

“It seemed like an annoying woman”: On the Perception and Ethical Considerations of Affective Language in Text-Based Conversational Agents

Lindsey Vanderlyn, Gianna Weber, Michael Neumann,
Dirk V ath, Sarina Meyer, Ngoc Thang Vu

Institute for Natural Language Processing
University of Stuttgart, Germany

`lindsey.vanderlyn@ims.uni-stuttgart.de`

Abstract

Previous research has found that task-oriented conversational agents are perceived more positively by users when they provide information in an empathetic manner compared to a plain, emotionless information exchange. However, users’ perception and ethical considerations related to a dialog systems’ response language style have received comparatively little attention in the field of human-computer interaction. To bridge this gap, we explored these ethical implications through a scenario-based user study. 127 participants interacted with one of three variants of an affective, task-oriented conversational agent, each variant providing responses in a different language style. After the interaction, participants filled out a survey about their feelings during the experiment and their perception of various aspects of the chatbot. Based on statistical and qualitative analysis of the responses, we found language style played an important role in how human-like participants perceived a dialog agent as well as how likable. Language style also had a direct effect on how users perceived the use of personal pronouns ‘I’ and ‘You’ and how they projected gender onto the chatbot. Finally, we identify and discuss ethical implications. In particular we focus on what factors/stereotypes influenced participants’ impressions of gender, and what trade-offs a more human-like chatbot brings.

1 Introduction

Conversational Agents (CAs) have been an area of interest for many years. As technology has advanced, these devices have become more capable and further embedded in users’ daily lives. With the advent of personal assistants, like Apple’s Siri and Amazon’s Alexa, focus has begun shifting away from simply conveying correct information to additionally considering the user experience (Yang et al., 2019) and *how* that information is conveyed.

Rather than seeing a CA as an inanimate tool,

users tend to view them as social actors (Nass et al., 1994) and infer or assign them personality traits. These perceived personas, whether intentionally designed into the CA or not, can influence how users perceive and interact with these systems. Therefore much research has gone into how to design appropriate personas (Kim et al., 2019; Nass et al., 1995) and investigate how users interpret them (Lee et al., 2019). However, while there is a study exploring *ethical* issues in text-based affective dialog systems (Kretzschmar et al., 2019), comparatively little research has been done in this area.

Open ethical questions thus remain unanswered around how users perceive and react to different dialog system designs. In this paper, we are particularly interested in how users perceive and emotionally react to a system mimicking human thoughts or emotional awareness, as well as what biases users project onto a dialog system based on cues in its language. One concrete example of this arises around the question whether conversational agents should use the pronoun “I” or convey information in passive voice as using a first person pronoun can imply to the user a false sense of self-perception.

Therefore, we explore the following research questions and hypotheses:

- **(RQ1)** How do users perceive and react to a conversational agent portraying human-like characteristics?
 - (H1a)** How users *perceive* the CA will depend on the language styles employed.
 - (H1b)** How users *feel* during the interaction will depend on the language style employed.
 - (H1c)** How users perceive the CA’s use of personal pronouns ‘I’ and ‘You’ will depend on the language style employed.
- **(RQ2)** Do users assign a gender to the CA as a result of the employed language style?

(H2) If/how users assign a gender to the CA will depend on the language style employed.

In order to answer these questions, we developed a dialog system to provide information about exam registration to students, an often emotional scenario. For this CA, we designed three different natural language templates, each providing identical information but each using a different language style. Two language styles were meant to be affective, reflecting either an empathetic or matter-of-fact personality, and the third was meant to be machine-like, using passive voice address. We then performed a user study where 129 participants interacted with and answered a survey about one of the three variants of the CA. The dialogs along with the survey responses from this user study will be publicly available for further research.¹

The main findings of this work are: RQ1) Users prefer an empathetic language style over a pure information exchange, and generally find the use of the pronoun ‘I’ natural, suggesting human-like attributes may not be inherently ethically problematic as long as there is transparency about the agent’s artificial nature. RQ2) A non-gendered, empathetic dialog agent allows users to project a gender (often their own) onto the chatbot, which could help avoid negative stereotypes associated with one gender over the other.

2 Related Work

2.1 Human-Like Conversational Agents

In their most basic form, conversational agents are software-based systems that communicate through natural language (Feine et al., 2019a). This means that CAs must solve the task of understanding natural language input, tracking information throughout a dialog, choosing the correct response, and generating a natural language output to communicate that response (Ortega et al., 2019). It is important to note that choosing the next response and communicating it to the user are two distinct steps: separating *what* is said from *how* it is said (Gatt and Kraemer, 2018). This allows for the generation of more human-like responses (Diederich et al., 2019).

Among these more human-like outputs are affective responses and responses designed to convey

¹https://github.com/DigitalPhonetics/ethics_in_chatbot_design

a particular persona (Gill et al., 2012). Affective responses fall into two categories: those that reflect the emotional state of the producer and those that try to induce an emotional state in the receiver (Gatt and Kraemer, 2018). In this work, we focus on the latter, particularly the feelings and reactions of users in response to different language styles.

2.2 Human Responses to Affective Conversational Agents

For a holistic understanding of user experience, research must go beyond pragmatically evaluating a system’s functionality/usability to also consider the hedonic/affective aspects of the user’s experience (Hassenzahl and Tractinsky, 2006; Hassenzahl, 2018). Studies examining the effects of affective dialog systems have found that a system’s affective output can impact how the user perceives the agent (Diederich et al., 2019; Fitzpatrick et al., 2017; Zhou et al., 2018; Brave et al., 2005; Verhagen et al., 2014). Tzeng (2004) found that when their system apologized during a malfunctioning scenario, users perceived the environment as more comfortable and less mechanical. Similarly, Verhagen et al. (2014) found that conveying friendliness increased the perception of personalization and social presence in users, leading to greater experience satisfaction. Additionally, chatbots have found growing popularity in the mental health sector after it was found that internet-delivered, guided cognitive behavioral therapy (CBT) can be as effective as face-to-face CBT (Andersson et al., 2014).

Based on these works, we chose to also implement an empathetic language style to investigate if similar trends hold in a more controlled setting, i.e., where only the language style varies between conditions, rather than also dialog content.

2.3 Ethical Perspectives on Affective Computing and Chatbots

Most studies demonstrating the effects of affective response generation disregard potential ethical questions that arise from human-like behavior produced by these systems. Dignum (2018) emphasizes that CAs are still manufactured artifacts whose primary aim is to help users accomplish a certain task, and that it is thus mandatory to identify ethical concerns related to human-like behavior and address them in their development.

To date, there are only few studies focusing on the emotional impact conversational agents have on users, including Brave et al. (2005); Huang et al.

(2001); Wuenderlich and Paluch (2017); Yang et al. (2019); Portela and Granell-Canut (2017); McDonnell and Baxter (2019). These studies found that dialog systems leveraging human-like cues can have not only a social, but also a psychological impact on users. For this reason, general recommendations about ethical concerns in affective computing are proposed in Institute of Electrical and Electronics Engineers (2017), e.g. transparency about a system’s affective nature, and considering the risk of causing emotional harm through manipulating, dampening, or amplifying a user’s emotions. In a study on the ethics of text-based mental health chatbots Kretzschmar et al. (2019) criticize how these agents give users the impression of interacting with an empathetic listener, but are not capable of identifying and considering a user’s individual needs. Another ethical concern stems from what sort of biases users bring in based on the design of a CA. For example, adding gender to a chatbot can emphasize biases and preconceptions from users (West et al., 2019).

3 Implementation

To create the text-based, task-oriented dialog system for this study, we used the open-source toolkit Adviser (Li et al., 2020). We created a domain ‘examination matters’ for the system built around the scenario of a ‘student advisor chatbot’, which can provide information on exam-related topics at the university. This domain provides a good lens through which to investigate users’ reactions to a dialog system as this topic is frequently related with negative emotions or stress.

To study affective response generation, we implemented pre-defined templates for three different response styles: two sets of affective responses, designed based on Myers-Briggs Type Indicator (MBTI) personality types; and one set of passive voice templates, meant to be purely informative, inspired by Huang et al. (2001). Depending on the chosen language style for the interaction, the corresponding response template was used to communicate the chatbot’s intent.

For this study, both affective response templates were designed to reflect the traits “Extraversion”, “Sensing”, and “Judging”, but varied so that one set of templates reflected the trait “Feeling” (ESFJ personality) whereas the other reflected the trait “Thinking” (ESTJ personality). These types are both described as decisive, realistic, organized and

practical. Adjectives describing the ESFJ personality are: supportive, outgoing, cooperative, warm, sympathetic, friendly, and loyal. In contrast, adjectives describing the ESTJ personality are: assertive, logical, responsible, efficient, straightforward, pragmatic, objective, and structured. These personalities were chosen to help investigate the perception of gender as the adjectives describing “Feeling” are more stereotypically associated with women while the adjectives describing “Thinking” are more stereotypically associated with men.

The **ESFJ language style** provides information in a warm, cooperative and empathetic manner. Responses contain personal pronouns (‘I’, ‘my’, ‘you’, ‘your’), hedging phrases (e.g. “I suggest”, “I recommend”) and modal verbs like “should”. Some expressions are vague and open to interpretation. Affective ESFJ responses demonstrate an empathetic understanding of student needs. Consequently, this language style can be viewed as affective in the sense that the users’ emotional state might be altered.

The **ESTJ language style** provides practical and matter-of-fact information in precise manner. Responses contain assertive and also passive-voiced phrases, thereby still addressing the user (e.g. ‘you’, ‘your’). Self-reference (‘I’) is less used, occurring half as many times as in ESFJ templates. Affective ESTJ responses are direct and remind students to follow the rules and requirements.

The **passive voice language style** reflects the artificial nature of a CA. Responses are phrased solely to provide information. They do not contain pronouns or emotional phrases, but consist mainly of passive infinitives and modals.

Example (1) depicts a passive voice, ESFJ and ESTJ response for the same user utterance.

- (1) **User:** Seems like I have missed the registration period. Can I also register for my requirements module later?
 - a. **Passive voice:** Belated registrations are not accepted by the examination office. Without a registration, the module cannot be sit and has to be taken in the semester in which it is offered next.
 - b. **ESFJ:** I’m really sorry, but belated registrations will not be accepted by the examination office and unfortunately, you’re also not allowed to take the module without a registration. For you, this means that you cannot take this module until the next time it is offered.

- c. **ESTJ:** You have to pay attention to registration deadlines! Belated registrations won't be accepted by the examination office and just for your information, without a registration, you are not authorized to take the module. Specifically, this means that you can only take this module in the semester in which it will be offered next time.

4 Pilot Study

4.1 Procedure

Prior to the main experiment, we ran a pilot study with 12 university students. Each participant interacted with all three CA variants (within-subject design) and answered a questionnaire after each interaction. Originally planned as an on-site, supervised experiment, the project was adapted to an online study and survey due to restrictions caused by the Covid-19 pandemic.

Participants were instructed to imagine themselves in a situation where they urgently needed to get exam-related information, but the student advisor was not available, so they had to chat with the "Exams chatbot". The instructions included a script with 11 conversational goals for the user to accomplish in the form of statements and questions to be typed to the chatbot. The user was directed to fulfill one goal per turn. The conversational goals remained the same for all three dialog system variants. Using identical instructions ensured that any differences participants noticed were as a result of differences in the language style and not in the presented scenarios. Participants were instructed to formulate the statements and questions in their own words, but to use specific highlighted words, to make sure their intent was understood by the system. In the case that the CA could not recognize the user intent or provides a thematically inappropriate response, participants were encouraged to retry using the exact wording from the script.

4.2 Results

The main findings from the pilot study are that an empathetic, self-referencing language style is generally preferred and perceived as most realistic (10 out of 12 participants commented that the ESFJ CA's increased usage of 'I' led to the perception of a more realistic and human-like agent). Further, affective language appears to affect users emotionally (both positive and negative) and influences the perceived gender of an agent. Only 25% of par-

ticipants assigned a male or female gender to the passive CA compared to 58% for the ESFJ and ESTJ variants each.

5 Main Study

5.1 Procedure

For the main study, we recruited participants from Amazon Mechanical Turk. Therefore, to reduce the time needed for the experiment and to reduce biasing effects, we changed to a between-subject design. This meant that each participant only interacted with one dialog system variant and were not aware that there could be other variants. As in the pilot study, each participant was asked to imagine that they were a student in need of information about their exam registration and were shown a set of conversational goals they should accomplish through a text-based dialog with a chatbot. Participants were again required to fulfill all of the listed conversational goals. To help encourage this, when a goal was accomplished, the instructions for it would change color (see figure 1 on the following page). This system indicated to the user when they could move on to the next goal and also ensured that users had to complete all goals in the correct order before they could take the survey. Once the dialog was complete, participants were asked to fill in a survey about the experience and submit a verification code to us. In total the experiment took an average of 21 minutes to complete and participants were compensated \$3 (\$9/hr).

5.2 Participants

129 participants were recruited for the survey; after removing duplicates and bad faith dialogs/surveys, 127 participants (62 female, 63 male, 2 other) remained. All subjects live in the United States and reported English as their native language.

5.3 Evaluation Methods

5.3.1 Questionnaires

To explore our research questions, we created a post-dialog questionnaire, combining quantitative measures (semantic differentials, Likert-scales and a "well-being" metric) and qualitative measures (free-form comments). A full version of the questionnaire can be seen in Appendix A.

RQ1(a): General perception and language style. The general perception of the conversational agent was assessed with three semantic differentials from

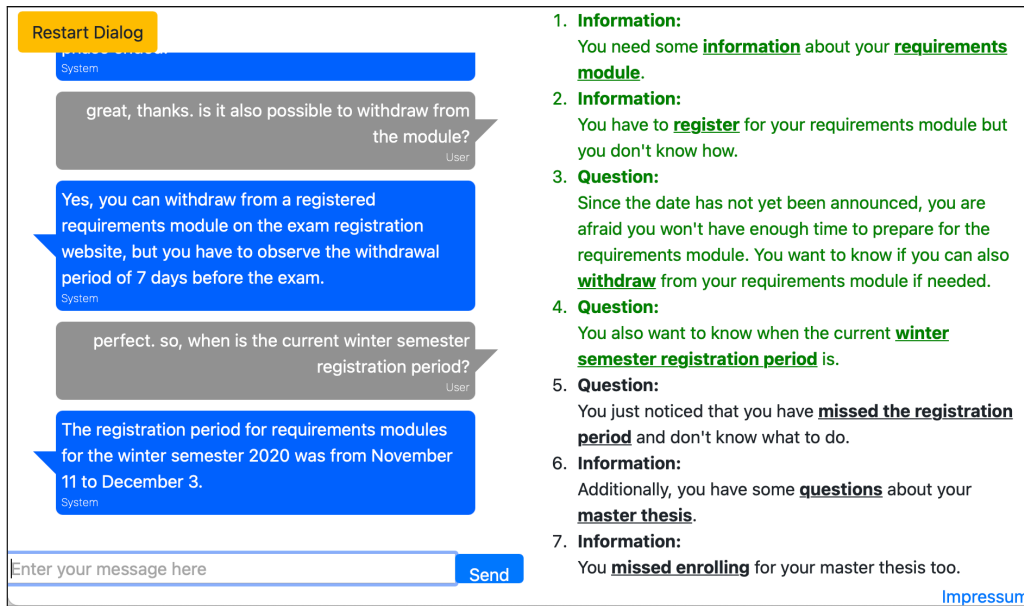


Figure 1: Screenshot of the web-based dialog system. The chat is shown side by side with the task instructions.

| | 1 | 2 | 3 | 4 | 5 | |
|--------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|--------------------|
| Disruptive | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | Calming |
| Exaggerated | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | Appropriate |
| Unrealistic | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | Realistic |

Figure 2: SD for the question ‘Please rate how you perceived the chatbot’s *language style*.’

the Godspeed questionnaire (Bartneck et al., 2009), which was designed to measure users’ perception of robots. We asked about humanness (anthropomorphism in the Godspeed questionnaire), likability, and intelligence of the CA.

The perception of the CA’s language style was assessed with a three-item semantic differential scale (SD), shown in figure 2.

RQ1(b): Feeling during interaction. Participants’ feelings during the interaction were measured with the Scale of Positive and Negative Experience (SPANE) (Diener et al., 2010). Participants were asked to indicate how much they experienced positive (six items: positive, good, pleasant, happy, joyful, contented) and negative (six items: negative, bad, unpleasant, sad, afraid, angry) feelings on a scale from 1 (“not at all”) to 5 (“a lot”). The sums of these scores form the SPANE Positive and Negative scores (ranging from 6 to 30 each). By subtracting the Negative from the Positive score, an overall well-being score, the SPANE Balance

(SPANE-B), is calculated (scale from -24 to 24).

RQ1(c): Perception of ‘I’ and ‘You’. From both a usability and an ethical perspective, we wanted to find out how users perceive self-reference by the CA (expressed by the personal pronoun ‘I’) and being addressed directly with ‘you’ or ‘your’. For this, we created a three-item SD: “How did you perceive the chatbot’s use (or lack of use) of the personal pronoun ‘I’ / ‘You’?” – eerie ↔ natural, dispensable ↔ indispensable, unsuitable ↔ suitable.

RQ2: Perception of gender. To find out whether users project a certain gender onto the chatbot (which was designed without any cues for gender), we asked, “If you had to assign a gender to the chatbot, what would it be – male, female, genderless?” Participants were also asked to explain their choice in a free-form comment.

Additional questions The last section of the survey contains two questions about *satisfaction of the interaction* in terms of overall interaction and task completion (5-point Likert scales). Finally, participants were asked to write about any positive and negative experiences as free-form responses.

5.3.2 Quantitative Analysis

To analyze the quantitative survey feedback, we conducted ANOVA analysis per question category to determine if there were any significant differences between the different linguistic styles. For

any question categories where the null hypothesis was rejected, we performed a post-hoc Tukey test to determine which differences existed between the groups. Results with a p -value less than 0.008 represent significant differences after applying correction for testing multiple hypotheses.

5.3.3 Qualitative Analysis

The free-form responses were evaluated manually in a qualitative content analysis. First, one author derived coding categories for each free-form question from a subset of the responses. Then, another author independently assigned each response to one or more of these codes, with the option to create a new category if needed. This procedure is referred to as conventional content analysis in Hsieh and Shannon (2005). It allows us to obtain the overall content of all responses in a categorized and countable way. Examples of these codes can be seen in Appendix B.

6 Results

6.1 RQ1(a): General Perception and Language Style

Although both the ESFJ and ESTJ language styles were perceived as significantly more human-like than the passive style, only the ESFJ style was seen as significantly more likable. When asked how appropriate they found the language style, users additionally saw only the ESFJ chatbot as significantly more appropriate than the baseline. These differences can be seen in table 1 on the following page.

In the free-form responses about general, positive and negative impressions, participants were overall satisfied with the *detailed answers* (46)² and appreciated the *efficiency* (11) of the conversation. Obtaining immediate responses without lag was generally regarded positively (27), but three subjects disliked it because it broke the illusion of having a conversation with a human. Further negative impressions concerned the CA's *limited understanding* (36) due to its dependence on keywords in the user input.

Regarding differences between the CA variants, survey responses mainly confirm the personality traits described in section 3. Only the ESFJ variant was experienced as *friendly and engaging* (10),

²Coding categories are in italics; numbers in parentheses show how many user responses were mapped to a particular code; one response can be mapped to multiple codes

whereas the other two were judged to be *cold and without empathy* (5 ESTJ, 7 passive). Seven subjects complained about the ESTJ agent responding in a *rude or harsh way* (ranging from “negative in tone” (P42) to “curt and sort of accusatory” (P81)). Compared to the passive agent, more subjects of the ESFJ and ESTJ variants mentioned they liked the answers they received (*detailed/good answers*; 10 passive, 18 ESFJ, 18 ESTJ). Similarly, more participants perceived the ESFJ and ESTJ agents to be *human-like* (5 and 8) than the passive one (3).

6.2 RQ1(b): Feeling during Interaction

Looking at the mean difference in SPANE Balance score (see table 1), users rated the ESFJ chatbot on average 6 points higher (scale -24 to 24) than the passive voice variant ($p=0.014$). This indicates they felt more positive emotions when interacting with the ESFJ chatbot compared to the passive one. As for the mean SPANE Positive and SPANE Negative scores (scale 6 to 30), we observed that participants experienced overall less negative emotions than positive ones across all three CA variants (SPANE Positive=20.8; Negative=8.8).

6.3 RQ1(c): Perception of ‘I’ and ‘You’

Regarding the free-form responses about the use or lack of use of ‘I’, most subjects either *did not notice anything unusual* (26), judged the responses to be *natural* (28), or found the usage *appropriate* (23). No major differences were found between the CA variants, although the passive agent does not use personal pronouns at all. Frequent usage of ‘I’ seems to increase perceived humanness: 13 subjects of the ESFJ group consider the agent to be *human-like* due to the pronoun (e.g. “It almost made me forget that it was a bot!” (P48, ESFJ)) whereas only two subjects noted this about the ESTJ variant, which uses ‘I’ half as often as ESFJ.

Similarly, the use or lack of use of ‘you’ was judged to be *appropriate* (35) or was *not noticed as unusual* (23) for all CA variants. In contrast to ‘I’, the use of ‘you’ (ESFJ and ESTJ) was more often reported as *natural* (25) than the lack of it (passive, 5). Participants also reported that the pronoun ‘you’ *made the conversation more personal* (10), e.g. “Using you made me feel like it cared” (P116).

6.4 RQ2: Perception of Gender

85 of 127 participants perceived the CA to be genderless. The most frequent explanations fell into three categories: *did not notice a gender bias* (37),

| | $\mu_{\text{ESFJ}} - \mu_{\text{Passive}}$ | $\mu_{\text{ESTJ}} - \mu_{\text{Passive}}$ | $\mu_{\text{ESTJ}} - \mu_{\text{ESFJ}}$ |
|--------------------------------|--|--|---|
| Humanness | 0.9436** | 0.7668** | -0.1767 |
| Likability | 0.7446** | 0.0838 | -0.6608** |
| Language Style (acceptance) | 0.4917** | -0.0497 | -0.5414** |
| I Pronoun (acceptance) | 0.4655* | 0.0147 | -0.4508* |
| SPANE-B | 6.0855* | 1.3387 | -4.7468 |
| Task Satisfaction | 0.4513* | 0.1676 | -0.2837 |

* $p < .05$; ** $p < .008$

Table 1: The values in the table represent the differences in mean scores ($\mu_a - \mu_b$) between the different language styles (column headers) for each of the scales (row labels). All scales except SPANE-B range from 1 to 5, p -values were calculated using a Tukey test. $\alpha = 0.05$ represents our original confidence level, $\alpha = 0.008$ reflects the p value corrected for testing multiple hypotheses. All entries with ** are significant after correction.

perceived as a genderless machine (13), and *chatbots do/can not have a gender* (17). In the remaining 42 answers we observed two patterns: The passive CA was mainly perceived as male (10 out of 12), whereas for the ESFJ CA a high positive correlation exists between the assigned and the participant’s own gender (Spearman $r = 0.73$). For the ESTJ agent no apparent pattern can be seen in the results.

The free-form responses provide insights about which language styles trigger specific gender associations. For instance, *short responses* were associated with a male gender (2), whereby *long and detailed answers* let subjects think they were talking to a woman (2). A *friendly* (4) and *calm* (4) tone, as used especially in the ESFJ agent, and the scenario of *talking to someone working in a university office* (1) also triggered a female perception. Being *rude, harsh, or naggy* (adjectives associated with the ESTJ variant), however, was equally often given as explanation for male/female (3 and 3). Interestingly, four of those six subjects associated this negative impression with their own gender (e.g. “It was bossy and rude and seemed like an annoying woman.” (P22, female subject)).

6.5 Comparison to the Pilot Study

The findings presented above coincide with and strengthen the results from our pilot study: an empathetic CA is preferred (higher likability) and the language style can affect users emotionally (differences in SPANE-B score). In the pilot study, we also observed the pattern that participants tend to assign their own gender to the ESFJ CA (aside from the majority selecting genderless for all variants).

7 Discussion

7.1 Ethical Implications

In section 1 we outlined two major ethical concerns regarding chatbot design that we aimed to investigate: the illusion of humanness (related to the use of affective language templates and ‘I’ as self-reference) and gender biases and preconceptions.

7.1.1 Humanness

From the user feedback, adding personality (ESFJ and ESTJ affective language styles) led participants to view the chatbot as more human-like and more likable. This however, raises the question how to make human-like dialog agents ethically correct. There is obviously a tension here, between making a chatbot seem more human, and thus more likable, and the ethical issues inherent in convincing a user, either consciously or unconsciously, that they are talking with a sentient being. That is, by creating a chatbot that seems empathetic, there is the ethical concern that users will assume it is capable of actually understanding their emotional needs.

Additionally, there is the question of whether or not chatbots should use self-referential pronouns (e.g., ‘I’), which can imply a level of sentience from the chatbot that it does not actually possess. Previous research found that such pronouns can be poorly received in spoken dialog systems (Huang et al., 2001). However, in our text-based scenario, it was clear for most participants that they were chatting with a machine. Personal pronouns (both ‘I’ and ‘You’) did not mislead users nor were they perceived as out of place (as many users stated they *did not notice* them or that it *felt normal*).

Therefore, we can infer that certain human-like traits are not inherently problematic so long as

transparency about the artificial nature of the dialog system is ensured. The guideline should be to create enjoyable human-like agents that clearly identify themselves as artificial and transparently state their limits (cf. Kretzschmar et al. (2019)). In our study, users knew they were talking with an automated system because the chatbot introduced itself as such and in our limited domain this appears to have been sufficient. However, the exact level of transparency necessary will depend on the application in which the chatbot is employed and how likely a user is to mistake the nature of their dialog partner. For example, in mental health applications transparency is critical (Kretzschmar et al., 2019) and simply introducing the CA as non-human at the beginning of the interaction may not be sufficient.

Another important aspect to consider in ethical design is the potential harm a system may cause users. While the empathetic persona was well liked and well regarded by participants, this was not the case for the ESTJ style which was often perceived as rude. One participant even suggested the system should “refrain from judgements” (P100). Therefore, while increasing humanness can increase user engagement, it is important to consider the emotional well-being of the users when deciding on a persona and it may be preferable for the system to remain neutral rather than judging participants. Well-being metrics such as the SPANE score should be included in the evaluation of dialog systems that convey a certain persona. In our case, the results showed that on average none of the chatbot variants evoked strong negative feelings in the users (SPANE Negative score of 8.8).

7.1.2 Gender

A recent UNESCO report (West et al., 2019) revealed that most popular spoken dialog systems are designed as female and discussed some how it is ethically problematic to have this type of subservient, error-prone technology portrayed as a woman. These findings were augmented by McDonnell and Baxter (2019) and Feine et al. (2019b) who find that even in text-based chatbots, the large majority are designed to present as female through their avatar, name, or the pronouns they are described with. However, it is unclear if non-gendered chatbots are the solution.

Our work shows that even without any of these markings, user biases and stereotypes may still cause them to project a gender onto a dialog system. For example, although we designed the pas-

sive style chatbot to be gender neutral, it was perceived by 25% of participants as male (and 5% as female), with statements indicating users associated the factual information delivery style with men. In contrast, the affective chatbot which used an empathetic linguistic style was perceived by 35% of participants as their own gender (and 5% as the opposite), with users providing positive attributes of the dialog system (e.g., ‘calm’ or ‘helpful’) as their justification. Although these numbers represent less than half of users, given the scale at which chatbots are being employed, this is still a large population whose experience cannot be ignored. Therefore we propose that a non-explicitly gendered, empathetic language style can allow users to project a gender (often their own) onto the chatbot, which may help avoid negative stereotypes associated with one gender over the other and provide an alternative to explicitly gender neutral chatbots.

7.2 Design Recommendations

Based on our findings and the ethical implications discussed above, we identified the following design aspects to play an important role in creating a conversational agent:

- (I) An empathetic, human-like language style should be preferred over pure information exchange – the language style might even affect task satisfaction. Direct language that is command-like or even judgemental is least accepted by users.
- (II) Personal pronouns and a self-referential language are not necessary for task-oriented chatbots, but they contribute to a positive, more personal, and natural experience.
- (III) Without other cues for gender (avatar or voice), language style alone affects user perception. Using a non-marked empathetic style may help avoid negative gender stereotypes.

8 Future work

As this study made use of crowdsourced workers who were roughly within the same age range and from the same cultural background, in the future, it would be interesting to see what role the age and cultural background of the user plays in how they perceive a conversational agent. In particular, it would be interesting to explore if cultural background adds more nuance to the general design recommendations we derived.

9 Conclusion

We presented a user study on the perception of affective language in conversational agents (CAs) and discussed ethical concerns in this context. Based on participants' survey responses, we showed that users prefer an empathetic, human-like agent, and that a third of the users viewed the agent as gendered, partially based on linguistic cues. From these findings, we infer that human traits in a CA are not inherently problematic from an ethical standpoint, as long as transparency is always ensured. Further, gender biases need to be considered when designing a CA, as the lack of explicit gender cues does not ensure a neutral perception.

10 Ethical Considerations

Through this work, we took care to ensure informed consent from participants and anonymized all collected data so ensure it could not be traced back to participants. Participants were compensated approx. 9\$/hr for their time based on US minimum wage of \$7.25. As no identifying information was collected, this study was not deemed to require review from a data protection officer. To ensure quality of the data collected, we went through by hand and removed any participants whose free-response answers indicated they had not actually engaged with the dialog system. As participants came only from the United States and were all in the age range 21-70, data from our experiment can not be certain to generalize to other cultures or populations.

Acknowledgments

This work was partially funded by the Carl Zeiss Foundation.

References

- Gerhard Andersson, Pim Cuijpers, Per Carlbring, Heleen Riper, and Erik Hedman. 2014. Guided Internet-Based vs. Face-to-Face Cognitive Behavior Therapy for Psychiatric and Somatic Disorders: A Systematic Review and Meta-Analysis. *World psychiatry: Official Journal of the World Psychiatric Association (WPA)*, 13(3):288–295.
- Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*, 1(1):71–81.
- Scott Brave, Clifford Nass, and Kevin Hutchinson. 2005. Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. *International journal of human-computer studies*, 62(2):161–178.
- Stephan Diederich, Max Janßen-Müller, Alfred Brendel, and Stefan Morana. 2019. Emulating Empathetic Behavior in Online Service Encounters with Sentiment-Adaptive Responses: Insights from an Experiment with a Conversational Agent. In *Proceedings of the Fortieth International Conference on Information Systems (ICIS)*, pages 1–17, Munich, Germany. Chapman and Hall/CRC.
- Ed Diener, Derrick Wirtz, and William Tov. 2010. New Measures of Well-Being: Flourishing and Positive and Negative Feelings. *Social Indicators Research*, 39:247–266.
- Virginia Dignum. 2018. Ethics in Artificial Intelligence: Introduction to the Special Issue. *Ethics and Information Technology*, 20:1–3.
- Jasper Feine, Ulrich Gnewuch, Stefan Morana, and Alexander Maedche. 2019a. A Taxonomy of Social Cues for Conversational Agents. *International Journal of Human-Computer Studies*, 132:138–161.
- Jasper Feine, Ulrich Gnewuch, Stefan Morana, and Alexander Maedche. 2019b. Gender bias in chatbot design. In *International Workshop on Chatbot Research and Design*, pages 79–93. Springer.
- Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. 2017. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Mental Health*, 4(2):e19.
- Albert Gatt and Emiel Krahmer. 2018. Survey of the State of the Art in Natural Language Generation: Core Tasks, Applications and Evaluation. *Journal of Artificial Intelligence Research*, 61(1):65–170.
- Alastair Gill, Carsten Brockmann, and Jon Oberlander. 2012. Perceptions of alignment and personality in generated dialogue. In *INLG 2012 Proceedings of the Seventh International Natural Language Generation Conference*, pages 40–48.
- Marc Hassenzahl. 2018. The thing and i: understanding the relationship between user and product. In *Funology 2*, pages 301–313. Springer.
- Marc Hassenzahl and Noam Tractinsky. 2006. User experience—a research agenda. *Behaviour & information technology*, 25(2):91–97.
- Hsiu-Fang Hsieh and Sarah E Shannon. 2005. Three approaches to qualitative content analysis. *Qualitative health research*, 15(9):1277–1288.

- Amy Huang, Francis Lee, Clifford Nass, Young Paik, and Luke Swartz. 2001. Can Voice User Interfaces Say “I”? An Experiment with Recorded Speech and TTS. *Unpublished Manuscript*.
- Institute of Electrical and Electronics Engineers. 2017. *Affective Computing*. In *Ethically Aligned Design: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems, Version 2*, pages 162–181. IEEE Standards Association.
- Hankyung Kim, Dong Yoon Koh, Gaeun Lee, Jung-Mi Park, and Youn-kyung Lim. 2019. Designing personalities of conversational agents. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–6.
- Kira Kretzschmar, Holly Tyroll, Gabriela Pavarini, Arianna Manzini, and Ilina Singh. 2019. Can Your Phone Be Your Therapist? Young People’s Ethical Perspectives on the Use of Fully Automated Conversational Agents (Chatbots) in Mental Health Support. *Biomedical Informatics Insights*, 11:1–9.
- Sunok Lee, Sungbae Kim, and Sangsu Lee. 2019. “what does your agent look like?” a drawing study to understand users’ perceived persona of conversational agent. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–6.
- Chia-Yu Li, Daniel Ortega, Dirk V ath, Florian Lux, Lindsey Vanderlyn, Maximilian Schmidt, Michael Neumann, Moritz V olkel, Pavel Denisov, Sabrina Jenne, Zorica Kacarevic, and Ngoc Thang Vu. 2020. *ADVISER: A toolkit for developing multi-modal, multi-domain and socially-engaged conversational agents*. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 279–286, Online. Association for Computational Linguistics.
- Marian McDonnell and David Baxter. 2019. Chatbots and gender stereotyping. *Interacting with Computers*, 31(2):116–121.
- Clifford Nass, Youngme Moon, Brian J Fogg, Byron Reeves, and Chris Dryer. 1995. Can computer personalities be human personalities? In *Conference companion on Human factors in computing systems*, pages 228–229.
- Clifford Nass, Jonathan Steuer, and Ellen R Tauber. 1994. Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 72–78.
- Daniel Ortega, Dirk V ath, Gianna Weber, Lindsey Vanderlyn, Maximilian Schmidt, Moritz V olkel, Zorica Karacevic, and Ngoc Thang Vu. 2019. *ADVISER: A Dialog System Framework for Education & Research*. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 93–98, Florence, Italy. Association for Computational Linguistics.
- Manuel Portela and Carlos Granell-Canut. 2017. A new friend in our smartphone? observing interactions with chatbots in the search of emotional engagement. In *Proceedings of the XVIII International Conference on Human Computer Interaction*, pages 1–7.
- Jeng-Yi Tzeng. 2004. Toward a more Civilized Design: Studying the Effects of Computers that Apologize. *International Journal of Human-Computer Studies*, 61(3):319–345.
- Tibert Verhagen, Jaap van Nes, Frans Feldberg, and Willemijn van Dolen. 2014. Virtual Customer Service Agents: Using Social Presence and Personalization to Shape Online Service Encounters. *Journal of Computer-Mediated Communication*, 19(3):529–545.
- Mark West, Rebecca Kraut, and Han Ei Chew. 2019. I’d blush if i could: closing gender divides in digital skills through education.
- Nancy Wuenderlich and Stefanie Paluch. 2017. A nice and friendly chat with a bot: User perceptions of ai-based service agents. In *Proceedings of the 38th International Conference on Information Systems (ICIS 2017)*, pages 1–10.
- Xi Yang, Marco Aurisicchio, and Weston Baxter. 2019. Understanding affective experiences with conversational agents. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–12.
- Hao Zhou, Minlie Huang, Tianyang Zhang, Xiaoyan Zhu, and Bing Liu. 2018. Emotional Chatting Machine: Emotional Conversation Generation with Internal and External Memory. In *The Thirty Second AAAI Conference on Artificial Intelligence*, pages 730–738.

A Post-dialog questionnaire

Please read and answer all of the questions below. Once all questions have been answered, you may hit submit to generate your completion code for Mechanical Turk.

Demographic Information

Please indicate the gender you most identify with.

- Male
- Female
- Other

Please indicate your age:

- Less than 20
- 21-30
- 31-40
- 41-50
- 51-60
- 61-70
- 71-80
- More than 81

Please indicate which country you have lived in the longest:

Please indicate your native language(s):

Survey Questions

1. Impressions

1.1 Please rate your impression of the chatbot's HUMANNESS on the following scale:

| | 1 | 2 | 3 | 4 | 5 | |
|--------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|------------|
| Fake | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Natural |
| Machine-like | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Human-like |
| Unconscious | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Conscious |
| Artificial | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Lifelike |

1.2 Please rate your impression of the chatbot's LIKABILITY on the following scale:

| | 1 | 2 | 3 | 4 | 5 | |
|------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|----------|
| Dislike | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Like |
| Unfriendly | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Friendly |
| Unkind | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Kind |
| Unpleasant | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Pleasant |
| Awful | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Nice |

1.3 Please rate your impression of the chatbot’s INTELLIGENCE on the following scale:

| | 1 | 2 | 3 | 4 | 5 | |
|---------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|---------------|
| Incompetent | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Competent |
| Ignorant | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Knowledgeable |
| Irresponsible | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Responsible |
| Unintelligent | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Intelligent |
| Foolish | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Sensible |

1.4 Please describe anything from your interaction that influenced how you perceived the chatbot. You may also provide any additional impressions you had which were not already covered in the above questions.

2. Linguistic Style

2.1 Please rate how you perceived the chatbot’s LANGUAGE STYLE on the following scale:

| | 1 | 2 | 3 | 4 | 5 | |
|-------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|-------------|
| Disruptive | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Calming |
| Exaggerated | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Appropriate |
| Unrealistic | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Realistic |

3. Personal Pronouns

3.1 Please rate how you perceived the chatbot’s USE (OR LACK OF USE) OF THE PRONOUN “I” on the following scale:

| | 1 | 2 | 3 | 4 | 5 | |
|-------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|---------------|
| Eerie | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Natural |
| Dispensable | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Indispensable |
| Unsuitable | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Suitable |

3.2 Please provide a short explanation of your choices from the previous question.

3.3 Please rate how you perceived the chatbot’s USE (OR LACK OF USE) OF THE PRONOUN “YOU” on the following scale:

| | 1 | 2 | 3 | 4 | 5 | |
|-------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|---------------|
| Eerie | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Natural |
| Dispensable | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Indispensable |
| Unsuitable | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Suitable |

3.4 Please provide a short explanation of your choices from the previous question.

4. Gender

4.1 If you had to assign a gender to the chatbot, what would it be?

Male Female Genderless

4.2 Please provide a short explanation of your choice.

5. Feeling during the interaction

5.1 Please reflect on your conversation with the chatbot and rate HOW MUCH YOU EXPERIENCED EACH FEELING on the following scale from 1 (NOT AT ALL) to 5 (A LOT)

| | 1 | 2 | 3 | 4 | 5 |
|------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| Positive | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Good | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Pleasant | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Happy | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Joyful | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Contented | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Negative | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Bad | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Unpleasant | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Sad | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Afraid | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Angry | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

6. Trust

6.1 Please indicate how much you agree with the following statements:

| | Strongly disagree | Rather disagree | Neither agree nor disagree | Rather agree | Strongly agree |
|--|--------------------------|--------------------------|----------------------------|--------------------------|--------------------------|
| I already know similar systems. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| I rather trust an automated system than mistrust it. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| One should be careful with unfamiliar automated systems. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| The developers take my well-being seriously. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| I trust the system. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| The developers are trustworthy. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| I can rely on the system. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| I have already used similar systems. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Automated systems generally work well. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

7. Satisfaction

7.1 Please indicate how satisfied you were with the chatbot in terms of the OVERALL INTERACTION:

Very dissatisfied | | Completely satisfied

7.2 Please indicate how satisfied you were with the chatbot in terms of the TASK-COMPLETION:

Very dissatisfied | | Completely satisfied

7.3 What things did you find positive about the interaction.

7.4 What things did you find negative about the interaction.

B Content Analysis

Table 2: Exemplary codes from the content analysis. For every code, an example of participants' free-form feedback is given. Every response is from another participant. The code label and the CA variant (assigned gender for category D) are given in parentheses.

| | |
|----|--|
| A1 | If you hadn't told me it was a chatbot, I would have thought that someone was on the other end, answering my questions. (human-like, ESTJ) |
| A2 | It was very straight forward, very cut and dry information. (matter-of-fact, passive) |
| A3 | The chatbot was very thorough in their explanations and were also very friendly and upbeat. (friendly, ESFJ) |
| A4 | It did not have any empathy for my deadlines being missed. (cold, ESTJ) The bot was purely mechanical and only regurgitated information back at me. It felt cold and unfriendly. (cold, passive) |
| A5 | I thought that some of the responses the chatbot gave were a little passive aggressive. That made me view the chatbot as being a little unfriendly. (mean, ESTJ) |
| A6 | I was impressed at the speed of the bot. It was very quick to answer my questions. No waiting and it knew exactly how to answer. (fast responses, ESFJ) |
| B1 | I didn't really notice how the chatbot used pronouns, so it must have been natural-sounding. (unnoticed, ESTJ) |
| B2 | It seemed normal and appropriate to me. Nothing was odd or eerie. (normal/appropriate, passive) I didn't even give it a second thought while chatting with the bot. Seems natural to me! (normal/appropriate, ESTJ) |
| B3 | The bot communicated like a real person would. The 'I' pronoun is nice in these scenarios because I think it makes the bot sound more human (human-like, ESFJ) |
| B4 | It's not human, so it isn't necessary. (not needed, ESTJ) |
| C1 | I didn't notice anything out of the ordinary (unnoticed, passive) |
| C2 | I also didn't notice the chatbot not saying 'you'. I felt like it was talking like a machine would talk. it seemed normal. (normal/appropriate, passive) 'You' is perfectly natural for the chatbot. (normal/appropriate, ESFJ) |
| C3 | It sounded just like any other human using the word 'you'. I can't draw a distinction between human and chatbot in this regard. (human-like, ESTJ) |
| C4 | Using you made me feel like it cared not just a plain example of what I was asking. (personal, ESFJ) It felt like the chatbot was talking directly to me and made me feel seen. (personal, ESTJ) |
| D1 | I have never associated a gender with a chatbot. (genderless bot, genderless) |
| D2 | I just got a male feeling from the chatbot I have no real explanation why (intuition, male) |
| D3 | I think males are a bit more harsh in chats than females. (rude, male) |
| D4 | It nagged a bit, which I associate with women. (rude, female) |
| D5 | The chatbot was not emotional which is what males are stereo typically thought of as. (cold/not emotional, male) I usually think of females being more enthusiastic and friendly. (friendly, female) |
| E1 | Most questions were easily, thoroughly and competently answered. (good answers, ESFJ) |
| E2 | I appreciated the detail answers the chatbot gave without a lot of unnecessary information. (efficient, passive) |
| E3 | It answered my questions as soon as I typed them. There was no lag. (fast answers, ESTJ) |
| F1 | It seemed to answer based on keywords, not a real understanding of what I was asking. (limited understanding, ESTJ) |
| F2 | There was one somewhat rude response to me about not paying attention to deadlines. (rude responses, ESTJ) |
| F3 | It felt very mechanical and cold. I felt like I was talking to a robot and not a human. (cold, passive) |

A = General impressions, B = Perception of pronoun 'I', C = Perception of pronoun 'you'

D = Perception of gender, E = Positive impressions, F = Negative impressions