



UiO : **Institutt for informatikk**  
Det matematisk-naturvitenskapelige fakultet

# **IN5480 - Interaction with AI**

Specialization in Research in Design of IT

Final report - Fall 2021



<b>Table of content</b>	<b>2</b>
<b>1.1 About us</b>	<b>3</b>
<b>1.2 Area of interest</b>	<b>3</b>
<b>1.3 Background section: existing knowledge and practice</b>	<b>4</b>
<b>1.4 Research questions</b>	<b>5</b>
<b>1.5 Methods - overall approach, design process</b>	<b>5</b>
<b>1.7 Findings (progress, initial outcomes)</b>	<b>7</b>
How did we do our analysis?	7
Literature 1: Kvale et al. (2021)	7
Literature 2: Brandtzaeg & Foelstad (2017)	8
Literature 3: Brandtzaeg & Foelstad (2021)	9
Literature 4: Brandtzaeg & Foelstad (2018)	9
Literature 5: Mohit Jain, et al. (2018)	10
Our takeaways from the literature	11
<b>Critique of literature</b>	<b>12</b>
<b>Results from the observations</b>	<b>12</b>
Observation of DNB's chatbot	13
Observation of NAV's chatbot	14
Observation of Morgenlevering chatbot	15
Observation of Posten's chatbot	16
<b>What we have learned from the observations</b>	<b>16</b>
<b>Discussion</b>	<b>18</b>
<b>Conclusion</b>	<b>19</b>
<b>Feedback and work</b>	<b>20</b>
Iteration 1	20
Iteration 2	20
<b>1.8 References</b>	<b>21</b>

## 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 degrees from Westerdals in Oslo.

Our group consists of the following members:

- Ashwin Rajeswaran: [ashwinsr@uio.no](mailto:ashwinsr@uio.no)
- Bjørn Langseth: [bjorlan@uio.no](mailto:bjorlan@uio.no)
- Christina Grønhaug: [chrgron@uio.no](mailto:chrgron@uio.no)
- Maren Ingunnsdatter Berge: [marenib@uio.no](mailto: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-automatization 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.

**Update from iteration 3:** we decided that the mentioned literature was not sufficient to answer our research hypotheses when we started conducting our literature analyses. These articles provided a good starting point for our research group to grasp what interaction with AI can be and consequences of not providing enough feedback. The points made by Norman (1990) still hold value for our observations as it corresponds with guideline 11 by Amershi et al. (2019) in requiring the system to ‘make clear why the system did what it did’. However, as we later specified the research to orient around service chatbots this became less relevant for our study.

### **1.3 Background section: 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 its 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:

- **H1: Why do 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.
  
- **H2: 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

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.

### **How did we do our analysis?**

For the second iteration, the group decided to read articles that we received from Foelstad, but for the third iteration we also included articles which we found through Google Scholar. For the literature from Google Scholar, we were urged to find and investigate literature written by other authors so that we could receive diverse content and opinions from different researchers. Hence we also wished to find literature that touched upon the same topics, but hopefully had different insight from the literature provided by Foelstad. We decided to read one article each and write down interesting themes that may be of importance for our research. In other words, we used thematic coding to analyse our literature and grasp themes for each article. Following the individual analyses, the themes and examples were discussed by the group to establish common takeaways from all the literature, conclude with whether the data corresponded with our research hypothesis or not, and finally prepare for our observation which we received the opportunity to conduct during the third iteration.

### **Literature 1: Kvale et al. (2021)**

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., 2021).

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., 2021). 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 & Foelstad (2017)**

Brandtzaeg and Foelstad argue that there is a lack of studies on why people use chatbots (Brandtzaeg & 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 (Brandtzaeg & 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 Brandtzaeg & 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 Brandtzaeg and Foelstad 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.

### **Literature 5: Mohit Jain, et al. (2018)**

The article by Mohit Jain, et al (2018) is based on a usability evaluation the researchers conducted by Context View (Convey) in chatbots. Their starting point prior to this evaluation was several research articles where the users' interaction experience with chatbots has been evaluated. These have revealed that there is a large gap between experience and expectation of chatbots and their intelligence. It is often considered that the system does not have the mental model to be able to carry out a conversation of a more complex nature - this in turn makes users worried about using it (Mohit Jain et al., 2018, p.1).

Based on these research articles, the researchers conducted a user evaluation with Context View in chatbots. In short, it is about the users having a better understanding of the context along the way because you can easily see what you have answered and change it. The chatbot changes, for example, suggestions for the type of shoe depending on what you have chatted about earlier and during the conversation, and can derive more from the context. The result from the user evaluation showed that the users preferred to use chatbots with context view. In particular, it has been pointed out that the chatbot and humans can have the same mental model, and that this helps to counteract the negative aspects of chatbots that have been shown in previous studies (Mohit Jain et al., 2018, p.5).

## **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. If a chatbot works well it will give answers to many questions faster than if the user was to contact customer service. Another motivation is that they are available ‘all-the-time’ for users.

Kvale et al. (2021) 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 & Foelstad (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 impromptu. 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.

## **Critique of literature**

The observation may contribute to users having a different perspective than the researchers. The general question to be asked is there is bias present in the literature that may be outweighed by the findings from observational studies. We know that involving users in the design process, perhaps as in participatory design, is a good way to investigate what could be improved in the chatbots and open dialogue about desires. It could also contribute to diminishing particular traits and improvements for a specific chatbot, as chatbot evaluations can be one-sided issues from a specific contributor. It is always possible that the researchers and users may have a bias in relation to how the chatbot should be. Maybe the developers for example wanted the chatbot to be human-like, to make it emulate the personality of a customer service worker, but then users considered it to be too much like a human. There is a disagreement as to how the chatbot should be. If the chatbot is too impersonal that could also be bad for the user experience. The chatbot could for example say “have a good day” or “goodbye” at the end of the chat to make it nice and friendly, or it could just say nothing if that is too human-like. This is an example of what could be disagreed about and discussed among researchers and users. Lastly, we would also like to air the thought that people may be satisfied with chatbots that are humble about its shortcomings, in line with what has been observed using Amershi et al. (2019)’s guidelines.

## **Results from the observations**

We decided to perform observations on four different service chatbots from different organisations, Posten, DNB, Morgenlevering and NAV. The observations were performed in natural environments, for example in front of the computer at home. The users recruited

were family and friends, because they were able to participate on a short notice and they had some experience with chatbots, but they are not experts.

First of all we needed some guidelines for the observations. These were used as requirements for the specific observation and how we collect the data.

- One user.
- The user should perform a simple task with the use of a service chatbot.
- Note how the chatbot tries to complete the task given by the user.
- Request that the user speaks out loud during the observation.
- Note the time used to complete the given task.
- Take notes of: the user's use of words, response time from the chatbot and number of steps to fulfill the task.

To observe the user's use of words was important, because we wanted to explore the use of natural language in chatbots, as discussed in the article by Brandtzaeg & Foelstad (2017). The observation of response time is inspired by the same article by Brandtzaeg and Foelstad, where they point to "ease, speed and convenience" as important to users when using chatbots (Brandtzaeg & Foelstad, 2017). The observation of the performance of the chatbot and user satisfaction is an important aspect to the success of a service chatbot according to Kvale et al.. (Kvale et al., 2021).

We decided on a specific task to each chatbot that we assumed were common tasks the user might ask a chatbot.

## **Observation of DNB's chatbot**

*Task: Order a new credit card.*

The user is presented with the task: "bestille nytt bankkort". The user opens the chatbot on the frontpage. The chatbot starts the conversation by presenting itself by name and a request that the user should use "få ord med hva man trenger hjelp til er ønskelig". The user types "bestille bankkort". The chatbot interprets that the user wants a new credit card, and presents how the user can order a new credit card in their online bank, along with other

alternatives presented as buttons. The alternatives are: “bestille nytt bankkort”, “erstatt kort”, “kortet utløper snart” and an alternative to “aktivere kontaktløs funksjon”. The user clicks “bestille nytt bankkort”, the chatbot sends the user to a new page to order a new credit card. This sequence lasted for 24 seconds.

## **Observation of NAV’s chatbot**

*Task: Request an appointment with a NAV consultant.*

The chatbot is represented with a cartoon face and a guiding message to indicate this is a chatbot. Once the user opens the chat interface, they are greeted with a introduksjon as simple as “Hi I am Frida and I am NAV’s chatbot”. Right after the message, the chatbot sends to alternatives, “For English, click here” and “Jeg er lege eller samarbeidspartner”, before sending a last message asking what the user needs help with. The participant found this transition of conversation-themes to be “out-of-place”, but also understood that it must be asked so that the chatbot can “answer the questions as good as it can”. For demonstration purposes, the participant chose to click “For English, click here”, which triggered an automatically written message saying “English, please”. Unfortunately, the chatbot does not support answering questions in English, and it guides the user to pursue help by contacting by phone (as the observation took place after the staff’s working hours). Luckily for the participant, there is an option to restart the conversation with the chatbot (which indicates that the user can restart a conversation when they feel like they do not receive the responses they want).

Once restarting the chatbot, the participant tries to solve the task in Norwegian. As a response to “Hva lurer du på?”, the user types “bestille konsultasjon”. The chatbot immediately replies that it is unable to comprehend the question and tells the user that it works the best with short questions. This makes our participant a bit dumbfounded, as “bestille konsultasjon” already is short enough! The following message urges the user to use keywords instead of full questions, and also presents a link called “Hvordan få svar fra en chat-robot”.

The participant tried to repeat the question by typing “bestille time” instead of konsultasjon, which triggered the chatbot to respond with an appropriate response. It tells the user that if they have a digital activity plan, then they may contact their consultant through their digital messaging service. However, this is not the case for our participant, and they are urged to call the customer service desk and present their phone number. Right after, it presents four links; “endre møte”, “syk eller i karantene?”, “videomøte” and “åpningstid telefon”

## **Observation of Morgenlevering chatbot**

*Task: Order a breakfast delivery offer from Morgenlevering.*

The user is given the following task: order a breakfast offer through the chatbot for morning delivery. When the user opens the chatbot for morning delivery, he is first told to "logg inn". The chat solution therefore requires login via schibsted. Users have questions about how to order breakfast and it is an advanced setup to get a schibsted account with, among other things, verification through bankid.

The user has a schibsted account and can log in. When you re-enter the chatbot, it says that you must log in even if you are logged in - a little further down you get the opportunity to send them a message, but it takes time before the user discovers this.

When the user enters the actual conversation with the chatbot, the chatbot suggests a number of things to "help" the user. The user then selects "Når og hvor leverer dere?". The chatbot then asks a series of questions that can only be answered by auto-questions, such as "Hvor bor du?". When it stops asking with only a pre-filled answer, it manages to tell very briefly about what morning delivery is. Finally, the user is asked if they got an answer to what they were wondering about, this can only be answered with the answer "Ja" or "Nei". The conversation ends when the user says "Ja" even though the user has several questions to ask. The user is not given the opportunity to change their answer and the only opportunity to get more help to order breakfast is to start a new conversation with the chatbot.

The user tells during and after the observation that the experience with morning delivery chatbot was less good. In particular, the user emphasizes that it is along the way where you do not get out of a loop, do not get the opportunity to edit the answers and finally where the conversation only ends when the user is left with several questions - that you experience frustration and that the chatbot is too "stupid" .

## **Observation of Posten's chatbot**

*Task: Order a sign for a mailbox from Posten.*

The user has to navigate to the chat icon on the bottom of the page, and select "start chat". The chatbot starts the conversation with "Hei! Jeg er chatbot hos Posten og Bring. Jeg svarer best på korte og konkrete spørsmål. Bruk gjerne bare stikkord.

Hva kan jeg hjelpe deg med?". The user writes "postkasseskilt". The chatbot provides the user with three alternatives after three seconds: "bestille skilt?", "gratis postkasseskilt" and "ikke mottatt skilt". The user chooses "bestille skilt" and after three additional seconds the chatbot presents four alternatives: "postkasseskilt", "dørskilt", "skilt til husveggen" and "gratis postkasseskilt". The user chooses "gratis postkasseskilt" and the chatbot provides the user with the following information: "Når du registrerer varig adresseendring på nettsiden vårt får du tilbud om gratis postkasseskilt, slik at du kan merke din postkasse på ny adresse. Logg deg inn med referansenummeret fra adresseendringen for å bestille det. Referansenummeret finner du i bekreftelsen du har fått på epost eller ved å logge inn på "Min adresse". The user chooses "min adresse" and the chatbot guides the user to the login page for electronic id. The total time spent on the interaction with the chatbot was about 50 sec and five steps until the user was sent to the login page.

## **What we have learned from the observations**

Regarding the DNB chatbot we argue that the chatbot did not fulfill the user's request. The chatbot referred the user to a different site, and was not able to order the new credit card



from the chat dialogue. Based on this assumption we see that the chatbot only guided the user in the right direction, and it would be interesting to investigate if the user was satisfied with the result of the chatbot. Related to the challenges in Kvale et al. they argued that the users were generally satisfied when the chatbot solved issues that were “specific and concrete with a simple-to-understand answer that solves the problem” (Kvale et al., 2021). In the case of DNB’s chatbot we argue that the chatbot did not solve the problem, it only guided the user in the right direction, and therefore this could cause poor user satisfaction.

The interaction with NAV’s chatbot took 54 seconds for the participants, though they were not able to get the help they yearned for since the chatbot is incapable of planning a meeting with a consultant on behalf of the user. Since the participant does not have a digital activity plan, they were also urged to contact by phone, and they felt as though they wasted their time contacting through the chatbot since they would have had to contact by phone anyways. Again we see a case where the chatbot is not able to fulfill the given task and therefore the user satisfaction is low, based on Kvale et al. conclusions (Kvale et al., 2021).

The user of NAV’s chatbot did feel like they received messages quite quickly, even though the chatbot periodically sent a few messages and paused to create a “natural” change of theme in the conversation. This could be an attempt on natural language, as Brandtzaeg & Foelstad argue could help with the communication barrier (Brandtzaeg & Foelstad, 2017). In the end, the user had to go through four stages before being told that the chatbot cannot help the participant, and in regards to that they wished the chatbot could immediately tell it is unavailable in English to begin with. This could give us an indication that the demand of “ease, speed and convenience” is not fulfilled and the consequence could be that the user refuses to use the chatbot (Brandtzaeg & Foelstad, 2017).

The article by Mohit Jain, et al (2018) explains how giving a user the opportunity to change the context along the way makes the user and the chatbot's mental models more balanced. In the observation of the use of morning delivery, it is clear that there are large discrepancies between the user's expectation of what a chatbot can do and what it actually offers. When the user according to the chatbot answers "wrong" by pressing "yes, I have received the information I need" then the conversation ends. The user is not given the opportunity to correct the answer and has to start a conversation again. This is an example of how a chatbot does not meet the user's expectation of being able to ask for more things.

The same issue from the DNB chatbot is present in Posten's chatbot, because it guides the user to a login page. The specific task is not fulfilled by the chatbot itself, but it needs some additional work from the user. On the other hand we see that the given task of "postkasseskill" at Posten and "bestille bankkort" at DNB includes an addition of personal data from the user, and this requires the user to access their profile with information the chatbot does not have access to. Therefore the chatbot is not able to complete the task, and it is important to state that this is because of the user's sensitive data. If the user had given the chatbot an easier task it is more likely the chatbot could solve it, and it could result in a greater user satisfaction, as stated by (Kvale et al., 2021).

## Discussion

Based on our findings from both the literature analyses and our observations, we have investigated several common characteristics and traits that have been apparent in the studies.

Recap of our research hypotheses:

**H1: *Why do people decide to not chat with chatbots?***

**H2: *What deciding factors make chatbots successful?***

Our assumptions for this study were based on Amershi et al. (2019)'s guideline G1 'Make clear what the system can do' and G5 'Match relevant social norms'. We assumed that the chatbot needs to imply the level of detail that is needed, but also needs to comply with the natural language for its use context so that the user is able to have a successful conversation with the chatbot.

Our literature study points to natural language being a cofactor for personal chatbots, but does not cover for service oriented chatbot. We find the natural language to be an important aspect in service oriented chatbots, but there is also a subconscious desire for formal but understandable language to keep a serious facade for the service that is providing the chatbot. In personal chatbots, the issue is relatively tied to the chatbot not representing itself with a personality and critical thinking. In service oriented chatbots we have learned the

opposite; that adding too much personality and informality may make people refrain from interacting with the chatbot as its main purpose is to help customers, not entertain them.

One important implication that was common in our chatbot observations is the way the users chose to respond to their assigned chatbots. They adapted to the social norms and use context that the service chatbot was operating under, and naturally answered in short, formal language to ensure the integrity of the conversation. However, when the chatbot was unable to process user input, like with the NAV chatbot, the users quickly adapted by changing to keyword based communication. And when even that failed, they manipulated their language (like changing words with synonyms) to make the chatbot understand.

Other cofactors we have investigated aside from the chatbots language is task-completion satisfaction, ease-of-use, and quick replies. These three aspects have been touched upon in our literature analysis, and are also prominent factors in our observations. We therefore see that whenever the chatbot performs suboptimally in any of these three aspects, the user satisfaction scores lower. These factors are however also tied to other independent factors like internet speed (how quick the system is able to send replies) and if the system is able to provide help to the given issue to begin with.

## **Conclusion**

In summary, we have concluded with that factors that make chatbots undesirable to use in service oriented settings is when the system...:

- ... has unclear or complicated language;
- ... is unable to complete its task in helping the user;
- ... is not straightforward to maneuver in chat-sequence dialogue;
- ... is responding slowly.

The opposite of these traits are hence naturally what makes a chatbot successful when measured for user satisfaction and usage.

## **Feedback and work**

### **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.

### **Iteration 2**

Having in mind the feedback we got on Iteration 2 we have chosen to dive deeper into what we have done in the literature review. It was requested in the feedback that we have a more thorough review of what we have done. In addition to this we have included critique and discussion about the subject in the assignment to show our thoughts about the subject. We chose to focus on having the observation in the third iteration because we were made aware of a need for more findings on our area of interest. We hope this helped us to improve our research on the issue.

## 1.8 References

### From the mandatory literature:

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., ... & Teevan, J. (2019). Guidelines for human-AI interaction. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (paper no. 3). ACM.  
(<https://www.microsoft.com/en-us/research/uploads/prod/2019/01/Guidelines-for-Human-AI-Interaction-came-ra-ready.pdf>)

Bratteteig, T., & Verne, G. (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: Inappropriate 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:

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](https://doi.org/10.1007/978-3-319-70284-1_30)

Brandtzaeg, P. B., & Foelstad, A. (2018). Chatbots. Interactions (New York, N.Y.), 25(5), 38–43.  
<https://doi.org/10.1145/3236669>

Brandtzaeg, P.B., Skjuve, Dysthe, & Foelstad, A. (2021). When the Social Becomes Non-Human: Young People’s Perception of Social Support in Chatbots. Association for Computing Machinery (ACM).

Jain, M., Kota, R., Kumar, P., & Patel, S. N. (2018). Convey. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 1–6. <https://doi.org/10.1145/3173574.3174042>

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](https://doi.org/10.1007/978-3-030-68288-0_14)

Verne, G. B. (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>