**Putting Chatbot Interactions on the Right TrAKX: Examining Whether Slight Delays in Conversation Increase a User's Trust in Recommendation Systems**

- Khellaf, Tarek (t.khellaf@students.uu.nl)

- Daggelinckx, Anneline (a.daggelinckx@students.uu.nl)

- Zhang, Xinyu (x.zhang16@students.uu.nl)

- Schoneveld, G, Kuil (k.g.schoneveld@students.uu.nl)

# Research **question**

Will a short delay in a chatbot's responses make the user trust the chatbot and it’s decisions more? More specifically, do slight delays in text conversations increase the trust a user places in their interlocutor? The number of user-chatbot interactions have increased dramatically as businesses include such systems on their websites. In accordance with intuition, the growing literature on the humanlike perceivability of customer service chatbots shows one’s trust in a chatbot is an innate part of an improved user experience. This trust can be exemplified in two senses; trusting the bot to function properly and entrusting it with sensitive information. For example, the paper *Faster Is Not Always Better: Understanding the Effect of Dynamic Response Delays in Human-Chatbot Interaction*, found that “dynamic response delays can not only increase users’ perception of humanness and social presence, but also lead to greater satisfaction with the overall chatbot interaction”. We will therefore attempt to replicate this study’s results to test whether the correlation holds between the first sense of trustworthiness (reliable function) and a response delay in our chatbot.

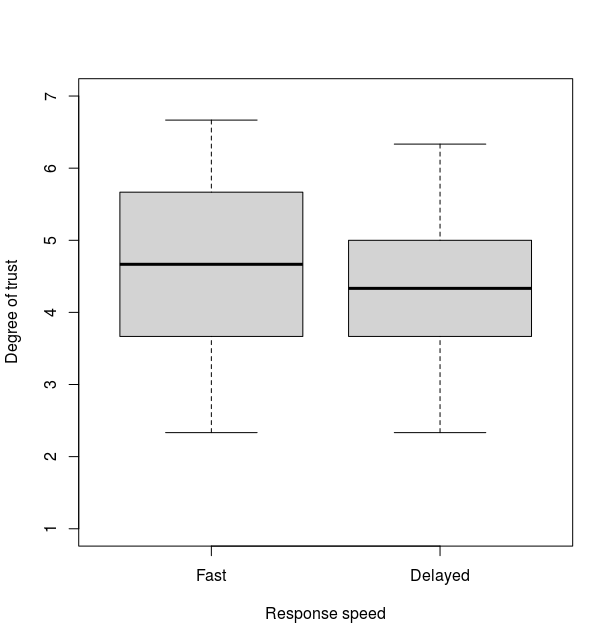
1. **General progress**

First, we set up a website hosting our chatbot implementation. The website is structured such that it first shows the participant general instructions. Then, one of the two chatbots is randomly selected to be shown first. After the users spend two rounds interacting with the bot by fulfilling a predefined query, the user is shown a survey by Google Forms. The same procedure is then repeated with the remaining bot. Finally, users fill in a final survey with general questions. In order to ensure that users do indeed interact with the bot for a sufficient amount of time, we ask them to enter the phone number of a specific restaurant into a field which enables the “next” button. However, we do not require a correct answer key, but rather let them continue no matter what they type in. To ensure that the participants fill in the surveys, we provided a numeric code that is displayed after submitting the survey which lets them continue on to the next page.

So far, we have collected more than half of our data (24 participants). Our goal is to have at least 30, but we await responses from many potential participants. When the data collection is finished, the result can be produced easily for analysis and discussion.

We have begun writing a first draft of our introduction and theoretical background. The other sections consist only of bullet points of important discussion points, but these still need to be made into full text.

**3. Graph and analysis**



**Figure 1: TrAKXing our participants’ distributions of trust across response speeds**

The figure above shows boxplots of the data distribution. A score of 1 means the participant didn’t trust the bot at all, while a score of 7 indicates that the participant had a lot of trust in the bot.

These results have already begun to show that the average user’s trust is slightly lower for the delayed answers than for the fast answers. This is the opposite of what we expected. A one-sided, paired t-test showed no significant raise in trust by delaying the responses of the chatbot, *t(13) = 0.58662, p = 0.7184*. One reason for this result could be the currently uneven distribution of groups. Currently 15 out of 24 participants were shown the delayed chatbot first compared to 9 out of 24 for the fast bot. It is likely that participants generally prefer the second bot as they have inadvertently adapted to the conversation style from their first interaction. Further analysis is necessary to clarify the strength of our conclusions. We may forcibly equalize the survey distribution of fast versus delayed responses by altering the code to ensure the delayed bot receives fair representation as the second exposure.

1. **Distribution**

This week, Tarek finished the test website, and we all tested it multiple times to try to catch and fix some problems. Kuil, Xinyu and Anneline finalized the questionnaires. Then we all invited participants and had them perform the experiment. We all started annotating the final report, and Kuil started writing the introduction section. In our opinion, everyone’s contribution is equal.

1. **Question**

* Our results aren’t significant, but we have taken knowledge of a few possible reasons for this (e.g. a lot of our participants didn’t notice the difference between our bots, or thought the delay was due to their internet connection and therefore ignored it). We were wondering whether we have to mention these things in the paper (and if so, we presume in the discussion part)?
* We have also collected data about participants’ age, experience of chatbot and gender so we can explore whether these characters have influence on people’s preference. Is this view worth displaying early in our final report? Or just show them in the discussion part? And does it matter that the age distribution is uneven?