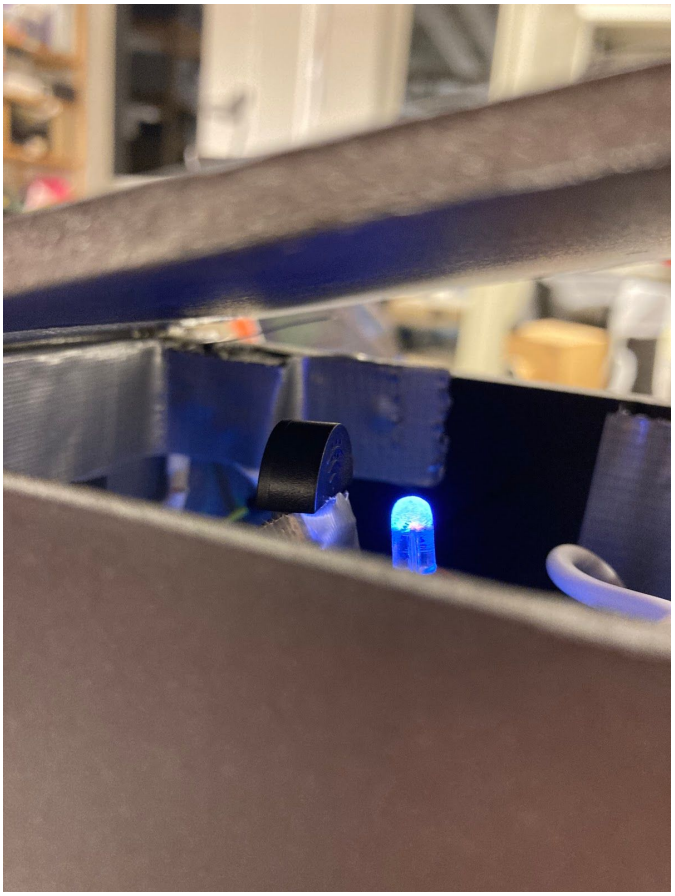
Group assignment

Final report

Fall 2020

“Black box”



Group 3

Niels Gabriel Theissen - nielsgt@uio.no
Stian Grimsrud - stiangri@uio.no
Gwendolyn Borchsenius - gwendob@uio.no
Henriette Elde - henreld@uio.no
Harald Maartmann-Moe - haralmaa@uio.no

Table of contents

1. About us

Our group consists of five master students and one Ph.D. candidate, four of whom have backgrounds from the bachelor program Design, use and interaction at IFI, and one with a bachelor in both Media Studies and Media- and Interaction Design from UiB.

2. Area of interest

Over the past decades, interaction via Graphical User Interfaces (GUI) has become a common form of interaction. Most technologies we use in our daily life enable interaction via GUIs. However, for some people with cognitive or sensory impairments, interaction via GUIs may be challenging, decreasing user experience and resulting in exclusion of particular user groups. Therefore, people living with cognitive, developmental, intellectual, mental, physical, or sensory impairments have repeatedly ended up as an afterthought, excluded by technology. About 15 percent of the world's population lives with cognitive, social, or physical impairments (WHO, 2020). As these groups make up a smaller part of the total population, they are frequently less represented, or less prioritized during technology development and design. This lack of inclusion in design and development can lead to social exclusion (Foley & Ferri, 2012, p.192). Inclusive design requires additional arrangements or a different type of recruitment, which many technology developers and designers tend to evade.

Fortunately, some technology has also been developed with these groups in mind, benefiting them. Audiobooks, video captioning, remote controls, video conferencing, eye tracking/detection, and many more technologies have increased the quality of life for people living with impairments. Recently, AI has contributed with improved speech recognition, speech synthesis, sign language translation, simplifying content for people living with cognitive impairments, and visual aids describing whatever the user shows their camera. These developments open up technological possibilities and can contribute to the autonomy of previously excluded groups.

Through our groups' previous experience, we have found a possible beneficial application of AI in assistive technology/UD. In Maartmann-Moe's (2019) work with older adults, many participants suggested that sending messages would improve their communication with family and friends whom they value immensely, and often talk about as one of their greatest sources of joy. Message sending has also become an increasingly large part of using services and systems in phones designed for older adults: Some phone manufacturers still use buttons and touch screens that require fine motor skills, accurate timing of actions, and either great vision or extraordinary memory. This makes the message sending inaccessible.
We have been thinking about speech recognition and its possible benefit to communication with friends and family. Speech recognition has been suggested as an opportunity to enable

and include older adults in the evolving communication praxis in contemporary day-to-day conversational exchange.

As a group, our impression is that there is an appeal for more work surrounding these difficulties and opportunities for older adults and AI. We aim to contribute to this area within AI and universal design with our project work with identifying potential problem areas for further investigation.

3. Background

Recently, speech recognition and speech-based interfaces have grown in popularity as they offer a different approach to HCI. Over the past few years, the conversation has become a key research area within the field of HCI as many authors argue that conversation is the most natural form for interaction (Luger & Sellen, 2016, p. 5286). Arguably, humans interact most naturally with each other through verbal speech which is why there has been a rise in so-called Conversational Agents (CA) in the HCI field (Luger & Sellen, 2016, p. 5286). CAs can be defined as "dialogue systems often endowed with humanlike behavior" (Luger & Sellen, 2016, p. 5286). Examples for CAs are chatbots, interface agents, or virtual assistants such as Alexa, Google Home, and Siri. Speech-based interaction offers several advantages compared to conventional GUIs such as the improved user experience in multitasking situations, reduction of exposure to a display and more natural and efficient interaction. For instance, the average person can speak 150 words per minute compared to typing 40 words per minute (Boyd, 2018). CAs may also let more users use services that previously have only been available on smartphones and less accessible technology. Especially people with cognitive or sensory impairments could benefit from such systems.

However, the dynamics of how and why CAs are used are still poorly understood. There exists also uncertainty around why certain CAs meet user acceptance whereas others are rejected by users. Nevertheless, AI still has opened up possibilities to include and empower people with cognitive, developmental, intellectual, mental, physical, or sensory impairments. We want to look closer at the Universal Design of AI-infused systems and if Conversational Agents could improve older people's communication.

3.1 Questions we want to address

Throughout this study we would like to address these following questions:

1. Could AI-powered speech recognition extend older adults' communication capabilities with family and friends?

2. Additionally, speech recognition is not perfect. There will be errors on the part of the AI's recognition of speech. How are these incomprehensive handled by systems and users today? How should/could errors be handled?

To get a better understanding of the existing context surrounding speech recognition and older adults we began by taking on the following questions

- How do our participants prefer to communicate?
    - What could be better?
- Is AI used in the design for older adults?
    - If not: Why?
    - Is speech recognition used in the design for older adults?
        - If not: Why?
        - Within communication?
- Are the existing speech recognition technologies inclusive of older adults as users?
    - Are the voice characteristics of older adults compatible with existing speech recognition technology?
        - If not, why/how?

4. Methods

4.1 Overall approach

Our case study follows an explorative and descriptive approach where we want to investigate how older people communicate. We do not aim to explain why they communicate the way they do or how factors are related to each other but rather explore communication capacities themselves. In order to address our research questions, we will initiate a design process where we apply AI technology to extend the communication capacities of older adults. By describing and exploring the communication process of eldery, we will be better equipped to discuss the intricacies and potential application of AI in design for older adults and their communication. We do not aim to contribute with generalized theories or frameworks, rather with a case study with exemplifying challenges and potential solutions in applying AI. By working with a concrete case, we hope to discover challenges and opportunities that would not necessarily have surfaced from a literature review or conceptual/hypothetical discussion.

Further, we focus mainly on qualitative data in order to get a rich and detailed picture of how an older person communicates. We could have chosen a more experimental approach for our study. However, as our goal is not to generalise or identify any causal relations, a descriptive, explorative approach seems more appropriate. In addition, the ongoing pandemic would have made experiments with a large number of samples unethical, especially with older people as the main research group. Nevertheless, we include a bit of quantitative data in our study as well.

The initial phase of the project focused on getting to know the current state of the use of AI and older adults as users, specifically speech recognition by examining existing literature and by interviewing older adults and experts. With this improved understanding, we uncover and describe areas with challenges or potential applications of AI that apply to specific participants. In our discussion, we also present suggestions that could extend the communication capabilities of older adults in their everyday lives. Our suggestions might also serve as inspiration for further beneficial application of AI to extend the capacities of older adult

4.2 Data collection method

We have divided our data collection into three different phases: exploratory interview, semi-structured interview with follow-up task, and evaluation of prototype and dictation functionality.

Our participants in this project consisted of an elderly person referred to as an elderly user (EU) and a family member of the elderly person which we refer to as family member (FM). Both participants had a relation to one of the group members, which served as an advantage considering the situation around the pandemic, where at least one of us could meet with the participants in person. The other group members took part in the research through phone, and/or video chat to ensure the safety and wellbeing of the participants.

In the exploratory phase of the project, we identified a possible participant and reached out to one of the elderly user's (EU) family members to gain more insight about how they communicate, how they would like to communicate, and in what context.
The interview with the family member (FM) was conducted via phone, where we followed a semi-structured interview guide and rounded off with a small exercise. The exercise was inspired by a future workshop approach where one invites the participant to consider a certain situation in an imaginary setting with no limitations and everything is possible. The purpose of this exercise was to better understand the ideal form of communication for the participant and prevent the participant from withholding ideas and input due to them not imagining a way to implement these. Even ideas and input from the participants that are not technologically feasible to implement can help us understand their goals, values, and preferences. Both the interview and exercise were useful to better understand how our participant communicates with her relative, in order to find potential beneficial/useful applications of AI. In addition - by tailoring the application of AI to the participants of this design process, we further ensure that participants also get something out of the design activities.

After the initial contact and the first interview with FM, we scheduled another interview, this time with the elderly user (EU) to collect more information on her thoughts and experiences

on communicating with family members, her preferred ways of communicating, and her thoughts on AI and voice-activated systems. This session took place in the home of the family member where one of the group members physically took part, and another group member joined in online through Zoom. In this session, we conducted a semi-structured interview and followed up with a similar exercise as we did in the interview with FM, where we invited EU to consider a certain situation with no limitations and how she would want to be able to communicate in that situation. Further, we presented a task to help the participant gain some experience in interacting with AI. This was done with a simple online dictation interface (Chrome web speech API demonstration) and the built-in dictation interface on EUs own tablet. On the tablet, we tested both the dictation function in "notes" and by sending a text message using the dictation. We ended the session by asking some questions regarding her experience using this technology. The interview and exercise were voice and screen recorded in order for us to analyze the session in hindsight. The experience with applications like these can contribute to expanding the design space and serve as examples.

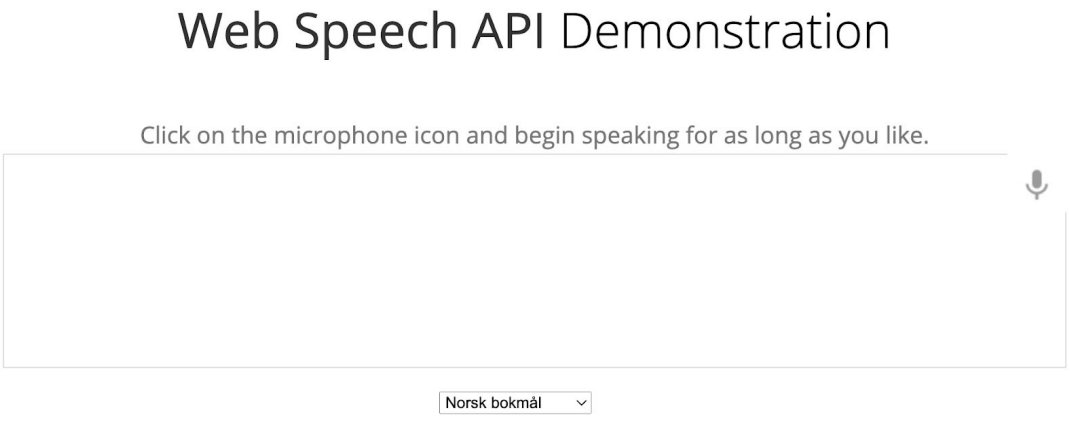
Further, we were able to supplement this with another interview/evaluation concerning the participants' experience with AI the following week. In the last session we wanted the participant (EU) to engage with a prototype so she could have the opportunity to gain more experience with the speech-to-text technology, which will help us in our further understanding if this technology is something that could improve the way EU communicates with family and relatives. We were also interested in identifying potential challenges for EU in the interaction with the AI-infused system. The prototype, described in part 5, would enable us to evaluate the dictation without any distractions such as opening apps, pushing buttons, and interpret a screen interface. To enable EU to articulate her perceived pros and cons with various technologies we structured the evaluation around comparing different ways of sending a message (future phone, voicemail, text message on a tablet, dictation on tablet, and a speech-to-text prototype). The messages sent would contain the same sentence. After each message was sent we would ask some follow-up questions concerning how she experienced sending a message with that particular technology.
In the first interview with EU, we identified a need for tools to enable EU to express abstract ideas, or rich descriptions of the use.
Drawing from this we prepared scales that could serve as "conversational tools".
We believed this would enable EU to express her opinions about the technology in finer detail. After the first follow-up questions, we asked her to range the experience on a scale from 1 to 10 on different matters like: "was it tiring to send the message this way?", "did it require a lot of focus and attention?". One of the group members facilitated the evaluation in person with EU, while another group member joined through Zoom as an observer. The approach to this session was to create a workshop-like environment where the conversation would be the main focus and get a hold of EU subjective meaning, rather than following a structured plan.

5. Sketches and prototypes

In our process, we utilized existing technology such as google chrome web speech demo and iOS dictation, to give our participant experience with the concepts and technology we wanted to discuss and explore. In addition to existing technology, we created a speech recognition message sending box named “Black Box” that would allow us to customize the speech recognition model, and allow the participant to use speech recognition with a different interface. Together, this set of prototypes served as valuable tools in discussing and exploring interaction with AI with our participants. We aimed to create a prototype that ideally can be tried out by real users, and iterate on the prototype to overcome initial challenges and discover further challenges and potential. We used this prototype to conduct different qualitative data collection methods and a quantitative data collection where we also look at errors in speech recognition that affect the overall user experience.

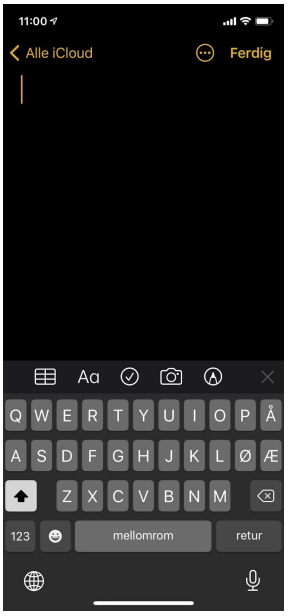
Following, an overview of the different devices and technologies used in interviews and evaluation:

Google chrome web speech demo
Google’s API demonstration (https://www.google.com/intl/en/chrome/demos/speech.html) website served as a simple way to let the participant interact with and explore speech recognition. We preselected language, increased the font size, and initiated the speech recognition for the participant.



iOS dictation

The participant owned an iPad that (when connected to the internet) allowed dictation. This also served as an alternative implementation of speech to text that supported our inquiry with the participant. This implementation was especially valuable as it allows the interpreted text to be sent as a text message, demonstrating that speech to the text message sent was a real option, not just a conceptual discussion.



iOS virtual keyboard

We utilized the participant's iPad virtual keyboard as an alternative interface/method of composing a message. We did not alter any settings such as font size but kept the preferences in their current state.

Voice messages

We utilized the existing voice message service that can allow users to leave an audio message for the recipient if they do not accept an incoming call.

Physical phone keypad

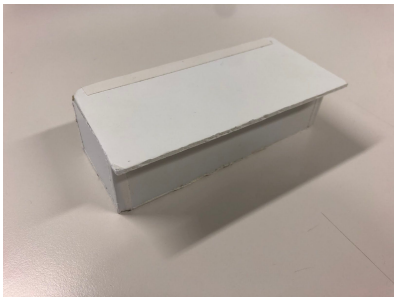
We utilized the participants' own phone in the process as well. We encouraged her to compose messages with her phone's physical keypad, in order to be able to compare and reflect on the different ways of composing messages.

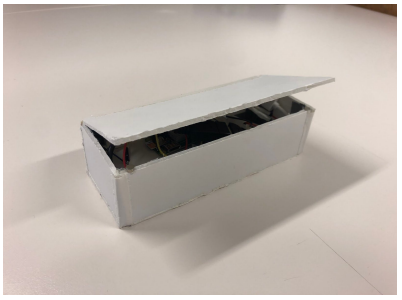
“Black Box”

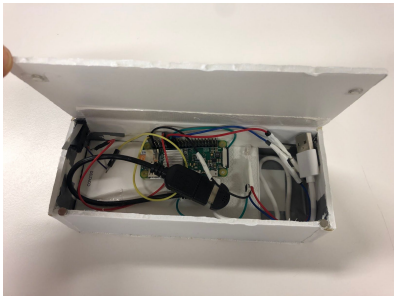
The first iteration of the prototype was simply a modified implementation of the google speech-to-text API where we displayed the APIs best interpretation of the speech, followed by less likely alternatives. The script could then send this text as an email when the right keywords were said.

This version of the prototype was not used with any participants as we found the existing published demo of chrome web speech to have a better interface. This prototype gave us some insight into the quality and inner workings of a new speech recognition model.

The second iteration took physical form as a box that one could open and close to send a message. It utilized the same API and a button that would be released when the box lid was opened. This combined with a power bank and a LED for feedback when recording created a simple way to let participants try out message sending built on speech recognition.



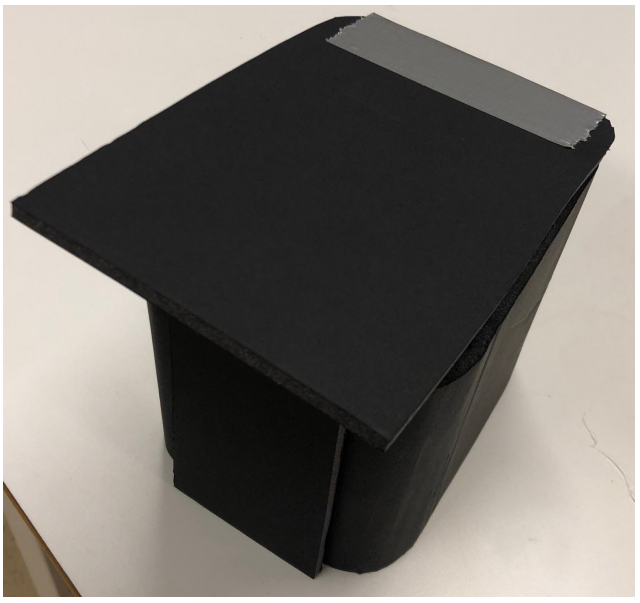


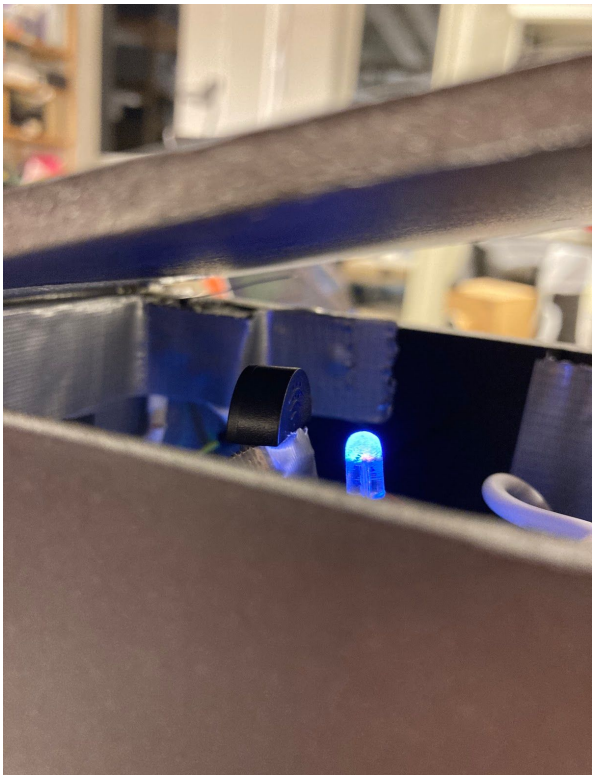


This prototype offered no feedback with details in how the speech recognition model interpreted the participants’ speech. Still, it offers an alternative implementation of speech recognition, as well as an alternative model which may be useful in demonstrating the different shapes speech recognition may take.

The third iteration focused on creating a prototype that could be used over longer periods of time. We intended to have our participants use this prototype as well as other methods of sending messages to keep in touch with us for a week, combined with a diary. As some older

adults have expressed a dislike of connecting USB chargers (Maartmann-Moe, 2019) in charging-ports, this prototype was built on a power bank that can be recharged wirelessly. This would hopefully make it easier to keep the prototype running for longer periods of time. The lid opens further and more easily to better allow the microphone to pick up speech. An RGB led offers more feedback on the state of the script through colors: No-internet, loading, listening, sending, successfully sent, error, and the lid closed. Though the prototype is literally a black box, we intentionally kept the wiring and electronic components exposed in order to avoid creating a black-box-like design (metaphorically). A black-box design would be especially unsuitable with an internet-enabled device that records audio in a participant's home.





For further details on the implementation, the project is published on UiOs github: https://github.uio.no/haralmaa/speechRecognitionMessageSender

6. Findings

6.1 Exploratory interview

In this section, we present the initial outcomes from the interview with a family member (FM).

From the conversation with FM we identified aspects of communication between FM and EU that could be better, factors that motivate participation in their communication activities, and also unwanted forms or aspects of communication.

Aspects/elements of communication that could be better

- The elderly user (EU) has access to two devices. One feature phone and one iPad. EU can use the phone unassisted but needs assistance to operate the iPad.
- The feature phone does not have the capability to receive or send images that hinder EU to get updates from her immediate family utilizing rich media.
- FM states that this is a feature that EU would be likely to appreciate.

Unwanted forms or aspects of communication

- FM explains that she sometimes avoids making phone calls to EU because they take up too much time. The effectiveness and possibilities for more frequent communication are mentioned as an upside to using text messages.

Motivation/value to participant in communication activities:

- FM would much rather make an appointment by SMS and have a longer conversation face to face with EU.
- Furthermore, FM emphasizes that text messaging would be beneficial for EU
    - in promoting communication with her grandchildren.
- FM explains that EU probably would like to write longer messages, but due to the keypad /interface on her feature phone, it is tiresome for her.

6.2 Exploratory semi-structured interview with a follow-up task

In this section we present the initial outcomes from the exploratory semi-structured interview with an elderly user (EU). In addition to the interview we introduced EU to some AI-infused tools as follow-up tasks.

From the semi-structured interview, we learned that EU has regular and frequent contact with family and friends (text messages, phone calls, tablet). EUs immediate family, EUs children and grandchildren, usually initiate contact. If EU receives a SMS from a family member, EU will answer with the same technology (if EU gets a SMS - EU answer with sms). When asked, EU says that they prefer the personal aspect of a phone call and prefer the phone call over SMS if there is time. Furthermore, EU prefers SMS when they need to get something done quickly or do not have a lot to say.

Aspects/elements of communication that could be improved

- EU has noticed that friends rarely answer voicemail.
- Says it would be nice to see the ones you are talking to
- Is reluctant to write messages on the tablet touch screen keyboard since EU struggles with locating the correct letters or symbols.

Unwanted forms or aspects of communication

- Voice mail

Motivation/value to participant in communication activities:

- Contact and make appointments with friends and family

Follow-up task (Dictation)

- Opposed to writing with a qwerty keyboard on a touch screen EU prefer dictation.
- EU stated that “I would like to have this” (technology/functionality)
- EU could not see any immediate obstacles using this technology.
- EU need assistance to use built-in dictation on ipad (a lot to remember)

Experience with AI-infused systems

- Introduction to speech-to-text tech baffled EU
- Had a clear physical reaction, jumping back in her chair laughing.
- Feeling of mastery. Seeking eye-contact maybe to express “did you see that” / “i did that”
- After texting FM with dictation on iphone EU proudly pointed at her chest and then pointed at FM in a “from me to you” gesture.
- Seemed more eager/active/engaged after testing speech-to-text
- EU stated multiple times that they would like to schedule a lesson in using dictation on tablet.

Analysis of transcription

- Error rate of ca 19% (punctuation, dropping some words)
- Errors did not alter the meaning or made the text appear nonsensical.

6.3 Evaluation of prototype and dictation functionality

In this section, we present the initial outcomes from the evaluation with an elderly user (EU). Below we present an overview over EUs error handling strategies to each specific device / technology and the what EU considered when defining ease of use. The participant concluded by rating the methods of composing and sending a message in the following (prioritized) order:
1) Sms on future phone 2) “black box” 3) Dictation on tablet 4) text on tablet 5) Voice mail

SMS (future phone)

- Ease of use attributed to habits/experience/size
- Error handling strategy: Delete all text even the misspelling / wrong word.

“Black box” prototype

- Ease of use attributed to recording and sending the message.
- Error handling strategy: Inconclusive

speech-to-text on tablet

- Ease of use attributed to navigation.
- Error handling strategy: Inconclusive.

Text on tablet:

- Ease of use attributed to qwerty keyboard and touch screen.
- Error handling strategy: Inconclusive

voice mail

- Ease of use attributed to lack of response from the recipient
- Error handling strategy: Correcting mistakes by addressing them at the end of the message.

Most of the problems during the evaluation occurred when using the tablet with dictation. These problems were connected to navigation, mapping, and visibility, not recording the message. This suggests that a clutter-free speech-to-text interface would be perceived as “better”. We also noticed a change in mood (smiling, laughing, gestures) when using devices with speech-to-text functionality enabled. This indicates that recording speech-to-text messages is an enjoyable experience for EU.
The dictation functionality on the tablet had no errors. This indicates that EUs voice is not a source of error and the (ML)model fits her voice. This invalidates a suspicion some of us had going into this project: that older adults voice could have been left out of the dataset that speech recognition models are built on, thereby excluding older adults form use of speech to text.

After EU had tried the prototype, the prototype was rated as easy to use and EU thought it was easier to send messages using this type of technology. Even though it was no option to correct errors using our prototype, it was a preferred way of sending text messages, second to phone calls, and SMS.

We realize that EU might avoid criticizing the *one* thing we have made ourselves (the prototype) and that enjoying new and "magic" technology is not necessarily equal to a better technology for EUs intended use.

7. Discussion

The questions we aimed to answer with our process were:

1. Could AI-powered speech recognition extend older adults' communication capabilities with family and friends?

2. Additionally, speech recognition is not perfect. There will be errors on the part of the AI's recognition of speech. How are these incomprehensive handled by systems and users today? How should/could errors be handled?

Our process has produced initial outcomes that suggest that AI-powered speech recognition can extend older adults communication capabilities. Though we have not found ways to better handle errors in speech recognition, we have improved our understanding of how speech recognition can benefit older adults in communication with family and friends.

Identified challenges
Text messaging is a part of the communication our participant uses. Writing long messages with existing technology can be tiresome, and is somewhat avoided.

The tools our participant is comfortable using excludes her from use of certain technology that could contribute to EUs inclusion. E.g.

Typing on iPad OS virtual keyboard can for some require tiresome searching for letters, which can discourage use.

Dictation in iPad OS requires significant memorization of steps, and simplifications/improvements should be explored with older adult users.

The evaluation of correcting errors in dictation and with the iPad virtual keyboard concluded that the interface and interaction was complicated and difficult. As speech recognition is imperfect, better ways of correcting mistakes should be explored.

Design implications
Speech recognition may serve as a valuable tool in extending/empowering communication capabilities as it may be a quicker, less tiresome method of composing text/messages that places less demands on fine-motor skills.

Use of speech recognition may demonstrate to older adults that different ways of communication exist, and that they are not limited to the tools that require fine motor skills.

The participant found the way of interacting with speech recognition by opening and closing a lid to be easy, and not demanding in memory. This demonstrates that implementations of speech recognition may be improved.

Research implications
Using the prototype as a tool for a longer period of time with some complementary tasks would give us a better understanding of EUs ability to use the prototype and potential learning curve with the prototype.

8. Ethics

Personopplysningsloven covers how information that can be traced back to an individual needs to be handled. We handed out an informed consent to all of our interview objects, ensuring legal and ethical treatment of personal data. The informed consent explained our goal with the study as well as the interview objects rights. We also communicated that the collected data will be deleted after the project. As mentioned earlier, the interview objects were family members of one of our group members which made the process of recruiting easier and more efficient. At the same time, we are also aware that knowing the interview objects personally may impact on the data. It is possible that both FM and EU committed to participate, motivated by doing us a personal favor.

When working with older adults one should make sure that their participation should be on their own terms. We inquired how our participants would prefer to participate, respecting their preferences regarding time and place, and whether we should meet in person. During interviews and evaluation we frequently proposed taking breaks, or ending the activity early if they prefered. We made sure to come to their preferred place when meeting in person, and only one member of the group visited to minimize infection risk.

To complete the evaluation of the prototype with the elderly user, we needed to connect to EUs personal wifi, obtaining EUs SSID and personal password. These are concepts that the elderly user did not fully understand. Therefore, we collaborated with the participants family in retrieving and temporarily connecting a device and storing the password in its configuration.

9. Conclusions and lessons learned

As we conducted this investigation towards a specific area of AI-design and AI-to-user interaction we have learned a great deal about how to execute a study to further expand our knowledge on the distinct research topic. By using relevant methods described earlier in the report.

Throughout the different iterations of this document, we have performed analysis focusing on relevant research literature, articles, practices, and other methods that we found applicable for our project. Among other things, we have increased our knowledge of chatbots, concepts in AI, which led us to know how to explain the literature and conduct data collection with relevant methods.

When we started our study we made two issues we wanted to investigate. These issues were; “Could AI-powered speech recognition extend older adults’ communication capabilities with family and friends?” and “How are the incomprehensive errors in AI-infused systems handled by the system and the users today?”. Through our study, we explored that identifying error handling with our user would be difficult because we only had one person included in the study. The reason for only having one person was access to elderly people in this pandemic, we didn’t want to risk their health on the behalf of the study.

We cannot conclude that our prototype can help “all” elderly adults in their communication, but we can see the possibilities the prototype can help our participant in daily communication with family and friends. During the study, we have seen that speech-text with AI can easify the task of sending a message and keeping up with the person’s family and friends.

References

AI Design Guidelines (2019) G9 - Support efficient correction. Make it easy to edit, refine, or recover when the AI system is wrong.
Foley A. & Ferri B (2012). Technology for people, not disabilities: Ensuring access and inclusion. Retrieved from https://www.researchgate.net/publication/263752183_Technology_for_people_not_disabilities_Ensuring_access_and_inclusion

Guidelines for Human-AI Interaction
In CHI Con- ference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland Uk. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3290605.3300233

Lohr, S. (2018, February 9). Facial Recognition Is Accurate, if You're a White Guy. The New York Times. https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html

Luger, E., & Sellen, A. (2016). Like having a really bad PA: the gulf between user expectation and experience of conversational agents. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (pp. 5286-5297). ACM. (https://www.microsoft.com/en-us/research/wp-content/uploads/2016/08/p5286-luger.pdf)

Maartmann-Moe, Harald (2019). "These things take time" Unhurried explorations and independent prototypes In participatory design with older adults https://www.duo.uio.no/handle/10852/70963

Norman, D. A. (2020). The "Problem" with Automation: Inappropriate Feedback and Interaction, not "Over-Automation." 10.

Schulz, T. (2020). Exploration of Moving Things in the Home. 136.
WHO (2020). World report on disability. Retrieved from https://www.who.int/disabilities/world_report/2011/report/en/

Shneiderman, B., Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy, arXiv.org (February 23, 2020). Retrieved from https://arxiv.org/abs/2002.04087v1

Boyd, C. (2018) Speech Recognition Technology: The Past, Present, and Future. Medium.com. Read: 22.10.2020 Retrieved from

https://medium.com/swlh/the-past-present-and-future-of-speech-recognition-technology-cf13c179aaf

Appendices

1. Chatbot design

What did we do?

We first had a group meeting where we decided which of the chatbot tools we wanted to use. We then had to decide on the purpose of our chatbot and how much time we wanted to spend on learning the tools. Harald, one of the group members showed how we could manage the task in the different tools available (Chattron, Chatfuel, and DialogFlow), and the group agreed on using DialogFlow because it better supports word recognition. In addition, we liked the user interface of DialogFlow better. One group member suggested making a chatbot about Stian and his handball team, the rest of the group agreed. The purpose of the chatbot is to give the user info about Stian and his endeavors playing and practicing handball.

How did we do it?

Before we started with the implementation of the chatbot, we had to identify some key tasks that the chatbot should handle. We identified three key tasks that the chatbot should handle in a good manner, and these were;

- give the user insights to Stians matches and give links to Stians stats
- give links to upcoming matches on the streaming platform sumo
- give links to Stians team (Haslum) stats

The first task is for the user and will give insight into Stian's matches, like the results and statistics for Stian in the given match. This will give the user options to ask the chatbot, named StianBot, about his stats from the previous matches. The next key task we identified was the chatbot to give suggestions on where to watch Stians matches, etc. TV2 Sport 2 or handballtv.com. This so the user easily can access Stians matches. The last key task for StianBot was to give the user statistics for Stian's team (Haslum HK), so the user could see how the team performed and see where in the table they are.

After we identified these key tasks, we started with the implementation of the chatbot. We then started with making entities and intents, so the chatbot would have some "keywords" it looked for, for example, match, stream, number, etc. We also used the default intents; Welcome intent and Default Callback, so we didn't need to implement these by ourselves. After some testing of the chatbot, we identified a need for buzzwords, because of the lack of knowledge of what a user could ask the chatbot. We then choose four buzzwords the user could interact with to get an understanding of what the chatbot could answer. We then found appropriate places to "set" these buzzwords, for example at the beginning of the conversation and when the chatbot didn't get what the user asked for. After making the chatbot we wanted

to integrate it into a conversation app (like Telegram and Messenger) and choose to integrate it in Telegram, because of the encryption of the messages.

Reflections on the outcome

While we were implementing the chatbot, all of the group members had the same experience in the making of the conversation. We had to design the chatbot for specific conversations and had to guess what a user actually would ask the bot. This is something we thought was a bit hard, to clearly identify what types of questions and words a user would use in the conversation. We then identified we needed to design the answers, in the same manner as we had to design the questions.

When thinking about the above, we also found out that a user could interact with the chatbot in a matter we didn't think about. This could result in a conversation with the chatbot that didn't give any answers to the users' questions. So we had to think about the whole conversation, so a user could get all the answers it needed.

2. Machine learning task - MovieChatbot

The second task in Module 2; “Design of interaction with AI” was to take an existing chatbot program and test it with different learning attributes and compare these. The goal of the assignment was to give us a deeper understanding of how AI and chatbots work internally.

```
n_steps = 1200
for i in range(n_steps):
    y_pred_train = n(x_train)
    loss_train = loss_fn(y_pred_train,y_train)
    optimizer.zero_grad()
    loss_train.backward()
    optimizer.step()
    if (i % 25) == 0:
        print(loss_train.detach().numpy())
```

We first began with the code we were given and tested this. Then we changed the ML model with 10 layers of neurons and a pyramid of these(128, 256, 512, 256, 128). That didn’t help the model, it actually got dumber. Then we tested with 5 layers of neurons with the same pyramid structure. It helped, but not as much as we wanted. From there we changed the structure of the neurons and added one layer at a time until we got to 8 layers and linear structure.

```
self.fc1 = nn.Linear(128,256)
self.fc2 = nn.Linear(256, 512)
self.fc3 = nn.Linear(512, 1024)
self.fc4 = nn.Linear(1024, 2048)
self.fc5 = nn.Linear(2048, 4096)
self.fc6 = nn.Linear(4096, 8192)
self.fc7 = nn.Linear(8192, 16384)
self.fc8 = nn.Linear(16384, num_classes)
```

We tested different combinations of training iterations and layers neurons and different combinations of neurons to see the difference in the outcome.

This is our result after trying out eight layers of neurons and making these layers linear, with 600 training iterations. Its 3.2% margin of error.

```
0.24798004
0.12711085
0.08269722
0.063733995
0.05397605
0.048242543
0.04454889
0.04200633
0.04016741
0.038785547
0.037715115
0.0368651
0.036176357
0.035608448
0.035133433
0.03473102
0.034386307
0.034088142
0.03382811
0.03359948
0.03339719
0.03321715
0.03305599
0.03291079
```

These are the results with the same neurons, but we doubled the training iterations. We got a little bit less of a margin of error. But not as big as we hoped.

```
0.054432813
0.04851372
0.044721466
0.04212207
0.040248066
0.03884355
0.037757687
0.036897074
0.036200695
0.03562721
0.035148077
0.03474246
0.034395278
0.034095276
0.03383367
0.03360397
0.033400718
0.033219766
0.033057917
0.032912336
0.03278069
0.03266121
0.0325523
0.032452863
0.032361384
0.032277264
0.032199632
0.032127835
0.032061145
0.031999145
0.031941302
0.031887244
```

Our conclusion after trying out these different training iterations and layers of neurons is that we didn’t see a clear connection between the margin of error and how well the chatbot’s

responses when we talked to it. For instance, we had a 3% margin of error, but the chatbot still repeated the same answers.

3. Evaluation

In this appendix task, we have chosen to evaluate our own prototype, which forms the base of our scope. We have chosen this because it is more relevant for our study than evaluating a random chatbot. For the first part, we will evaluate our prototype, "Black-box", and Abusability test. For the second part, we will evaluate "Black-box" through the Human-AI Guidelines developed by Microsoft and Shneidermann's framework.

3.1 Abusability test

Evaluation plan
We will go through three phases in the abusability test; 1) benefits, 2) Vulnerabilities, and 3) Abuse scenario.

1. *Come up with 3-5 value propositions statements that describe the (potential) benefits Replika brings to individuals or society overall.*
2. *Review the LBB and benefits above. Come up with 3-5 specific examples that highlight how Replika could be intentionally abused or result in unintended consequences.*
3. *Review the LBB, benefits, and vulnerabilities above. Pick a vulnerability from the section above or come up with your own. Begin to develop a story. Ask yourself: what could go wrong next? Try to imagine a snowball effect or the worst-case scenario*

Benefits

1. Doesn't need any form of fine movements to send a message. This is one of the biggest benefits because elderly people tend to lose their fine movements, and with our prototype, the user doesn't need fine movements.

2. Lets the user elaborate the messages in a faster way. This is an issue we found with our participant, our participant wished to elaborate on the messages, but found it difficult with the current owned phone. In that manner, our prototype will help the participant with this, by the speech-to-text prototype.

3. Easy to use. By being a box where you can open the lid, we find the prototype easy to use. By having limited functionality, can only send messages, it can get rid of disturbances.

Vulnerabilities

1. Anyone with access to the device can send messages posing as the owner. Because there it is not a passcode for the prototype, everyone that can get a hand on it would have the possibility of sending messages from it.

2. Others can listen in on you, either by physical presence or by accessing the device. So persons know what you send a message about.

3. Someone could change or add the destinations of your voice and message. If someone gets access to "Black-box" it's possible to change the destination of the message, so the person may send a message to the wrong person.

Abuse scenario

In the case of someone listening to what the user is saying, it is possible to imagine that the voice also could be recorded. This recording could be used to synthesize a voice, resembling the user. This voice could be programmed to say anything. This could enable the attacker to engage in conversations, impersonating the user, gaining access to personal information, or financial assets. This information could be further exploited to damage the users' reputation.

3.2 Guidelines review

Guidelines review using the guidelines for Human-AI interaction (Amershi et al. 2019).

Evaluation plan

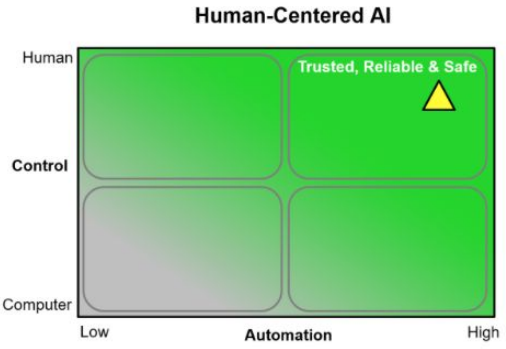
For this type of evaluation, we will conduct a guidelines review of "Black-box"(our prototype), using the guidelines for Human-AI interaction. In our review, we will state whether our prototype has applied some of these guidelines and which is not applied.

Findings

In our review of "Black-box" using the guidelines for Human-AI interaction, we found out that "Black-box" didn't meet many of them. For instance, Guideline 2, where it should be stated how well the system can do the specific task, this is not meet through our prototype.

But "Black-box" meets three of the guidelines; G1) "Make clear what the system can do", G7) "Support efficient invocation" and may support G13)" Learn from user behavior". Our prototype is a box, with a microphone, which indicates that you can talk to the box and then supports G1 in a some way. By allowing the user to chose when to start the interaction, when lifting the lid, "Black box" supports G7 fully. It is the user that chose when to interact with the system. G13 may be supported, but it depends on the API, we have used the Google API, and therefore think that the box will learn from the user behavior.

Evaluation of "Black Box" based on Shneidermann's framework for trusted, reliable and safe (TRS)



Based on the two-dimensional framework (Shneidermann, 2020, p.7) for achieving TRS AI-infused systems, we argue that the "*Black Box*" prototype lies in the lower right quadrant as google speech API offers a high automation degree. At the same time, the degree of human control is relatively low as the prototype does not support correcting speech errors or manipulation of the data. Once the prototype receives speech input and the box is being closed, "*Black Box*" will send the text message, not allowing for further manipulation of data. Increasing the level of human control could improve our prototype and increase reliability, safety and trust. This could be investigated in further research.

Critique of the "Black box" prototype from the perspective of Norman (2020)

Norman (2020) emphasises the vital role of feedback in systems with high automation. The "Black Box" prototype provides little feedback, only a RGB led indicating the state of the prototype, and nothing about the details of the message it is sending. Therefore, improving feedback would be a compelling next step if creating an actual product.

3.3 Lessons learned

During the process of conducting an abusability test and a guideline review of "Black-box" we have learned that the consequences of deviating from these guidelines possibly can lead to an undesirable outcome, like the one in our abuse scenario.

4. Interview Guide for FM

Notater
Hensikt: Begynn å svare på :

1. Could AI-powered speech recognition contribute to older adults' communication with family and friends?

Ved å bli kjent med deltakerne, kontekst, kommunikasjonen deres
semi-strukturert
→
Intervju med familiemedlem
→
Burde vel nevne at vi tar gjerne notater om hun synes det er greit og bruker det i en rapport til emnet (anonymisert), hun ikke trenger å svare/delta, kan avbryte når som helst, osv

Warm up
Introdusere caset, gruppe på fem studenter som går interaksjonsdesign på UiO; undersøker hvordan AI i tale-baserte teknologier kunne bidra til bedre/enklere kommunikasjon mellom eldre og familien/vennene deres; spesielt interessert i hvordan AI kunne forbedre hverdagen til eldre med kognitive og motoriske vansker.
vi er like interessert i hvorfor AI ikke blir brukt, og hvorfor det eventuelt ikke funker med AI for slikt

- Kan du fortelle litt om deg selv og om din mor?

bare skriv her om dere vil si noe uten å avbryte

hvorfor irritasjon med sms?
hvorfor ikke sms

Hovedel

- Vi har hørt fra Niels at du har mye kontakt med din mor, hvordan foregår den kommunikasjonen?
    - Kommuniserer dere med hjelp av teknologi?
        - [Vår oppgave: Finn et konkret eksempel som kan brukes i utforming av aktiviteter som kommer senere]

        - Hvor lenge har dere kommunisert slik?
        - Hva synes dere om å kommunisere på denne måten? glede/stress
            - Er det noe som kunne vært bedre
        - Var det noe opplæring for å kunne kommunisere slik?
- Hvordan foretrekker du å kommunisere?
    - hvorfor?
        - Hva er fordelene?
    - Hvis meldinger ikke nevnes: Bruker du meldinger til å kommunisere?
        - video-samtaler, bilder, talemelding, telefonsvarer, brev?
- Hvordan tror du EB foretrekker å kommunisere?
- Hvordan kunne de tenke seg å kommunisere?
    - (om det ikke var noen begrensninger på hva som var teknisk mulig å implementere
    - [Vår oppgave: Finn et konkret eksempel som kan brukes i utforming av aktiviteter som kommer senere]
- Hva er tre assosiasjoner du får opp når du tenker på AI?
    - knytter sammen med AI som konsept.
- Har de noe erfaring rundt AI
    - rundt talegjenkjenning?
    - hvis ja
        - hva slags
    - hvis nei
        - hvorfor ikke?
- Har de noen holdninger til AI
    - Til talegjenkjenning
    - diktering som eksempel

5. Interview Guide for EB

Hensikt

Bli kjent med hvordan og hvorfor den eldre deltakeren (EB) kommuniserer; hvordan kunne kommunikasjonen forbedres.

Praktisk

Samtykkeerklæring printes
To laptops - en til oppgaver og en til Zoom
Skriver notater underveis hvis du synes det er greit
Be noen av de som skal være tilstede sette opp diktering på forhånd
Video opptak under utprøving av diktering

H (og andre?) blir med på zoom
Om deltaker ønsker det tar vi skjermopptak på tlf og pc hvor hun tester ai + zoom opptak.

i tidsrommet 17-19, men sikte på maks 40 min?

Oppvarming

- Introduksjon
    - om oss og prosjektet
        - en gruppe studenter, vi tenkte å snakke om kommunikasjon, og teknologi som du eventuelt bruker i kommunikasjon
    - takk for at du vil være med
- Vi tenkte å snakke litt om hvordan du holder kontakt med din familie

Hoveddel

- Har du mye kontakt med familie og venner?
    - Hvorfor er det/ hvorfor ikke? Hvorfor er dette viktig/ikke viktig for deg?
- Hvordan pleier du å holde kontakten/kommunisere med disse, som f.eks. din datter, barnebarn?
    - Hva syntes du om denne/disse måtene å kommunisere på?
        - er det slik du foretrekker å gjøre det?
        - Er det noe som kunne blitt bedre rundt dette?
        - er det noe ved dette du ikke liker?
- Øvelse som er litt rar: forestille seg at alt man kan tenke seg var mulig å lage, ingen begrensninger, på/gjennom hvilken måte hadde du ønsket å kunne kommunisere med andre på? Hvis du ikke kan møte dem fysisk og snakket med dem ansikt til ansikt?
    - *[anerkjenn at det er vanskelig spørsmål, og at vi er interessert i å høre om hun allerede har noen tanker om det, eller om hun får det senere]*
    - Og hvorfor?
    - Eksempler om vi blir bedt om det:
        - jeg liker best å skrive med penn og papir, så jeg kunne tenke meg at jeg skrev meldingene mine på penn og papir, la det i postkassen min, og 3 sekunder senere så får mottakeren høre min stemme lese opp teksten jeg skrev automatisk
        - Jeg liker å se personen jeg kommuniserer med, så jeg kunne tenkt meg å snakke inn i et slags speil hvor jeg og den andre personen kunne se hverandre mens vi pratet

        - Jeg liker best å høre på meldinger, så jeg vil helst få meldingene mine lest opp automatisk, i stemmen til avsenderen
        - Jeg vil helst at meldingene mine skal dukke opp som tekst i himmelen, ved hjelp av stjernene
        - Meldingene mine skal kun dukke opp som tekst på fly-slept-banner
- Vi tenkte nå at det hadde vært spennende om du fikk se og prøve et program som gjør det du sier om til tekst på telefonen uten å bruke tastatur.

    - Referer til instruksene
    - Gå gjennom dikterings eksempel med deltaker ("Beskriv hvordan været har vært i dag" / "Beskriv hva du har spist til middag i dag")
      "Det er høst. Sommeren er forbi, den forsvant like så hurtig som den kom; akk hvor den gikk hurtig."
        - cmd + + for større font?
        - Chrome web speech
        - Google api
        - Tlf diktering i notat
        - tlf diktering til å sende melding
    - Hva synes du om det?
    - Hvordan var det å bruke?
        - Var det enkelt å bruke/forstå?
        - Hvordan var det å bruke i forhold til din egen telefon?
    - Er det noe du kunne sett for deg å bruke?
        - hvorfor/ikke?
- [noe om å komme fram til et konsept for tek vi eventuelt kan lage og la henne prøve videre med?]

Avrunding

- Takk, dette har vært veldig nyttig for oss
- Hvis vi rekker å lage noe mer og bedre som du kunne få testet ut, til f.eks å [noe som hun har nevnt som nyttig] kunne du/dere vært interessert i å prøve det og snakke med oss igjen?
- Vil du ha en kopi av det vi skrev ned av notater?

Abusability test

Benefits

- Enkel å bruke
- Enkel å lære
- Krever ikke finmotorikk for å skrive meldinger
- Lettere å bare snakke for å sende meldinger for noen
- Lar brukere skrive lengre og fortere
- Sender melding

- Med tekstmeldinger som er diktert med tale-til-tekst generelt
    - Man vet aldri hvem som sender meldingen
        - Man kan utgi seg for å være EB, kan ikke verifisere
    - andre kan lytte
    - Stjele din stemme, og bruke ml til å få din stemme til å si hva som helst
    - Ved å bruke stemme til å komponere meldinger vil andre kunne høre deg, hvor du er, og hva du sier
    - Stemmegjenkjenning-modellen kan velge å ignorere ord, dermed endre språket til brukeren
- Med boksen spesifikt
    - vanskelig å si hva som blir sendt
    - Lett å bytte ut innmaten, vanskelig å si om komponenter er byttet ut
    - Enkelt å koble seg til og endre innholdet
        - lett å endre hva brukeren sier uten at brukeren merker det siden boksen ikke viser hva den tolker og sender

Guidelines (amershi) gjennomgang

6. Form of consent

Beskrivelse av prosjektoppgaven

Vi er en gruppe studenter i kurset IN5480 - Spesialisering innen forskning i design av IT, ved Institutt for informatikk, Universitetet i Oslo. Prosjektgruppen består av Henriette Elde, Stian Grimsrud, Harald Maartmann-Moe, Niels Theissen og Gwendolyn Borchsenius.
Kursleder er Jo Herstad epost: johe@ifi.uio.no tlf: +47 91560563

Som en del av emnet skal vi gjennom observasjon og intervjuer undersøke ulike aspekter ved bruk av kunstig intelligens i talegjenkjennig. Vi skal også skrive en rapport om dette.

Vi ønsker å intervjue deg om dette temaet for å få en få en bedre forståelse av kommunikasjon via talegjenkjenning.

Frivillig deltakelse

All deltagelse er frivillig, og du kan når som helst velge å ikke svare på et spørsmål, avslutte intervjuet eller trekke tilbake informasjon som er gitt under intervjuet.

Gjennomføring

Vi vil gjennomføre et semistrukturert intervju, hvor vi har notert ned en del temaer og spørsmål på forhånd som vi ønsker å fokusere på, og en øvelse hvor deltakeren prøver ulike verktøy for talegjenkjenning. Det vil bli tatt lyd- og skjermopptak under intervjuet og øvelsen, om du samtykker til dette. Hele eller deler av lydopptaket vil bli transkribert og analysert. Notater og transkripsjon fra intervjuet og øvelsen vil bli brukt i kursrapport og publisert på Universitetet i Oslo sine nettsider. Lydopptaket blir slettet etter sensuren i kurset har falt. (desember 2020).

Anonymitet

Notatene og rapporten vil bli anonymisert: ingen andre enn prosjektgruppa vil vite hvem som er blitt intervjuet, og det du sier i intervjuet vil ikke kunne tilbakeføres til deg.

Behandlingsansvarlig

Behandlingsansvarlig er Niels Theissen

Kontakt: nielsgt@uio.no / Tlf: 988 61 317

Ved spørsmål/innsyn på data, kontakt behandlingsansvarlig.

Før intervjuet begynner, ber vi deg om å samtykke i deltagelsen ved å undertegne på at du har lest og forstått informasjonen på denne erklæringen, og ønsker å stille opp til intervju.

Samtykke til opptak av lyd og skjerm

I intervjuet ønsker vi å ta opptak av samtalen for senere å kunne skrive bedre notater om hva som ble sagt. I tillegg til lydopptak vil det være nyttig for oss å se videoopptak av hva som foregår på skjermen til teknologien du vil ta i bruk under intervjuet. Det vil ikke bli gjort noe videoopptak av personer, kun av det som foregår på skjermen til teknologien vi har med til intervjuet. Opptakene vil bli slettet i løpet av desember 2020.

Kryss av i avkrysningsboksene nedenfor om du samtykker til opptaksformene:

- ❏ Jeg samtykker til at det blir gjort lydopptak av intervjuet
- ❏ Jeg samtykker til at det blir gjort skjermopptak under intervjuet
- ❏ Jeg samtykker til at en enhet som gjør om tale til tekst ved hjelp av av internett er koblet på mitt nettverk under testperioden.

Samtykke

Jeg har lest og forstått informasjonen over og gir mitt samtykke til å delta i intervjuet:

_____

Sted og dato Signatur

Tusen takk for din deltakelse. Ved spørsmål angående prosjektet, vennligst kontakt hovedansvarlige Niels Theissen.