Designing for explanation-driven trust in chatbots

Abstract

Trust plays an important role in the implementation of chatbot technology. This study was also focusing on the user trust in chatbots, particularly focusing on the role of response delay and explanation driven subjective transparency. This research includes a pretest and a main test. In the pretest, we selected one explanation that was perceived by the participants that can raise the most social presence feeling as well as the subjective transparency of the chatbot. In the main test, a 2 × 2 between-subject experiment was designed and conducted to test the hypotheses. First, the findings revealed that while response delay did not significantly influence trust or social presence, clear explanations, especially in the context of instant delays, positively impacted subjective transparency and trust. Second, the study reinforced the positive correlation between social presence and trust, subjective transparency and trust. From a practical perspective, the research offers insights for chatbot design, emphasizing the importance of improving subjective transparency, and rendering a more natural and human-like interaction.

Research question and hypotheses

Research questions:

RQ1: How does the response delay influence the user's trust in chatbot?

RQ2: How does the explanation of the response delay influence the user's trust in the chatbot?

As shown in Figure 1, corresponding reaserch model was proposed based on previous reaserches.

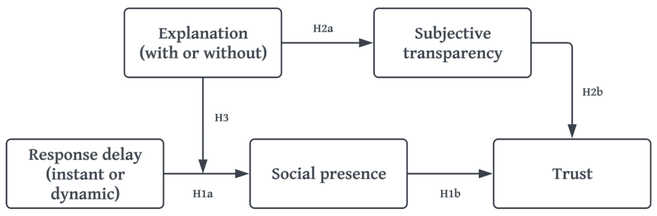

Figure 1. Reaserch model.

H1: User social presence mediates the positive influence between response delay and user trust in the chatbot.

H1a: The response delay of the chatbot positively influences the user's social presence while interacting with the chatbot.

H1b: User social presence positively influences the user's trust in the chatbot.

H2: Explanation of the response delay has a positive effect on user trust in the chatbot.

H2a: Explanation of the response delay has a positive effect on user-perceived transparency of the chatbot system.

H2b: User-perceived transparency of the chatbot system has a positive effect on users' trust in the chatbot.

H3: Explanation of the chatbot response delay can moderate the effect between response delay and social presence.

Pretest

Existing literature does not provide definitive evidence or consensus regarding which explanations are most beneficial or effective for chatbots. Consequently, our pre-test is designed to identifying an explanation that can enhances subjective transparency and user social presence. The chatbot that was used in the pretest is shown in figure 2.

Participants

10 participants (5 male, 5 female) were recruited with no compensation for the pre-test. They were all recruited offline in the Industrial design engineering faculty of TU Delft. Participants are all master students at the Industrial design engineering faculty of TU Delft, aged from 23 to 26, speaking English as a second language, and having experience with chatbots, especially with daily usage of chatGPT in the recent month.

Procedural

The pretest followed a within-subject experiment procedural. During the test, a text introduction was first shown to the participants and experimenter was there to help them understand what they needed to do. Participants were asked to interact with all chatbot settings. The chatbot setting included 6 different explanation conditions (none, basic, first-person, detailed, first-person + detailed, and humor) and 2 different delay conditions (instant and dynamic). So during the pretest, each participant was asked to interact with chatbots in 12 (6 * 2) different conditions. After finishing the interaction, the definition of social presence and subjective transparency was explained to the participants. Then, all explanations were presented to the participant, and the participant was asked to rank the explanations based on the degree they think the explanation fits with social presence and perceived transparency. In the end, a short interview was conducted to collect further feedback from the participants regarding understandability, clarity, etc,.

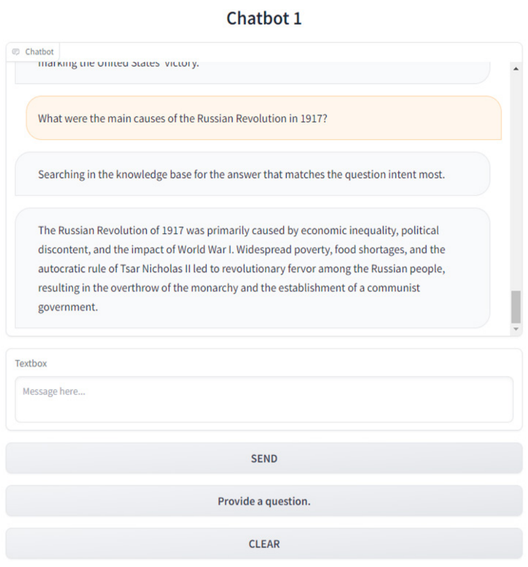

Figure 2. Pre-test Chatbot prototype.

Key takeaways

Key take aways from the pre-test include:

(1) The first-person + detailed explanation emerged as the most effective in enhancing both social presence and subjective transparency. **(2)** The specific words used in the explanation need some optimization to increase the understandability of the explanation. For instance, terms like "pre-trained model" were found to be potentially confusing and even distancing for some users. **(3)** The dynamic response delay setting was perceived to be too long by the majority of the participants and already influenced user perception of the capability of the chatbot. **(4)** Participants expressed a need for clearer demarcation between the explanation and the actual chatbot response. Given these findings and the feedback received, several adjustments were made to optimize the chatbot prototype and its interactions. These modifications, ranging from the content of explanations to the calculation of response delays, were aimed at ensuring the validity of the subsequent main study.

Main test

To test our hypotheses, we conducted a two-factor between-subjects experiment. The independent variables: response delay (dynamic or instant) and explanation (with or without), were manipulated in the experiment, hence 2*2 in total 4 groups were set for the experiment: instant/dynamic delay with/without explanation.

Participants

The experiment was conducted via Prolific (https://www.prolific.co/) – an online recruitment platform. In total, 201 subjects participated in the study. Participants in Prolific are paid in GBP (£) and studies are required to pay a minimum amount that is equivalent to USD ($) 6.50 per hour. This study followed a between-subject procedural. Four participants got excluded, because they answered the English language level check question as advanced (3 participants) or intermedia (1 participant). The final sample included 197 participants (94 male, 93 female, 7 Non-binary/third gender, and 3 prefer not to disclose/self-describe). Participants were between 19 and 54 years old (M = 34.42, SD = 10.30)

Procedural

First, participants need to read through the same introduction about the study and provide informed consent to joining the study. Participants were informed that this is a study related to the interaction design of an astronomy chatbot. But the details about what will be tested or evaluated was not disclosed to them. Then the participants were asked to follow the video carefully and randomly assigned to one of the four conditions. The randomization process was done automatically by Qualtrics (https://www.qualtrics.com/) - an online survey builder. For every condition, participants need to answer the same questionnaires afterward. Considering the cognitive load of the participant increased as the study went on, we decided to show the video stimuli and measure the factors in our research model first. Hence, the questionnaire was shown to the participants by order: perceived delay, social presence, subjective transparency, trust, chatbot experience, astronomy knowledge, and AIT scale. In the end, several demographic questions were asked to the participants. Everything was anonymized in this study. At the end of the questionnaire, we gave participants a chance to comment on this research. Both the four stimuli were evenly distributed among participants.

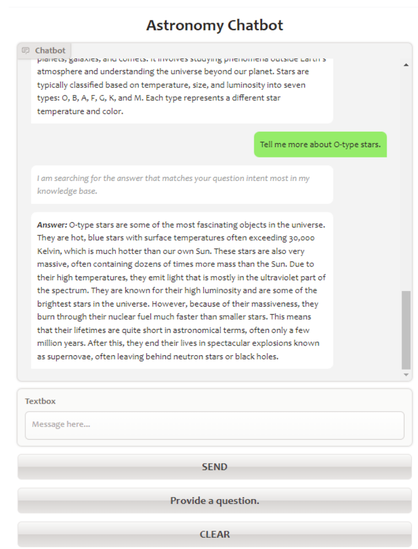

Figure 3. Main test Chatbot prototype.

Hypothesis testing

Social Presence (H1a & H3)

A Mann-Whitney U test was used to test the association between response delay and social presence. The test indicated no statistically significant difference between the two groups, U = 4834.500, Z = -.039, p = .969. This suggests that the delay type did not have a significant effect on subjective transparency. Hence, H1a was not supported. The two-way Bootstrap ANOVA conducted to investigate the potential moderating effect of the experimental type on the relationship between delay type and social presence showed no significant interaction (F(1, 193) = .234, p = .629). Hence, H3 was not supported.

Subjective transparency (H2a)

The test revealed a statistically significant difference between the two groups (U = 4029.000, Z = -2.054, p = .040). This suggests that the explanation type has a significant effect on subjective transparency, with the group receiving an explanation tending to rank higher in terms of transparency compared to the group without an explanation. Hence, H2a was supported.

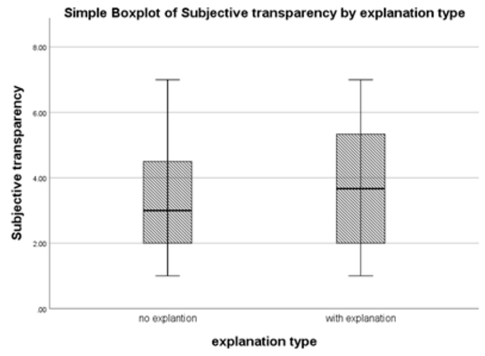

Figure 4. Boxplot: Subjective transparency level in different explanation types

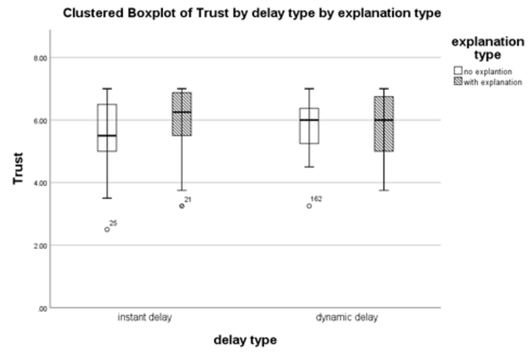

Figure 5. Boxplot: Trust level in different delay and explanation conditions

Trust towards chatbot (H1b, H2b)

For the relationship between social presence and trust, a significant positive correlation was found (Spearman's rho(197) = .165, p = .021). This suggests that higher levels of social presence are associated with higher levels of trust in chatbot interactions. For the relationship between trust and subjective transparency and trust, a significant positive correlation was identified (Spearman's rho(197) = .247, p < .001). In both analyses, p-values were less than .05, indicating the correlations are statistically significant. Hence, H1b and H2b were both supported.

Mann-Whitney U tests were conducted to evaluate the differences in trust between the two delay types, and two explanation-types. The test revealed no statistically significant difference between the two delay groups (U = 4845.500, Z = -.011, p = .991) and no statistically significant difference between the two explanation groups (U = 4212.000, Z = -1.602, p = .109).

A two-way Bootstrap ANOVA was performed to investigate the potential moderating effect of the delay type on the relationship between explanation type and trust. The interaction between delay type and explanation type was marginally significant, F(1, 193) = 3.106, p = .080. Mann-Whitney U tests were conducted for each combination of the delay type and explanation type. Under the instant delay condition, the Mann-Whitney U test revealed a statistically significant difference between different explanation groups (U = 945.500, Z = -2.113, p = .035). This reveals that providing an explanation in the instant delay condition significantly increased trust (M = 5.995, SD = .992) compared to not providing an explanation (M = 5.536, SD = 1.128).

Implications

1. This study replicated the positive correlation between subjective transparency and user trust in the knowledge chatbot domain. The result also suggested that providing explanations about how the chatbot is functioning can enhance the perception of transparency. For designers, giving users insight into how the chatbot functions can build a sense of transparency and likely lead to more trust towards the chatbot.
2. This study replicated the positive correlation between social presence and user trust in the knowledge chatbot domain. This indicates that designers can also try to enhance the sense of social presence in chatbot interactions.
3. This study revealed a marginally significant effect between explanation and response delay type on trust. For designers, this could mean that when designing for different types of knowledge chatbots (retrieving- or generating-based), they should implement different explanation strategies for a trustworthy chatbot.

Zhengquan Zhang
Designing for explanation-driven trust in chatbots
15-08-2023
Strategic Product Design

Committee Dr. Schneegass, C.
Dr. Tsiakas, K.

Faculty of Industrial Design Engineering

Delft University of Technology