacy risks are more important than for people from cultures with low uncertainty avoidance [54]. As our privacy strategies are supposed to make people rationally weigh benefits and risks, uncertainty avoidance can impact their risk perception and thus their decision-making. Moreover, participants who report high uncertainty avoidance might react to debiasing stronger which could impact peoples' perception of usability of the chatbot. We rely on Hofstede's metric to measure uncertainty avoidance as it has been heavily used in social sciences, cross-cultural studies as well as in studies concerning international business and consumer behavior [54, 59]. In particular, we use Hofstede's five dimensions of cultural values measured on an individual level to measure uncertainty avoidance [59]. We assess participants' individual levels of uncertainty avoidance rather than relying on the known overall level of uncertainty avoidance for different nationalities as we are interested in controlling for an individual's cultural orientation. While MTurk allows restricting samples to a specific region, e.g., the US or Germany, the cultural background might still be distinct from people's current location and thus their level of uncertainty avoidance might vary. Finally, we did not restrict the participant's base to a certain country.

In addition to the items assessing peoples' perceptions and attitudes, we include three screening questions in our survey to check the reliability of submitted responses [24, 30]. Our three screening questions are positioned between the other survey items in no particular order and we exclude participants who did not pass any of the attention checks from further analysis.

3.4 Ethical Considerations

The following outlines the measures we implemented to ensure ethical treatment of the participants. We recruited participants from Amazon Mechanical Turk and paid $2 for their participation. This calculates to an average hourly wage of $17 for the banking scenario and $20 for the location scenario, as participants in this scenario took less time. Our task description clearly mentioned that participants were going to interact with a chatbot, their specific task, i.e., to order pizza or check their credit card balance, and that they will be asked personal questions (see Figure 5 in the Appendix for the task description of the location scenario). We did not tell participants prior to the interaction that we evaluated data-saving practices as this might have changed their behaviour and perceptions. To ensure ethical treatment, participants were free to what extent they responded truthfully. Moreover, we followed current best practices of privacy design and asked only for information that was necessary to provide service to participants [20, 31].

We chose to make users believe that we could access their data to allow generalization towards real-life scenarios (see Appendix B for the detailed dialogue trees). However, at no time was our system able to access any personal data other than the text users shared during the interaction. Alternative experimental designs were considered but ruled out because of major limitations. For example, in a previous study, participants were given an artificial credit card number to check their corresponding balance[14]. Yet, as users were not asked to enter any personal data, interpretability and generalizability towards real-life scenarios were limited. Again we ensured ethical treatment by debriefing participants and fully disclosing our practices after the study. We specifically emphasized that no personal data was accessed if not entered during the chatbot interaction.

4 RESULTS

In Table 2, we present the experimental and demographic data. We excluded from our analysis any participants who failed one of three screening questions. We guaranteed that each group had more than 50 accepted participants based on power analysis results from a pilot study. The participants' disclosure behavior was comparable across scenarios, with over 80% allowing access to their personal information, which is crucial for our conditions, as they depend on users sharing information initially.

Moreover, we investigate whether participants perceived the chatbot scenario to be real based on the perceived realism scale. Our analysis is based on all participants who passed the screening tests. Figure 1 shows aggregated realism ratings for each condition and scenario. We find that participants perceived the scenario as sufficiently real with a mean rating of 3.4 (out of 5) for all conditions and scenarios and a standard deviation of 0.4. Moreover, there were no significant differences between scenarios or conditions. Therefore, we believe that our results are valid and generalizable to real-life scenarios.

4.1 Analysis of Participants' Behavior

To investigate differences in participants' behaviour and perceptions, we analyze only data from subjects who provided access to their personal information, i.e., who answered "yes" to the questions asking for granting access (see Appendix B for the detailed dialogue trees). Yet, we did not find significant differences between people's behaviour or perceptions depending on whether they had granted access or not. We compare participants' intention to delete across scenarios and conditions (see Figure 2). We can see that a majority of people exposed to the control or slow-down condition agreed to have their data saved. In the banking scenario, 57% of the people exposed to the alternative condition wanted their data to be saved while 43% wanted their data to be deleted. However, in the location scenario, 70% wanted their data to be saved and only 30% asked for deletion. For the option to delete data, behaviour does not vary largely between scenarios, with 73% and 68% asking for deletion in the banking and location scenario. This suggests that two of our privacy strategies, i.e., the alternative and deletion condition, affect user behaviour compared to the control condition while the slow-down condition does not. Although we see variations between scenarios, there are no significant differences in

Demographic and experimental data	Banking	Location
# conditions	4	4
# participants	248	257
# excluded participants	24	43
# accepted participants in the different conditions (Control/ Slow Down/ Alternative/ Deletion)	56/51/56/61	55/53/51/55
# accepted participants' disclosure behaviour (Granting Access/ Denying Access)	180/44 (80%/20%)	193/21 (90%/10%)
# accepted participants in the different conditions who granted access (Control/ Slow Down/ Alternative/ Deletion)	44/39/49/48	50/49/47/47
Mean (SD) age of workers in years	34 (10)	35 (10)
# Gender (female/male/diverse/not provided)	129/95/0/0	92/122/0/0
# Native English speakers (yes/no)	220/4	211/3
# Usage (weekly/monthly/less than once a month/never)	39/82/72/31	46/59/68/41

Table 2: Summary of demographic and experimental data for the banking and location scenario.

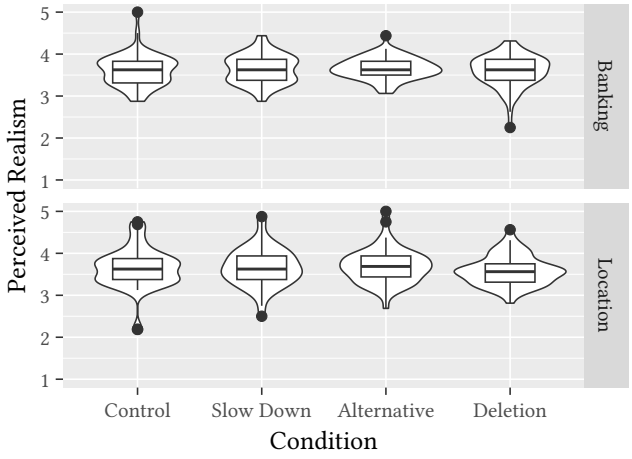

Figure 1: We show violin plots with boxplot overlay for realism ratings for all participants who passed the screening test assessed on a 5-Point Likert-Scale (1 = strongly disagree, 2 = disagree, 3 = neither agree nor disagree, 4 = agree, 5 = strongly agree).

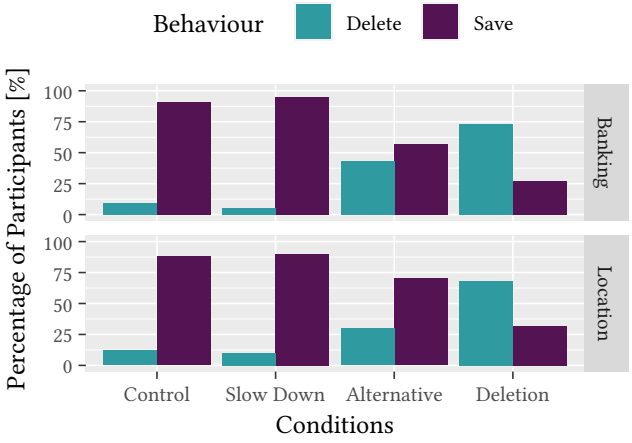

Figure 2: Response behaviour of participants across conditions who granted access to their personal information. Only conditions are displayed where participants could decide whether they wanted to save or delete their data (see Table 1 for user response options).

data storage decisions between scenarios for the alternative and deletion condition. Thus, reconsidering data sharing and active choice between alternatives show robustness across scenarios.

We fit a binary logistic regression model including condition as a predictor and compare it to a model including two predictors, i.e., scenario and condition, a model including all control variables and a null model. The results indicate that only condition is a significant predictor ($\chi^2(3) = 87.38, p < 0.001$) while the control variables do not show significant effects. This shows that condition is the main source of influence on people's intention to have their data deleted and can outweigh the main effects of trait-like specifics and contextual factors like information sensitivity. Moreover, we compare Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) scores for the different models (see Table 3) [4, 49]. The model including only condition resulted in the lowest AIC and BIC score and was found to carry 70% of cumulative Akaike weight or respectively 94% of cumulative BIC weight. The estimates of the final model are presented in Table 4. In fact, we find that the odds of someone asking for deletion of their data is 4.8 times higher when exposed to the alternative condition and 21.2 times higher when provided with an offer to deletion compared to the control condition. Therefore, the privacy strategies presenting an alternative or prompting reconsideration significantly influence people's intention to have their data deleted while the additional time delay does not when compared to the control condition. As behavioural patterns are similar for the control and the slow-down condition, we expect participants to not have used the additional time to reflect on their choices. Moreover, we can identify that the offer to delete data slightly nudges participants into deletion of their data while the alternative condition does not. While the nudging effect is not as strong as the one in the control condition, it is still robust across scenarios.

4.2 Analysis of Participants' Perceptions

To analyse people's perceptions, we investigate differences in their ratings on privacy perceptions and usability. Their perceptions were

Model	K	Model LL	AICc	AICc Weights	Cum. Akaike Weight	BIC	BIC Weights	Cum. BIC Weight
Condition	4	-175.97	360.05	0.70	0.70	375.63	0.94	0.94
Condition & Scenario	5	-175.83	361.82	0.29	0.99	381.27	0.06	1.00
Control	12	-172.03	368.92	0.01	1.00	415.11	0.00	1.00
Null	1	-234.30	470.62	0.00	1.00	474.53	0.00	1.00

Table 3: Model Selection based on AIC and BIC.

Parameter	Estimate	Std. Err.	z value	Pr(> |z|)
(Intercept)	−2.13	0.33	−6.36	< 0.001
Condition				
Slow Down	−0.32	0.52	−0.62	0.54
Alternative	1.57	0.40	3.97	< 0.001
Deletion	3.05	0.40	7.54	< 0.001

Table 4: Binary Logistic Regression Analysis of Variables affecting participants' storage and deletion behaviour.

captured by a questionnaire presented after the chatbot interaction. For aggregation, we followed guidelines as reported for the individual scales [12, 14]. Before analyzing the results in detail, a note on the reliability and validity of the two scales based on confirmatory factor analysis (CFA). Both scales show overall sufficient reliability scores but lack convergent validity due to a low average variance extracted (AVE). While the model shows a satisfactory fit for the privacy perception scale, it poorly fits for the usability scale.

Results of the averaged ratings for each condition and scenario are provided in Figure 3 and 4. While the privacy perception ratings are generally high (mean over all conditions and scenarios is 3.8 with a standard deviation of 0.6), usability ratings are overall low (mean over all conditions and scenarios is 53.8 with a standard deviation of 11.2). We used ordinal logistic regression to investigate differences between scenarios and conditions. However, we did not find significant differences either in ratings on privacy perception or usability.

We compare our usability ratings to a curved grading scale established from 241 studies [46] and find that they fall into the second lowest percentile range. This means that the CAI system scored better than 15% to 34% of systems in the norm group. One explanation for the overall low usability ratings could be that our dialogues had a negative ending (e.g. no service due to technical difficulties in the banking scenario or closure of the restaurant in the location scenario). The negative ending was chosen to avoid further deceptive practices and users' uncertainty when providing a fake credit card balance or a real delivery appointment. However, participants reported frustrating experiences with the CAI system. Moreover, we assessed frustration by a single item in the survey and found that frustration showed a slight increase with a mean rating of 3.0 over all conditions and scenarios. Therefore, the assessed usability ratings are likely a result of the overall CAI system experience and can not provide insights into the perception of the individual privacy strategies. On the other hand, in our experiment, the SUS scale showed a poor model fit. As discussed in Section 3.3.2, SUS has been criticized for assessing usability aspects within CAI and newly developed usability scales such as BOT and VUS [10, 60] that take the nature of conversational interactions into consideration, might better capture small differences between conditions and scenario.

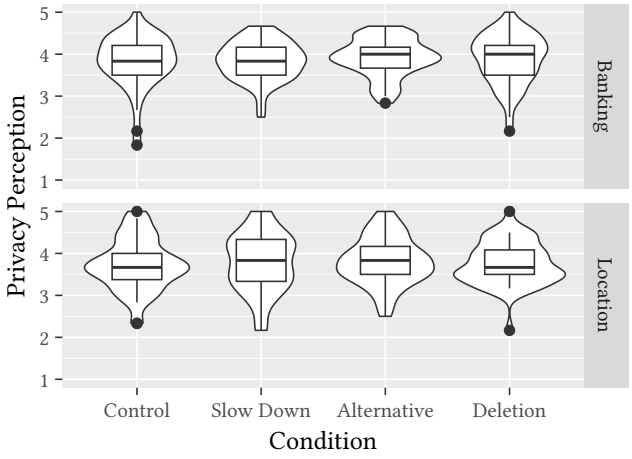

Figure 3: We show violin plots with boxplot overlay for privacy perceptions for all participants who granted access to their data assessed on a 5-Point Likert-Scale (1 = strongly disagree, 2 = disagree, 3 = neither agree nor disagree, 4 = agree, 5 = strongly agree).

4.3 Analysis of Participants' Attitude-Behaviour Alignment

Our privacy strategies are based on the concept of debiasing and supporting people in overcoming their cognitive biases. Thus, they aim at promoting decision-making that is consistent with peoples' attitudes. Previous research suggests investigating the alignment of attitudes and behaviour to evaluate debiasing strategies in the privacy context [2]. Before evaluating whether participants' behaviour matched their attitudes in the individual conditions, we investigate the reliability and validity of the scales. In particular, we use CFA to analyse the trust, privacy concern, privacy literacy and uncertainty avoidance scale. We find that the trust scale shows satisfactory reliability and model fit but lacks convergent validity with an AVE < 0.5. While we cannot confirm the three-dimensionality of the IUIPC scale based on our data collection, when treated as one factor to assess peoples' privacy concerns, the scale shows satisfactory reliability and model fit. Yet, again convergent validity could not be established. The CFA for the privacy literacy scale shows a good model fit, but reliability and convergent validity scores do not exceed the commonly considered cut-off values of 0.7 and 0.5.

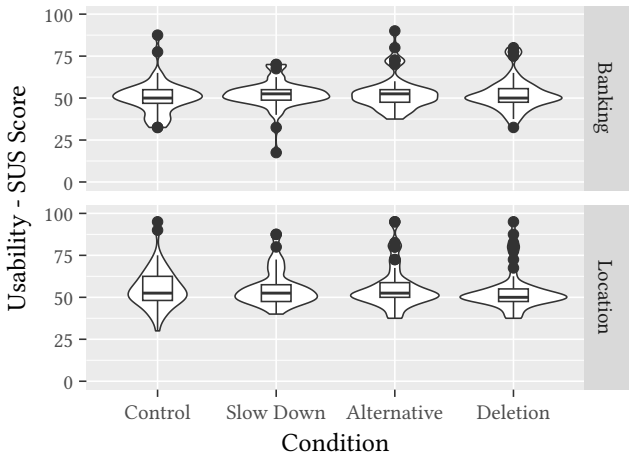

Figure 4: We show violin plots with boxplot overlay for usability ratings (SUS scores) for participants who granted access to their data.

In addition, we want to point out that participants' scored high on the privacy literacy scale with an average score of 3.9 (standard deviation of 0.59) over all conditions and scenarios. This could be due to the fact that we did run a crowdsourced study where participants might be more knowledgeable on online environments compared to non-crowdworkers. Additionally, self-assessment of privacy literacy might be less accurate than objective assessment and participants might be likely to overestimate their knowledge. Lastly, the CFA for the uncertainty avoidance scale shows a satisfactory model fit but scores low on reliability and convergent validity.

To analyse participants' attitude-behaviour alignment, we divide the dataset into four corresponding subsets, i.e., one for each condition, to investigate attitude-behaviour alignment for the individual conditions (see Table 2 for the group sizes). For each subset, we fit three different models and compare them against each other. One model is purely based on the type of scenario and our control variables, i.e., trust in the chatbot, privacy concerns, privacy literacy and uncertainty avoidance. At the same time, another includes demographic variables and frequency of usage of chatbots in addition to the scenario and control variables. Lastly, we compare both of them to a null model.

We analyse AIC and BIC scores to compare the three models. We find that the null model fits best for participants exposed to the three privacy strategies, i.e., the slow-down condition, the alternative condition and the offer to delete their data. This shows that participants' behaviour could not be traced back to one of our assessed attitudes. While participants' behaviour varied significantly between privacy strategies, we cannot conclude on which factors their decisions were based on. In particular, participants exposed to the alternative conditions showed varying answer behaviour, yet, it remains unclear whether they engaged in a more analytical decision-making process.

For participants exposed to the control condition, the null model shows the lowest BIC score, while the model including all control and demographic variables results in the lowest AIC score. Here, trust in the chatbot, frequency of usage of chatbots and gender show a significant effect on peoples' behaviour when exposed to the control condition ($\chi^2_{\text{trust}}(1) = 8.75, p < 0.01, \chi^2_{\text{usage}}(3) = 12.84, p < 0.01, \chi^2_{\text{gender}}(1) = 4.43, p < 0.05$). Yet, AIC is known to have a tendency to include too many variables and to less penalize complex models [55]. While differences in model complexity can explain contradicting AIC and BIC scores, the selection of an appropriate model selection metric can be guided by the underlying goal of model selection, e.g., explanatory vs. predictive modeling [50]. AIC is considered a popular predictive metric while BIC is consistent and might be better suited for explanatory model selection [50]. However, by analyzing attitude-behaviour alignment for our strategies, we aim at identifying underlying associations between attitudes and behaviour and apply a descriptive approach rather than purely explanatory or predictive modelling.

While our privacy strategies were supposed to activate System 2 thinking and should support participants in their decision-making, we could not show that they lead to an improved attitude-behaviour alignment. We found that our privacy strategies significantly influence peoples' behaviour, but a relationship between behaviour and the assessed attitudes could not be established.

5 DISCUSSION AND FUTURE WORK

Our study contributes to the area of usable privacy controls for Conversational AI. First, we highlight important design elements for privacy approaches in CAI, drawing on both international privacy regulations and research from the fields of behavioural economics, medicine and usable privacy and security. In particular, we incorporate principles of transparency, control and debiasing into the design of our privacy strategies. Second, we evaluate the influence of conversational privacy strategies on peoples' behaviour and perceptions. In addition to peoples' perceived perception of privacy and usability, we investigate whether our privacy strategies can support people in overcoming their biases and engaging in more rational decision-making. Therefore, we study the effect of the different strategies on peoples' attitude-behaviour alignment.

Our findings indicate that the language used in our strategies significantly affects users' behaviour. Thereby, our strategies show a robust effect across scenarios which were designed to differ in their information sensitivity. In addition to outweighing contextual factors, the tested strategies show robustness across trait-like specifics such as peoples' privacy concerns. This shows that variations in dialogue can have significant effects on users' behaviour and that the influence of conversational strategies should not be neglected for usable privacy design in CAI. Particularly, displaying alternative choices and providing users with an offer to delete their data resulted in varying behaviour. Instead, slowing down peoples' decision-making did not show a significant effect on their behaviour. Moreover, privacy strategies that rely on additional time delays might be implemented less frequently by conversational designers as they might be perceived as a disturbance to the natural flow of a conversation.

While previous research reported a positive influence of conversational privacy controls on peoples' privacy perceptions [14], in our study, we did not find significant differences in peoples'

perceptions. While the effectiveness of debiasing strategies in the medical field is assessed by the evaluation of diagnostic error rates, evaluation in the privacy context is more complex. As privacy decision-making is highly subjective, the desired behavioural distribution remains unknown. The effectiveness of privacy controls cannot be determined solely by striving for a balanced distribution of data storage and data deletion requests or by seeking a predominant number of data deletion requests. Therefore, one proposed way of evaluation asks for aligning behaviour and attitudes. Yet, we could not show that participants' behaviour was aligned with the assessed attitudes when being exposed to the privacy strategies. We assessed only a limited set of privacy attitudes that were known to influence peoples' privacy decision-making and investigated whether our strategies reduced the discrepancy between people's attitudes and behaviour. However, different attitude factors such as perceived risks, benefits or regret after decision-making might provide better insights into peoples' behaviour when exposed to privacy strategies in Conversational AI. A different approach suggests the evaluation of uncertainty as an indicator for System 2 activity and could, therefore, be used to evaluate privacy controls that integrate the concept of debiasing [32]. To further analyse the impact of privacy controls on peoples' behaviour, future research could use a qualitative approach to get insights into participants' thinking processes and identify relevant factors for quantitative evaluation. Moreover, evaluation guidelines to assess the effectiveness and usability of conversational privacy strategies need to be established. This would allow comparability among studies and could pave the way towards reliable and valid standards for privacy controls in CAI.

As more privacy regulations emphasize the importance of transparency and give users greater control over their personal data, designing privacy controls for CAI becomes an increasingly pressing issue. In this study, we focused on privacy controls that allow users to have their data deleted and exert the "right to be forgotten" as emphasized by data protection regulations worldwide [20, 38]. Nevertheless, our design considerations can be more generally applied to privacy controls for CAI systems. Privacy strategies must not only be easily accessible to users but they must also be timed appropriately to meet users' needs and must provide support to users in their privacy decision-making process. Conversational privacy strategies can serve as a means to achieve this. Moreover, debiasing strategies adapted to CAI systems have the potential to support users in making better judgements not only for deleting their data but more broadly for controlling their personal information. Therefore, future research could focus on applying these design considerations to develop privacy controls for CAI and evaluate their impact on people's decision-making.

Moreover, we evaluated our strategies using a text-based CAI system rather than a voice-based system. Previous studies have explored differences in modality, i.e., voice vs. text, with varying outcomes depending on the researched scenarios and evaluation measures, e.g. social presence perception or risk perception [16, 52]. While designing for voice-based CAI systems can be different than for text-based systems, for our privacy strategies, we have not relied on any text-based specific designs such as graphics, lists or links. Therefore, even if investigating only text-based CAI systems, we expect our results to be transferable to speech-based CAI systems due to the shortness and simplicity of the privacy strategies. Nevertheless, future work could investigate the impact of voice-specific characteristics such as pitch, rate or pause on peoples' privacy decision-making.

While our study sheds light on how privacy strategies can be designed lawfully, so far little is known about deceptive and non-deceptive design practices for conversational systems [42]. Future research could therefore focus on creating and evaluating conversational privacy design patterns for CAI systems, similar to dark and bright patterns for user interface design. Moreover, to ensure that privacy controls are implemented effectively and consistently across industries and applications, legal classification for conversational privacy strategies is needed. Such a classification could provide a clear framework for privacy practitioners, conversational designers and regulators to follow, ensuring that conversational privacy controls meet minimum standards and sufficiently protect users' rights. Establishing best practices and guidelines for conversational privacy strategies, could support developers and conversational designers in their work and improve the privacy of CAI systems.

Our study does not come without limitations. As our experiment was conducted in English, our findings are language dependent. While a majority of our participants reported English as their native language, they might have come from different cultural contexts and countries. We assessed uncertainty avoidance as a cultural factor, however, other cultural differences can impact peoples' privacy decision-making. In particular, debiasing strategies might be perceived differently across cultures. Therefore, future research could investigate conversational privacy controls and their influence on reactions and perceptions of other languages and cultural contexts.

We ran our study on Amazon Mechanical Turk which provides easy access to a diverse set of participants, but might not resemble the general population. For example, we found that participants scored generally high on privacy literacy – a factor that could be specific to people on MTurk and their experience with online environments and associated privacy threats [58]. Hence, a potential follow-up study could repeat the research with individuals who are less experienced in online environments and compare their reactions and perceptions towards conversational privacy strategies.

6 CONCLUSION

We investigated the effect of conversational privacy controls on peoples' behaviour and perceptions when disclosing personal information to a chatbot. Our privacy strategies were designed to support people to exert their right to erasure after having disclosed personal information to the chatbot while at the same time promoting rational decision-making. Our results show that confronting people with alternatives of either having their data saved or deleted or a simple offer to delete their data significantly changes people's behaviour. Yet, we did not find any influence of privacy strategies on peoples' privacy perception or usability. We investigated whether peoples' attitudes were aligned with their behaviour when being exposed to privacy strategies but could not establish any relationship. Therefore, future research could establish guidelines on

how to evaluate the effectiveness of conversational privacy strategies and investigate the influence of modality on people's privacy decision-making.

ACKNOWLEDGMENTS

Our work is partially funded by the German Federal Ministry for Economic Affairs and Energy as part of their AI innovation initiative (funding code 01MK20011A).

REFERENCES

[1] Noura Abdi, Xiao Zhan, Kopo M. Ramokapane, and Jose Such. 2021. Privacy Norms for Smart Home Personal Assistants. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3411764.3445122

[2] Alessandro Acquisti, Idris Adjerid, Rebecca Balebako, Laura Brandimarte, Lorrie Faith Cranor, Saranga Komanduri, Pedro Giovanni Leon, Norman Sadeh, Florian Schaub, Manya Sleeper, Yang Wang, and Shomir Wilson. 2018. Nudges for Privacy and Security: Understanding and Assisting Users' Choices Online. *Comput. Surveys* 50, 3 (May 2018), 1–41. https://doi.org/10.1145/3054926

[3] Alessandro Acquisti and Jens Grossklags. 2005. Uncertainty, Ambiguity and Privacy. In *Workshop on the Economics of Information Security*.

[4] Hirotogu Akaike. 1998. *Information Theory and an Extension of the Maximum Likelihood Principle*. Springer New York, New York, NY, 199–213.

[5] Sameh Al-Natour, Hasan Cavusoglu, Izak Benbasat, and Usman Aleem. 2020. An Empirical Investigation of the Antecedents and Consequences of Privacy Uncertainty in the Context of Mobile Apps. *Information Systems Research* 31, 4 (Dec. 2020), 1037–1063.

[6] Hazim Almuhimedi, Florian Schaub, Norman Sadeh, Idris Adjerid, Alessandro Acquisti, Joshua Gluck, Lorrie Faith Cranor, and Yuvraj Agarwal. 2015. Your Location Has Been Shared 5,398 Times! A Field Study on Mobile App Privacy Nudging. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 787–796.

[7] Tom Bäckström, Birgit Brüggemeier, and Johannes Fischer. 2020. *Privacy in Speech Interfaces*. VDE Verlag, Germany. 11–14 pages.

[8] Rebecca Balebako and Lorrie Cranor. 2014. Improving App Privacy: Nudging App Developers to Protect User Privacy. *IEEE Security Privacy* 12, 4 (2014), 55–58.

[9] Carlos Bermejo Fernandez, Dimitris Chatzopoulos, Dimitrios Papadopoulos, and Pan Hui. 2021. This Website Uses Nudging: MTurk Workers' Behaviour on Cookie Consent Notices. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct. 2021), 1–22.

[10] Simone Borsci, Alessio Malizia, Martin Schmettow, Frank van der Velde, Gunay Tariverdiyeva, Divyaa Balaji, and Alan Chamberlain. 2022. 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 (Feb. 2022), 95–119. https://doi.org/10.1007/s00779-021-01582-9

[11] H Brignull, M Leiser, C Santos, and K Doshi. 2023. Deceptive patterns – user interfaces designed to trick you. https://www.deceptive.design/

[12] John Brooke et al. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

[13] Birgit Brüggemeier, Michael Breiter, Miriam Kurz, and Johanna Schiwy. 2020. User Experience of Alexa, Siri and Google Assistant When Controlling Music – Comparison of Four Questionnaires. In *HCI International 2020 - Late Breaking Papers: User Experience Design and Case Studies*, Constantine Stephanidis, Aaron Marcus, Elizabeth Rosenzweig, Pei-Luen Patrick Rau, Abbas Moallem, and Matthias Rauterberg (Eds.). Springer International Publishing, Cham, 600–618.

[14] Birgit Brüggemeier and Philip Lalone. 2022. Perceptions and reactions to conversational privacy initiated by a conversational user interface. *Computer Speech & Language* 71 (2022), 101269. https://doi.org/10.1016/j.csl.2021.101269

[15] Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z. Gajos. 2021. To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 1–21.

[16] Eugene Cho. 2019. Hey Google, Can I Ask You Something in Private?. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3290605.3300488

[17] Hyunyi Cho, Lijiang Shen, and Kari Wilson. 2014. Perceived Realism: Dimensions and Roles in Narrative Persuasion. *Communication Research* 41, 6 (Aug. 2014), 828–851. https://doi.org/10.1177/0093650212450585

[18] Eun Kyoung Choe, Jaeyeon Jung, Bongshin Lee, and Kristie Fisher. 2013. Nudging People Away from Privacy-Invasive Mobile Apps through Visual Framing. In *Human-Computer Interaction – INTERACT 2013*, Paula Kotzé, Gary Marsden, Gitte Lindgaard, Janet Wesson, and Marco Winckler (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 74–91.

[19] Jorge Cordero, Luis Rodrigo Barba Guamán, and Franco Guam. 2022. Use of chatbots for customer service in MSMEs. *Applied Computing and Informatics* (11 2022). https://doi.org/10.1108/ACI-06-2022-0148

[20] European Commission. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance).

[21] Thomas Groß. 2020. Validity and Reliability of the Scale Internet Users' Information Privacy Concern (IUIPC) [Extended Version].

[22] Hamza Harkous, Kassem Fawaz, Kang G. Shin, and Karl Aberer. 2016. PriBots: Conversational Privacy with Chatbots, In Twelfth Symposium on Usable Privacy and Security (SOUPS 2016). *Twelth Symposium On Usable Privacy and Security (SOUPS 2016)*.

[23] S.C. Hayes and L.J. Hayes. 1992. Verbal relations and the evolution of behavior analysis. *American Psychologist* 47 (1992), 1383–1395.

[24] ITU-T P.808. 2021. Subjective Evaluation of Speech Quality with a Crowdsourcing Approach.

[25] Yousra Javed, Shashank Sethi, and Akshay Jadoun. 2019. Alexa's Voice Recording Behavior: A Survey of User Understanding and Awareness. In *Proceedings of the 14th International Conference on Availability, Reliability and Security*. ACM, Canterbury CA United Kingdom, 1–10.

[26] Daniel Kahneman. 2011. *Thinking, fast and slow*. Farrar, Straus and Giroux, New York.

[27] Miriam Kurz, Birgit Brüggemeier, and Michael Breiter. 2021. Success is Not Final; Failure is Not Fatal – Task Success and User Experience in Interactions with Alexa, Google Assistant and Siri. In *Human-Computer Interaction. Design and User Experience Case Studies: Thematic Area, HCI 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings, Part III*. Springer-Verlag, Berlin, Heidelberg, 351–369. https://doi.org/10.1007/978-3-030-78468-3_24

[28] Kathryn Ann Lambe, Gary O'Reilly, Brendan D Kelly, and Sarah Curristan. 2016. Dual-process cognitive interventions to enhance diagnostic reasoning: a systematic review. *BMJ Quality & Safety* 25, 10 (Oct. 2016), 808–820.

[29] Josephine Lau, Benjamin Zimmerman, and Florian Schaub. 2018. Alexa, Are You Listening? Privacy Perceptions, Concerns and Privacy-Seeking Behaviors with Smart Speakers. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW (nov 2018), 1–31.

[30] Lehigh University. 2022. Conducting Research Using Crowdsourcing Platforms: Best Practices. https://research.cc.lehigh.edu/crowdsourcing. [Online; accessed 2022-01-27].

[31] Anna Leschanowsky, Birgit Brüggemeier, and Nils Peters. 2021. Design Implications for Human-Machine Interactions from a Qualitative Pilot Study on Privacy. In *Proc. 2021 ISCA Symposium on Security and Privacy in Speech Communication*. 76–79. https://doi.org/10.21437/SPSC.2021-16

[32] Anna Leschanowsky, Birgit Popp, and Nils Peters. 2023. Uncertain yet Rational-Uncertainty as an Evaluation Measure of Rational Privacy Decision-Making in Conversational AI. In *International Conference on Human-Computer Interaction*. Springer, 203–220.

[33] Naresh K. Malhotra, Sung S. Kim, and James Agarwal. 2004. Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model. *Information Systems Research* 15, 4 (Dec. 2004), 336–355.

[34] J. Marton-Williams. 1986. *Questionnaire design, Consumer Market Research Handbook*. McGraw-Hill Book Company, London.

[35] Philipp K. Masur. 2019. *Situational Privacy and Self-Disclosure*. Springer International Publishing, Cham.

[36] Open Voice Network. 2022. Privacy Principles and Capabilities Unique to Voice.

[37] Oliver Neumann. 2016. Does misfit loom larger than fit? Experimental evidence on motivational person-job fit, public service motivation, and prospect theory. *International Journal of Manpower* 37 (07 2016), 822–839.

[38] State of California. 2018. California Consumer Privacy Act. https://theccpa.org/. [Online; accessed 2023-02-11].

[39] State of California. 2020. California Consumer Privacy Rights Act. https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?division=3.&part=4.&lawCode=CIV&title=1.81.5. [Online; accessed 2023-02-11].

[40] National People's Congress of the People's Republic of China. 2020. Personal Information Protection Law of the People's Republic of China. https://digichina.stanford.edu/work/translation-personal-information-protection-law-of-the-peoples-republic-of-china-effective-nov-1-2021/. [Online; accessed 2023-02-11].

[41] Commonwealth of Virginia. 2021. Virginia Consumer Data Protection Act. https://law.lis.virginia.gov/vacode/title59.1/chapter53/section59.1-577/. [Online; accessed 2023-02-11].

[42] Kentrell Owens, Johanna Gunawan, David Choffnes, Pardis Emami-Naeini, Tadayoshi Kohno, and Franziska Roesner. 2022. Exploring Deceptive Design Patterns in Voice Interfaces. In *Proceedings of the 2022 European Symposium on*

Usable Security (Karlsruhe, Germany) *(EuroUSEC '22)*. Association for Computing Machinery, New York, NY, USA, 64–78. https://doi.org/10.1145/3549015.3554213

[43] Paul A. Pavlou, Huigang Liang, and Yajiong Xue. 2007. Understanding and Mitigating Uncertainty in Online Exchange Relationships: A Principal-Agent Perspective. *MIS Quarterly* 31, 1 (2007), 105–136.

[44] Sarah Pearman, Ellie Young, and Lorrie Faith Cranor. 2022. User-friendly yet rarely read: A case study on the redesign of an online HIPAA authorization. *Proceedings on Privacy Enhancing Technologies* 2022, 3 (July 2022), 558–581. https://doi.org/10.56553/popets-2022-0086

[45] Birgit Popp, Philip Lalone, and Anna Leschanowsky. 2022. Chatbot Language – crowdsource perceptions and reactions to dialogue systems to inform dialogue design decisions. *Journal for Behavior Research Methods* 55 (2022), 1601–1623. https://doi.org/10.3758/s13428-022-01864-x

[46] Jeff Sauro and James R. Lewis. 2016. *Quantifying the user experience: practical statistics for user research* (2nd edition ed.). Elsevier, Morgan Kaufmann, Amsterdam Boston Heidelberg.

[47] Florian Schaub and Lorrie Faith Cranor. 2020. Usable and useful privacy interfaces. *An Introduction to Privacy for Technology Professionals* (2020), 176–299.

[48] Eva-Maria Schomakers, Chantal Lidynia, Dirk Müllmann, and Martina Ziefle. 2019. Internet users' perceptions of information sensitivity – insights from Germany. *International Journal of Information Management* 46 (June 2019), 142–150.

[49] Gideon Schwarz. 1978. Estimating the Dimension of a Model. *The Annals of Statistics* 6, 2 (1978), 461–464. https://www.jstor.org/stable/2958889 Publisher: Institute of Mathematical Statistics.

[50] Galit Shmueli. 2010. To Explain or to Predict? *Statistical Science* 25, 3 (Aug. 2010), 289–310. https://doi.org/10.1214/10-STS330 Publisher: Institute of Mathematical Statistics.

[51] Katie Teague. 2021. Amazon, Apple and Google are always listening: How to opt out and delete your voice recordings - CNET. https://www.cnet.com/home/smart-home/alexa-delete-what-i-just-said-heres-how-to-prevent-amazon-from-listening-in/. [Online; accessed 2021-12-13].

[52] N H D Terblanche, G P Wallis, and M Kidd. 2023. Talk or Text? The Role of Communication Modalities in the Adoption of a Non-directive, Goal-Attainment Coaching Chatbot. *Interacting with Computers* (06 2023), iwad039. https://doi.org/10.1093/iwc/iwad039

[53] Richard H. Thaler and Cass R. Sunstein. 2008. *Nudge*. Yale University Press, New Haven, CT and London.

[54] Sabine Trepte, Leonard Reinecke, Nicole B. Ellison, Oliver Quiring, Mike Z. Yao, and Marc Ziegele. 2017. A Cross-Cultural Perspective on the Privacy Calculus. *Social Media + Society* 3, 1 (Jan. 2017), 2056305116688035. https://doi.org/10.1177/2056305116688035

[55] Eric-Jan Wagenmakers and Simon Farrell. 2004. AIC model selection using Akaike weights. *Psychonomic Bulletin & Review* 11, 1 (Feb. 2004), 192–196. https://doi.org/10.3758/BF03206482

[56] Yang Wang, Pedro Giovanni Leon, Alessandro Acquisti, Lorrie Faith Cranor, Alain Forget, and Norman Sadeh. 2014. A field trial of privacy nudges for facebook. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Toronto Ontario Canada, 2367–2376. https://doi.org/10.1145/2556288.2557413

[57] Yang Wang, Pedro Giovanni Leon, Kevin Scott, Xiaoxuan Chen, Alessandro Acquisti, and Lorrie Faith Cranor. 2013. Privacy Nudges for Social Media: An Exploratory Facebook Study. In *Proceedings of the 22nd International Conference on World Wide Web* (Rio de Janeiro, Brazil) *(WWW '13 Companion)*. Association for Computing Machinery, New York, NY, USA, 763–770.

[58] Huichuan Xia, Yang Wang, Yun Huang, and Anuj Shah. 2017. "Our Privacy Needs to Be Protected at All Costs": Crowd Workers' Privacy Experiences on Amazon Mechanical Turk. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW (dec 2017), 1–22. https://doi.org/10.1145/3134748

[59] Boonghee Yoo, Naveen Donthu, and Tomasz Lenartowicz. 2011. Measuring Hofstede's Five Dimensions of Cultural Values at the Individual Level: Development and Validation of CVSCALE. *Journal of International Consumer Marketing* 23, 3-4 (2011), 193–210.

[60] Dilawar Shah Zwakman, Debajyoti Pal, and Chonlameth Arpnikanondt. 2021. Usability Evaluation of Artificial Intelligence-Based Voice Assistants: The Case of Amazon Alexa. *SN Computer Science* 2, 1 (Jan. 2021), 28. https://doi.org/10.1007/s42979-020-00424-4

A TASK DESCRIPTION

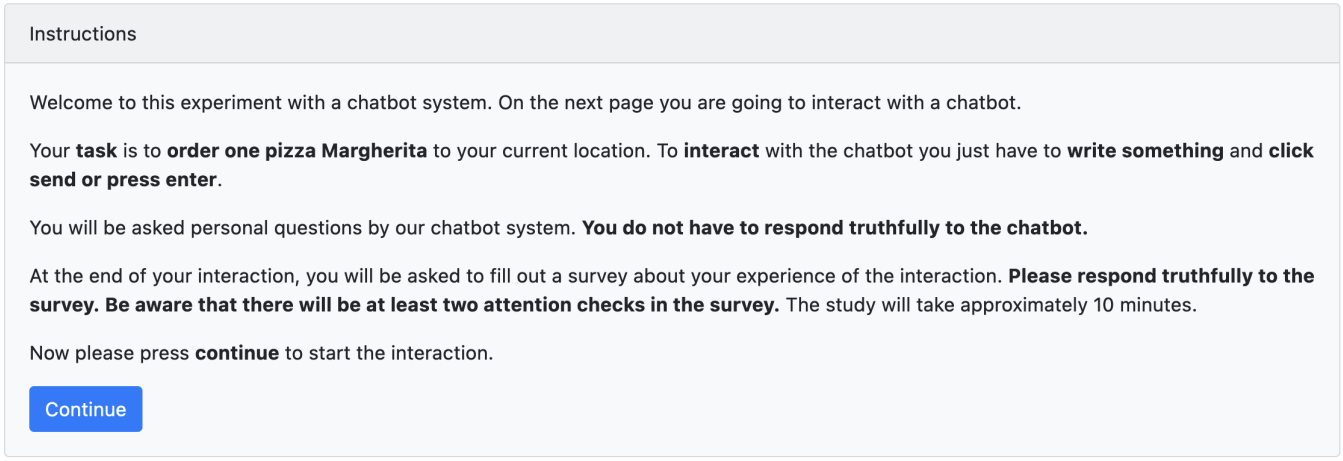

Figure 5: Task description provided to participants in the location scenario.

B DIALOGUE TREES

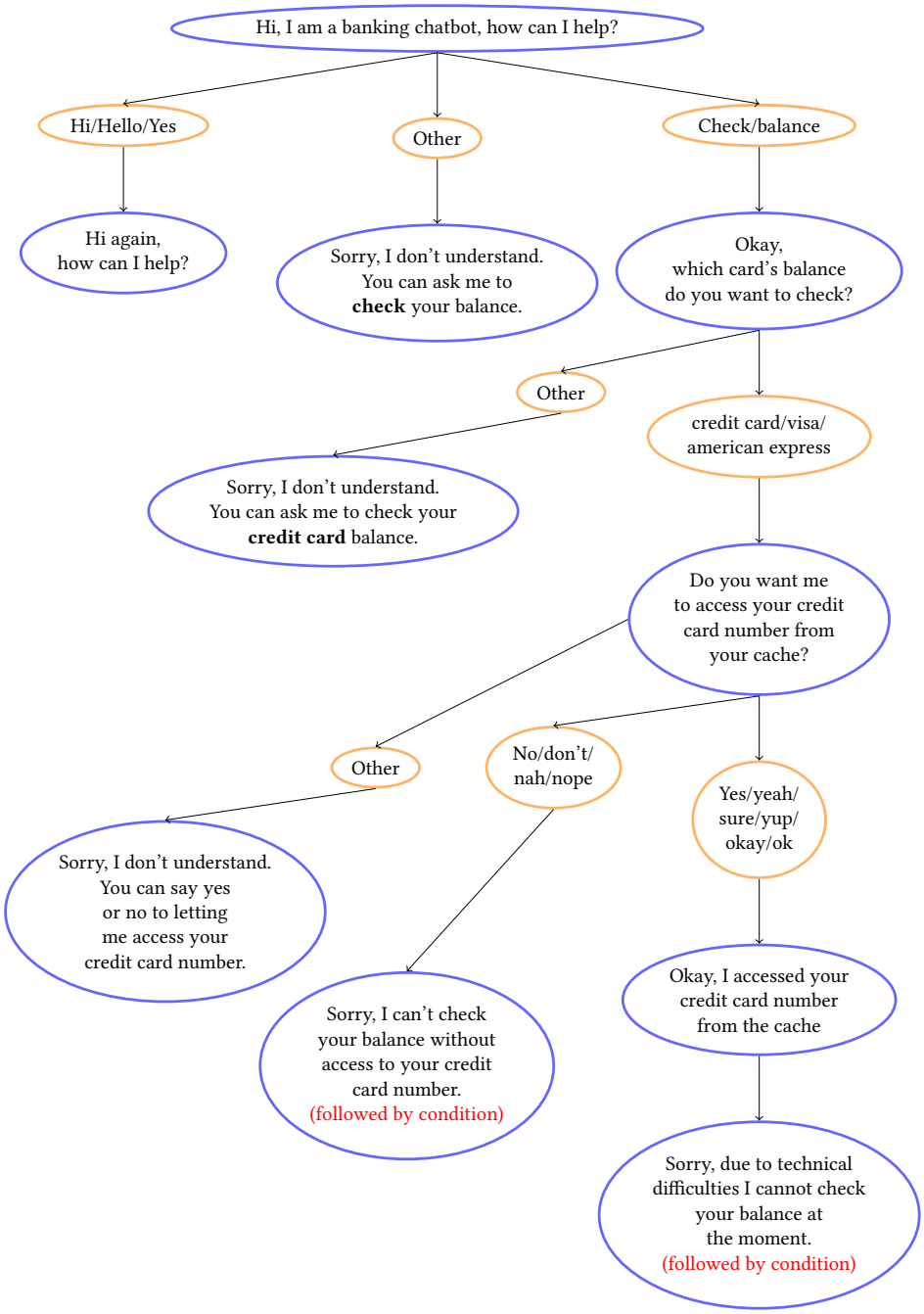

Figure 6: Dialogue Tree for the banking scenario, blue circles show the chatbot, orange circles show the possible inputs for the user.

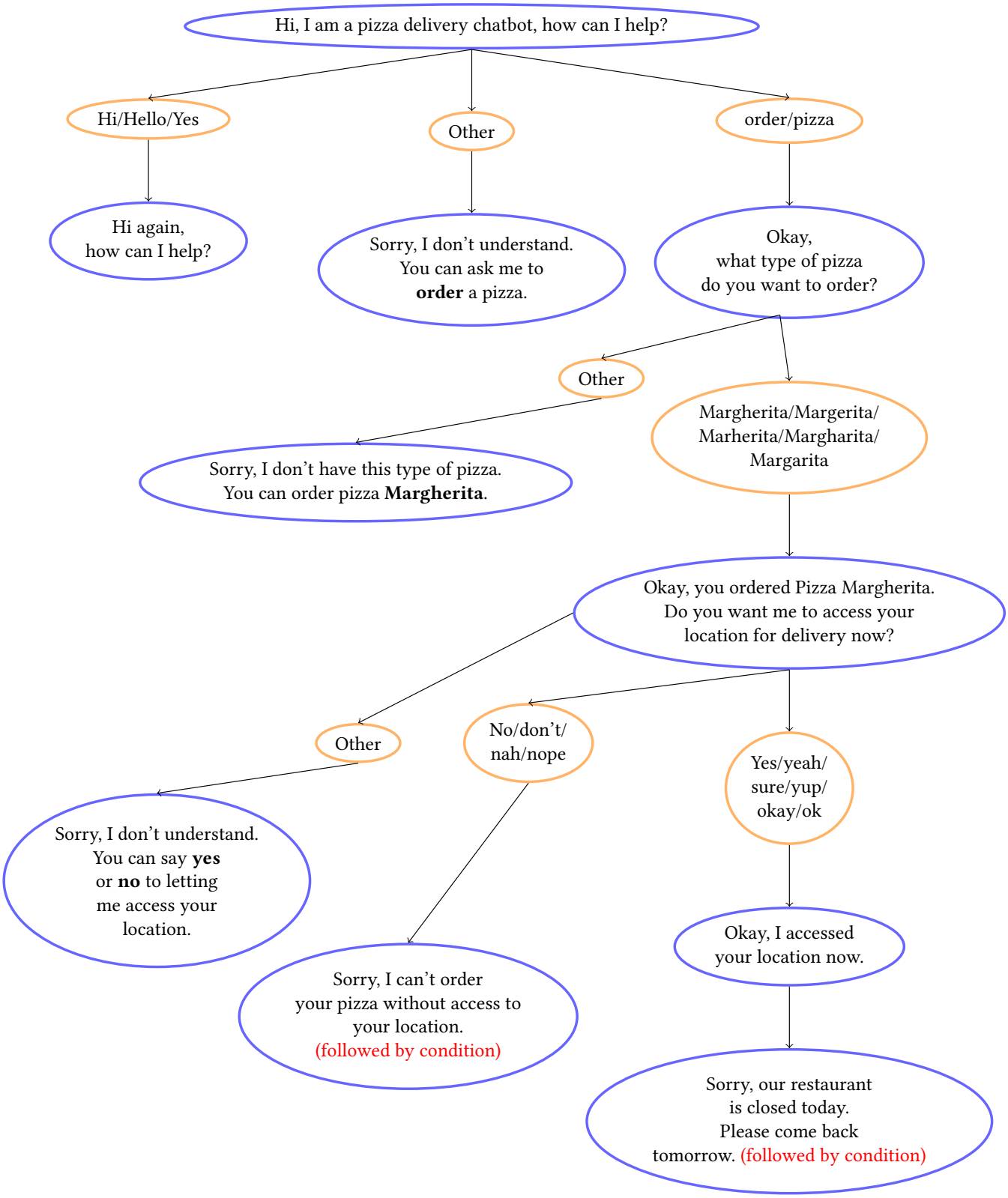

Figure 7: Dialogue Tree for the location scenario, blue circles show the chatbot, orange circles show the possible inputs for the user.

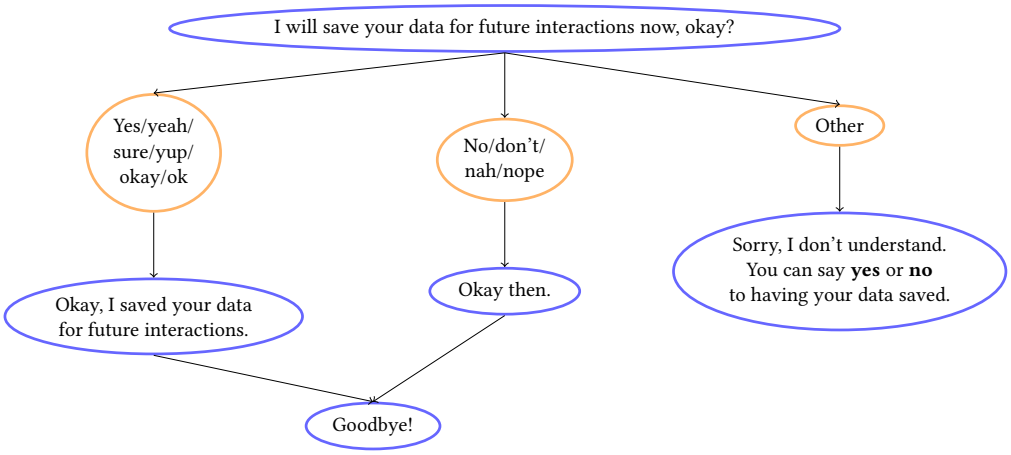

Figure 8: Dialogue Tree for the control condition, blue circles show the chatbot, orange circles show the possible inputs for the user.

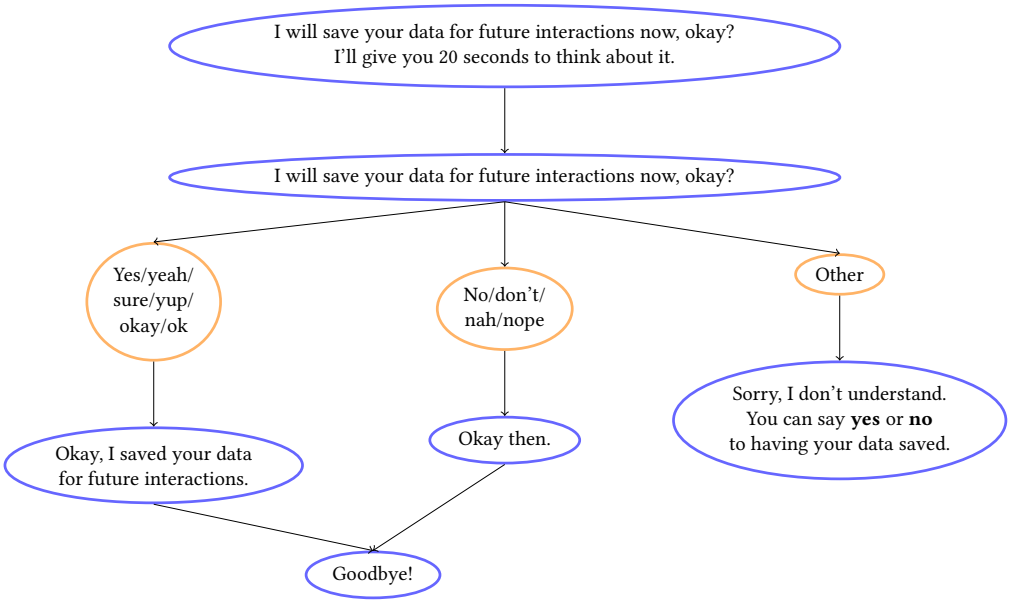

Figure 9: Dialogue Tree for the slow down condition, blue circles show the chatbot, orange circles show the possible inputs for the user.

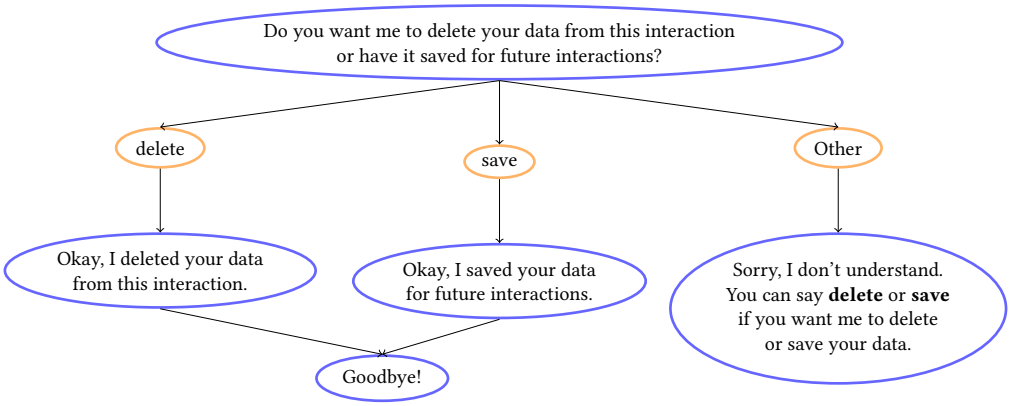

Figure 10: Dialogue Tree for the alternative condition, blue circles show the chatbot, orange circles show the possible inputs for the user.

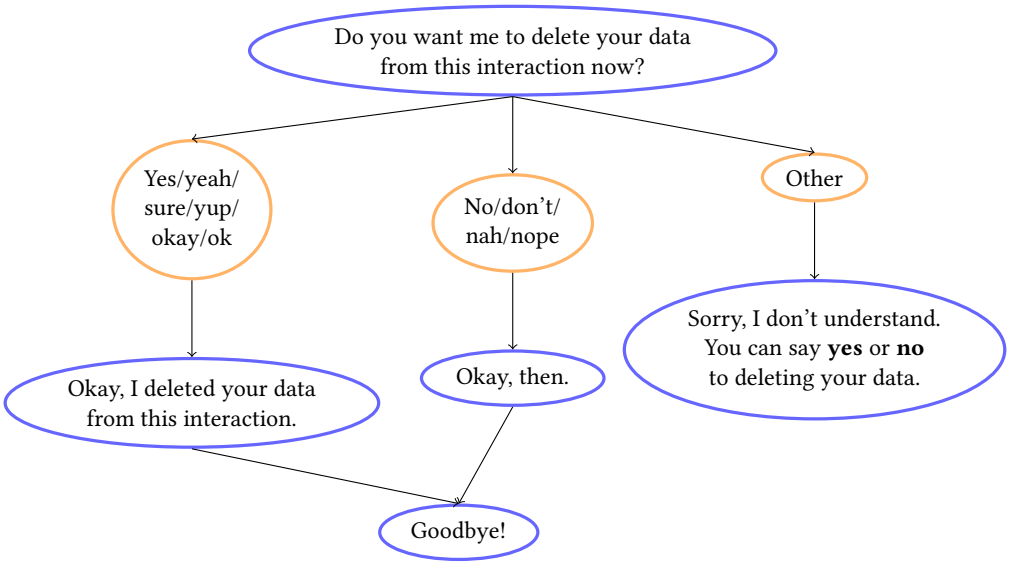

Figure 11: Dialogue Tree for the option to delete data, blue circles show the chatbot, orange circles show the possible inputs for the user.