Midterm report - second delivery

1.1 About us

We are a group of four students in the masterprogram informatics: design, use and interaction at University of Oslo. Three of the members have bachelor degrees from the University of Oslo, and the remaining member took their bachelor from Westerdals in Oslo.

Our group consists of the following members:

- Ashwin Rajeswaran: ashwinsr@uio.no
- Bjørn Langseth: bjorlan@uio.no
- Christina Grønhaug: chrgron@uio.no
- Maren Ingunnsdatter Berge: marenib@uio.no

1.2 Area of interest

An area of "interaction with AI" we are interested in working with is the psychological aspects of AI. In particular how users perceive AI-machines and how it affects human-AI interaction. We wish to research how this affects interaction with chatbots. Relevant literature for this study could be:

- **Verne & Bratteteig (2018):**
  - We want to look at this literature for their definition of AI in a human-AI perspective.
- **Verne (2020):**
  - This article is interesting to us because it shows how important the work situation is for an AI to automate well for users. We also want to have a look at how humans have to adapt to AI and vice versa.
- **Norman (1990):**
  - Can show the consequences of "bad" interaction, and shows how AI could need a dialogue to fulfil its purpose.
  - Over-automazation can become too "monologue" for the user and leave them "out of the loop" (Norman, 1990), and the users will have difficulties to grasp the underlying task chains that the system is performing.

1.3 Background section: Position your work relative to existing knowledge and practice

Our work is positioned relative to our academic literature but also previous experiences of our group. We have also based our report on our own hypothesis that people are unhappy with chatbots and we want to explore if there is some actual truth in our hypothesis and if it can be connected to existing literature.

We have through our experience learned that the main practice with engaging in contact with chatbots is to attempt to talk to real humans right away. This is illustrated from scenarios when people type keywords to get redirected to human customer-support. Are there any scenarios where the chatbot fulfils it purpose, or does the customer journey always end with human-human interaction? Regardless of what answers the chatbot provides, users may send a clarification mail or ask for contact by other means if they feel the question is too complex for a robot to answer.

Frustration with the use of chatbots is also prevalent, and it motivates us to explore any pain points that may cause the frustration to begin with. This frustration-aspect is an element we want to investigate also based on personal experience within the group.

1.4 Research questions

Based on the areas of interest we want to investigate:

- **Why people decide to *not* chat with chatbots?**
  - Our presumptions:
    - The chatbot needs to imply the level of detail needed to communicate. How basic commands can limit the usefulness in solving complex questions or problems from the users.
    - People usually do not know how chatbots actually work; it looks for keywords, and does not analyze entire text.
- **What deciding factors make chatbots successful?**
  - **In other words:** *What defines a successful chatbot and what does not?*
  - We wish to investigate what can be changed to improve chatbots.
  - Investigate through the chatbot we will construct in module 2.

1.5 Methods - overall approach, design process (optional, but encouraged), data collection methods

Initially, we planned to use four data collection methods to gain answers to our hypotheses. However, after discussing within the group and basing on relevant literature, we have concluded with only partaking in literature analysis and observation. We believe these methods will provide the most insight in our short project lifetime.

Literature analysis

- We will examine the articles mentioned above during the project to understand the different perspectives of a human-AI interaction. We also want to see factors that may lead AIs to 'fail' in their use in their context.
- The goal of using this method is to gather data based on older projects to support or, hopefully, confirm our hypothesis. We already have a quantity of existing data, and we believe it is crucial to elaborate earlier findings to guide our own project.

Observation

- We want to experience real chatbots and how users interact with them. We can compare the data from the different chatbots and see how they tackle the same problem.
- The goal of using this method is to get unique insight about the real user experience, and to witness frustrations that may emerge in real life situations. This may provide additional support for our hypothesis.
- For the time being, we do not prioritise this method, as the literature analysis will be shaping our core knowledge. We will follow up with this method in the next iteration.

The other methods we had planned were:

- **Wizard of Oz / thinking aloud**
  - We can use this as a method to see how it is experienced. What can we do to make the interaction better? The method will provide us information about what is "good" and "bad" responses.
- **Interviews**
  - At the beginning of the project we do not see the value of conducting interviews, but we may need it later in the project to clarify information.

These methods are discontinued due to the fact that they do not correspond with our current goal for the project. They are great data collection methods, but for this project we have decided to build upon the already existing data from previous studies.

1.7 Findings (progress, initial outcomes)

We decided to read literature about customer service chatbots to better understand any key characteristics that may point to bad user experience. In this section, we have summarized the main points of the literature and takeaways that are relevant for our project.

Literature 1: Kvale et al. (2020)

This paper investigates what users perceive as a good customer service chatbot. The findings suggest that the users' satisfaction with a service-oriented chatbot is correlated to how well the chatbot can solve the users' problems. As customer service chatbots' focal goal is to be "(...) low threshold channels for information and support (...)", the users' motivation to engage in a dialogue with chatbots is primarily that the service is cost-effective and accessible any given time of the day, as opposed to manual customer service.

The study bases the research questions upon existing literature, indicating that 'low customer satisfaction scores may indicate poor user experience'. Their main concern is whether or not the users' problems are solved, what kind of problem and what kind of intents have positive or negative association (Kvale et al., 2020).

The findings suggest that the problem itself is a substantial factor for customer satisfaction with the chatbot. An important point of the study is that customers were generally satisfied with chatbots when it solved issues that were 'specific and concrete with a simple-to-understand answer that solves the problem' (Kvale et al., 2020). Intents that are associated with negative feedback are, according to the study, intents that are aiming to deal with too many situations. This is especially true for situations where the users lack understanding of the problem's nature. Therefore they do not receive specific answers, but in turn receive information in a loop.

Literature 2: Brandtzaeg and Foelstad (2017)

Brandtzæg and Følstad argue that there is a lack of studies on why people use chatbots (Brandtzæg & Foelstad, 2017, p. 377). In their article they discuss different aspects of motivation that might be the reason why people use chatbots. Based on our area of interest, why people choose not to use chatbots, we think the aspect of motivation gives us an interesting perspective on the subject. We argue that if we explore why people use chatbots, this might give us a pointer on what user needs need to be addressed, and if there is a lack of these needs it might result in people not using chatbots.

An example of a motivation is "chatbots for Productivity" and in the study described in the article, this was the main aspect of why people use chatbots (Brandtzæg & Foelstad, 2017, 383-384). Furthered explained this consisted of the users need for "ease, speed and convenience", and that chatbots fulfill this purpose. Based on our view on the subject we argue that if the chatbot doesn't meet these demands, that might be a reason why people don't use chatbots. In appendix two, we experienced some delays in the conversation with the chatbot, which resulted in a lot of frustration on us as users.

We see that chatbots usually have to fulfill a goal based on its use context, and for customer service chatbots, this purpose is solving the users' problems. Users change their behaviour to reduce misconceptions (Brandtzaeg & Foelstad, 2017, p. 379-380), and by introducing a natural language in a service oriented context (which is supposedly formal and concrete) it may result in more usage as it may reduce a barrier to communication.

Literature 3: Brandtzaeg & Foelstad. (2021)

The article by Brandtzæg & Foelstad is based on research on social support through chatbots aimed at young people. The research question that is being explored is "How do young people perceive different types of social support in chatbots, and what are the social implications associated with such chatbot use?" (Brandtzaeg & Foelstad. 2021, p.1).

The main findings from the article are that young people use it because it is easily accessible and it ensures the young people's anonymity. They can easily talk about issues that are challenging and open up about it. In connection with our problem, the problem of actual use and how it is used is not necessarily relevant, but the findings are interesting. As we see it, these can be duplicated to

why people actually use chatbots in the first place - they are available and can ensure your anonymity.

Furthermore, there is another finding in the article we find very interesting and that is how users do not take in issues around chatbots and sharing sensitive personal data and that this can be problematic. As mentioned in the article, chatbots can pose a risk in the form that they give biases, insufficient and unsuccessful answers, as well as give privacy problems (Brandtzaeg & Foelstad. 2021, p.2). These are elements we believe are relevant regardless of who the chatbot is designed for and what use it is intended for.

Literature 4: Brandtzaeg & Foelstad. (2018)

One of the main arguments of the article by Brandtzæg and Følstad from 2018, is that the chatbots could end up becoming too human. A consequence could be that people try to exploit the chatbot's knowledge for other means than intended. One example from the article is the chatbot Tay,developed by Microsoft. The chatbot was subject to deep learning from any user who had access to it, and was exposed to extremist users on the internet. Once the chatbot got exposed to extremist values and profanity language, the chatbot also learned from this behaviour and displayed this to any user who tried to communicate, creating hostile environments for all the users.

Another point made in the paper that could relate to our case is the concern of the chatbot becoming too humane. If this happens, the robot will be treated in a different way than intended by the user. We see this as relevant for customer service chatbots also, as a too humane chatbot may disorient the user as the language will become less specific and harder to interpret, which defeats the purpose of customer service chatbots to begin with.

Our takeaways from the literature

The "ease, speed and convenience" of chatbots is a common theme in the articles, and based on this we will argue that this is one of the main arguments why people use chatbots. We see this as a focal point in all of the papers; that the main motivation is quick and convenient communication to solve their problems. Another motivation is that they are available 'all-the-time' for users.

Kvale et al. (2020) points to actors who have experienced not receiving the help they need, and thus become frustrated and ask to receive manual customer service for a problem that *they* perceive as

an easy-to-solve problem. Our chatbot, Bottern, also shows this interaction, where the user asks for quick, concrete and easy solutions to their problems but is instead thrown in an information loop.

We see that the availability is the starting motivation for users, but the coming response is what builds trust to the system later on, and is the 'make it or break it' for chatbots.

Two of the reviewed papers are also alluding to natural language as a factor for the users behavior. The users treat it as a human based on the language, but this also affects the customer service aspect because it could make the content more understandable for the user. Brandtzaeg & Følestad (2017) argue that many chatbots are based on natural language, and we see it as a tool for relatability and forming a relation between the trademark service and the user. One should still be careful of the language model material that is provided, however, to avoid non-specific and non-concrete language, and hostile language alluding to extremist values and profanity language.

In summary, our thoughts about motivations and needs for people who interact with chatbots are mainly based on availability in the initial state. This is to quickly get a response or solution to their problems to fix in-promptu. This means that how the chatbot answers and provides a specific solution becomes the breaking point in if the user is satisfied or dissatisfied with the service. There are therefore many personal motivations the chatbot has to satisfy. This is tied with the language of the chatbot, as when the provided answer by the chatbot is non-specific, it creates confusion and dissatisfaction for the user as they may not have received the help they anticipated to receive.

Feedback and work from iteration 1

Our wonder document was received very well by the other group. They wished that we further explain and narrow down on the methods we are going to use in this project, which we now have done extensively for the literature analysis. We did not see this as necessary for the other methods because we are not focusing on using those for the time being.

1.8 References and possible literature

Mandatory literature:

- Bratteteig, T., & G. Verne. (2018). Does AI make PD obsolete? Exploring Challenges from Artificial Intelligence to Participatory Design. Proceedings of PDC 2018, Belgium, 1-5. https://doi.org/10.1145/3210604.3210646
- Norman, D (1990). The problem of automation: Inappropirate feedback and interaction, not over-automation. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, Vol. 327, No. 1241, Human Factors in Hazardous Situations (Apr. 12, 1990), pp. 585-593 (9 pages)

Supplementary literature:

- Guri B. Verne. 2020. Adapting to a Robot: Adapting Gardening and the Garden to fit a Robot Lawn Mower. *In Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (*HRI '20*). Association for Computing Machinery, New York, NY, USA, 34–42. DOI:https://doi.org/10.1145/3371382.3380738
- Brandtzaeg, P., & Foelstad, A. (2018). Chatbots. Interactions (New York, N.Y.), 25(5), 38–43. https://doi.org/10.1145/3236669
- Brandtzaeg, P. B., & Foelstad, A. (2017). Why People Use Chatbots. Internet Science, 377–392. https://doi.org/10.1007/978-3-319-70284-1_30
- Brandtzæg, Skjuve, Dysthe, & Foelstad. (2021). When the Social Becomes Non-Human: Young People's Perception of Social Support in Chatbots. Association for Computing Machinery (ACM).
- Kvale, K., Freddi, E., Hodnebrog, S., Sell, O. A., & Følstad, A. (2021). Understanding the User Experience of Customer Service Chatbots: What Can We Learn from Customer Satisfaction Surveys? In Chatbot Research and Design (pp. 205–218). Springer International Publishing. https://doi.org/10.1007/978-3-030-68288-0_14