AI and human (dis)abilities

Sara Prestvik Kristensen - Sarapk@uio.no
Emma Tvinnereim - Emmatv@uio.no
Maria Roheim - Marinro@uio.no
My Schultheiss - Mysc@uio.no

AI for all

We are a group of four students who found a common interest in learning more about- and researching the field of AI and Universal Design. We have all done our bachelor's degrees at IFI and are glad that there is common ground, even though we are aware that it's beneficial with input from students with different backgrounds. Therefore it's great that we have our minors from pedagogical, psychological, and sociological fields.

AI and UD

As mentioned above, we are interested in working with AI and UD, and more specifically how everyone can be included, as well as excluded from AI. Almost no guidelines talk about AI in contexts of social responsibility, help, care, welfare, but are focused on more practical concerns such as privacy, accountability and fairness (Hagendorff, 2019, p. 103). In addition to this, Bratteteig & Verne (2018) reported that in multiple instances, people have a tendency to alter their behavior and help the systems perform better than they initially do (p.4). When this is not adequately investigated and regulated, the user might end up as the losing part. We initially started with the research questions *"Does AI create disabilities for the users?"* and "*What is the consequence when the user does not have the possibility to alter their behavior to accommodate the flaws in the AI system?"*. After some feedback on the project, we have decided to formulate new research questions: *"When does AI create disabilities for the users?"* and *"How does AI create disabilities for the users?"*. This allows us to examine in what situations AI becomes a problem, rather than the question *if* it does it at all.

To further our research and understanding of the questions, we have decided to focus on Siri and similar voice assistants. Our understanding is that these are often made by English speaking companies, and might make it harder to use with another language, even though they say that this is a possibility.

Methods and research

We want to research this question to learn more about the negative aspects of AI solutions. Further, we want to investigate if there are any areas of improvement and conclude with suggestive metrics that can be used to ensure usable and inclusive design. In our own experiences, certain voice assistant solutions do not necessarily work well in supporting us as users. They may create situations where we have to go through more intricate processes to reach our goals than we would have if we skipped the step of trying to simplify them through the voice assistant at hand. This creates a barrier where we feel unable to proceed with our goal.

To investigate the research questions, we will discuss problems with AI voice assistants presented by Amerish, Corbett & Weber, Luger & Sellen, and Hagendorf, and present some possible improvements. In addition, we have conducted a series of short interviews coupled with a user test to get a better understanding of how people interact with voice assistants. The reasoning behind this is that while an interview gives us information on how the subject interacts with voice assistants in everyday life, the user test allows us to get a better insight into the result of this interaction, and how well it is carried out. The study was conducted with 6 participants, and with both Siri, Get boX and Samsung Smart-TV as voice assistants. After every user test, we did a short debrief to investigate whether the results from the test was a regular occurrence, or if there was something different about these results compared to the ones the subject usually achieves.

We will now use previous research that addresses similar issues to discuss how AI can create disabilities for users. The section after this will explain the data collection and present the analysis and findings. Concluding this paper, we will combine these sections and discuss possible solutions.

When and how does AI create disabilities for users?

A voice assistant is an AI-infused system. This means the system has features that harness AI capabilities directly exposed to the end user (Amershi et al., 2019). AI-infused systems can be inconsistent because they learn from user behavior. This can lead to the system changing over time and behaving differently from one user to the next. Amershi et al. (2019) claim that AI systems can violate established usability guidelines, and this inconsistent and unpredictable behavior can confuse and create challenges for the user.

Other authors who discuss how conversational agents and AI can create problems for users are Corbett and Weber (2016). They argue that using voice assistants could provide a user with a high level of accessibility and independence, but then the implementation and design must be done right. They further this argument by stating that difficulty learning and discovering voice commands could negatively affect the user experience. Corbett and Weber (2016) explain these challenges for users with limited hand dexterity, but we think they discuss several important points that could also apply to other users. One of the problems presented is long-standing issues of discoverability and learnability with voice interactions. Learnability is here described as a key aspect of the overall usability of technology and contains how easily new users can begin effective interaction and achieve maximal performance. The problem with discoverability comes from users' inconsistent and inaccurate mental models of what a system can do and how to interact with it. This is particularly difficult for voice interactions as they are inherently invisible. Since affordances and metaphors do not aid interaction with conversational agents in the same way as other interaction forms, this creates a challenge with learning and discoverability. They claim that the main problem is that users will assume the system can understand more than it is capable of, and that users will be unaware of the available functionality. In addition, techniques and design principles that address the learnability and discoverability of voice interaction design are limited. One example is that the participants in the case study presented in the article expected to be heard and understood at any point in time. They thought the system would listen to them as long as they continued talking, even when pausing to reflect in between (Corbett & Weber, 2016, pp. 1-3). This shows how conversational agents might create disabilities for the user based on our mental models and expectations.

Luger & Sellen (2016) presents another problem with conversational agents (CA). Although much research exists regarding technical aspects of AI, or systems specifically created for certain contexts, little research has been conducted asking the question of which factors result in the general success of CA (Luger & Sellen, 2016). In their article, Luger and Sellen (2016) describes a use of CA that is relatively simple where the user usually starts their interaction with the system with natural sentences. However, the need for a simplified language is often needed to achieve the result the user initially was expecting. Furthermore, the article revealed how a majority of the women included in the study found that the CA had a problem with understanding their voice and accents, whilst the men did not report this as an ongoing issue. It was, however, the most common reason as to why the CA did not work as intended. Thilo

Hagendorff (2019) describes in their article how the AI community is lacking when it comes to diversity, and it is clear within the field of AI research and development, and the workplaces in which this takes place. They present how “In the end, a relatively small group of predominantly white men determines how AI systems are designed, for what purpose they are optimized, what is attempted to realize technically, etc.” (Hagendorff, 2019, p. 105). This statement, coupled with Luger and Sellens findings, may indicate how CA’s such as Siri and Alexa might create disability for the users based on their gender and pitch of their voice. We now want to present our findings and possible solutions to these problems.

Data collection and findings

The user test was planned as a simple, four step test to investigate common and easy use of a voice assistant. The reasoning behind this was to gain a better understanding of *if* and *when* errors occurred in common, everyday practices. The four tasks the participants were asked to perform were:

1. Find a movie you would like to watch.
2. Send a text message.
3. Google something.
4. Ask what time it is.

User	Gender	Age	Voice assistant
1	Male	47	Siri
2	Female	47	Siri
3	Male	18	Siri
4	Male	25	Siri
5	Male	25	Siri
6	Female	24	Get Box
7	Male	27	Samsung smart-TV

When asked whether or not the participants used voice assistants in their daily lives, the responses varied quite a lot. User 6 said that she never used voice assistants, mostly because she did not see any need for it. She added that lately she had started using the voice assistant "Siri" because the touch screen on her phone was poor after she dropped it on the ground, and that she was quite satisfied with how it worked. The Get Box assistant however, a built-in voice assistant in her TV-remote control, she pretty much stopped using after the first time she tried it because it never gave any satisfactory results. This also applied to user 7, who had tried using the smart-tv a couple of times, but experienced that the system rarely understood what he was saying. Further, we asked if they ever have experienced any problems with voice assistants. Participant 6 said that she had never had anything but problems with the Get Box voice assistant, and illustrated her point through trying to perform task 1 "Find a movie you would like to watch" several times. There were no results. The system did not pick up any words, as if she was not talking. We asked if it had always been like this, and she told us that the first time she tried it it must have picked up some words, because even though the results were not good, there had been some.

She then told us that she was never quite sure where she should point the remote control when using the voice assistant. The Get boX was close by and always responded when pressing a button on the remote, even if she did not point it straight at the decoder. At the same time she thought there was a microphone at one of the far ends of the remote, which happens to be the same end that bluetooth signals were sent from when pressing buttons. Therefore the remote would be turned opposite from the decoder if she talked into what she thought was the microphone, and if she pressed a button with the remote this way, no signal was received by the decoder. She then started scrutinizing the remote and uttered that she had no idea where one was supposed to talk or in which direction the remote needed to be pointed while talking. A quote from the participant highlights her frustration in this situation: "Now I'm not sure whether I'm just stupid or what.. Can you tell me how this works? I feel like a 94 year old woman now".

Possible improvements and solutions

Corbett and Weber (2016) claim that implementing an AI voice assistant requires research to address the usability challenges that follow voice interaction and assistants. They address these problems and propose improving the visibility and learnability of voice interactions to increase user experience. They claim that in the initial stages of voice interaction, the user

needs the most assistance and that over time, users gain efficiency in command discovery and usage. One of the solutions Corbett and Weber (2016) presents to this is to learn as you go, and the availability and effectiveness of tutorials and training material is a key element in learnability. They claim that it seems more effective if it occurs in the context of the actual use environment. Another solution is discovery-based learning because they claim that a trial and error approach is a natural human tendency to voice interactions. Users must be encouraged to learn by exploring and interacting with the environment (Corbett & Weber, 2016).

Based on our findings, we think this trial-and-error solution could help users feel safer when they interact with an AI voice assistant. For example, participant 2 said that he prefers to do tasks on his own because he wants to feel like he is in control. He also said that he feels stupid when using voice assistants. If the system facilitates learning as the user goes in the actual context of use, it could help them become familiar with the system and its possibilities to feel more in control.

Amershi et al. (2019) discuss another solution for designing human-AI interaction. They propose 18 generally designed guidelines that can serve as a resource for practitioners designing applications and features that use AI technologies. Even though the authors emphasize that some guidelines do not directly apply to AI systems without graphical user interfaces, like voice assistants, we think that some do. We believe that following these guidelines could limit the chance of AI voice assistants creating disabilities.

For example, two guidelines are “Make clear what the systems can do " and “Make clear how well the system can do what it can do”. These could improve the system for participant 2 and participant 6. Participant 2 explained the need to feel in control. If the user knows what the system can do, it may be easier to trust. After doing the user test, the participant explained that he was surprised by the use of Siri, both positively and negatively. He had used Siri before, but not in a long time, and was positively surprised that Siri had improved since he last used the system. At the same time, he also realized that if he asks a more complex question, it is up to chance what Siri will do. Therefore, he realized he needed to keep it simple to get the desired result. This shows how the guidelines could assist the users in showing how to talk to the system, and what it actually can and can not do. Participant 6 would from time to time test the Get boX voice assistant because it was to some extent

entertaining to see what feedback it would produce, but it had never provided the correct results. One example she gave was that she wanted to watch one of the “Harry Potter " films, and activated the assistant and said “Harry Potter”. No results were provided and she therefore tried saying it in different ways; with clearer pronunciation, in a British accent, in an American accent and so on and so forth. After a while she just felt ridiculous for spending so much time trying to get the voice assistant to work, so she decided to search manually instead. Combined with the fact that she did not understand it is physically set up as described in the data findings, clearer layouts, set up and feedback from the system could improve her user experience.

To conclude, we believe that in some situations, AI does create disabilities for users. Therefore, we agree with Amershi et al. (2019) and think additional guidelines may be necessary to design intuitive and effective voice assistant systems. If not new guidelines, perhaps an even stricter perspective on how a solution should behave to accomplish them. In theory, “make clear what the system can do” and “ make clear how well the system can do what it can do” is clear and straightforward. The issue, presumably, lies with the companies who follow them as they see fit, and present solutions that do not properly achieve what the guidelines aim them to do.

References

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for Human-AI Interaction. *Conference on Human Factors in Computing Systems*, 1-13. https://doi.org/10.1145/3290605.3300233(CHI '19)

Bratteteig, T. & Verne, G (2018). Does AI make PD obsolete? Exploring challenges from Artificial Intelligence to Participatory Design. https://doi.org/10.1145/3210604.3210646

Corbett, E., & Weber, A. (2016). What can I say?: addressing user experience challenges of a mobile voice user interface for accessibility. 72-82. https://doi.org/https://doi.org/10.1145/2935334.2935386

Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and machines (Dordrecht)*,*30*(1), 99-120. https://doi.org/10.1007/s11023-020-09517-8

Luger, E., & Sellen, A. (2016). "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. 5286–5297. https://doi.org/https://doi-org.ezproxy.uio.no/10.1145/2858036.2858288

Appendix

Chatbot project

The goal of our chatbot is to simplify the working tasks for nurses regarding structuring nutrition plans for elderly patients. One of the group members is a domain expert on what is needed when working with nutrition plans for elderly patients. Therefore, we used her expertise as input for the flow.

The chatbot has several important tasks like understanding the problem of current nutrition plans for the given patient and giving guidance to changes should be made. For this, it needs to calculate calories and nutrition for different food items. In addition, the chatbot should handle input like a security number and identify the patient from the archive.

During this project we have learned that the chatbot is not as smart as first anticipated, e.g. it does not understand small typos. Therefore, we had to change the workflow of the interaction and make it easier for the chatbot to understand by making default lists of options. In this way, we can minimize the chance of misunderstandings.

For this project we used WatNot as a software to make the chatbot.

