Second iteration

Module 1

1.1. Concepts, definition and history of AI and interaction with AI

1.1.1 The history of AI

The term AI - *Artificial Intelligence* was used for the first time in 1956 by John McCarthy, an American logician, and mathematician (Grudin, 2009, p. 49). This was in the light of a workshop where it was a lot of attention because of the optimistic forecasts of the participants. Even though this was the first use of AI as an official term, different issues and opinions around the technology have been discussed even earlier. Alan Turning's writing in the *London Times* in 1949 is an example of this (Grudin, 2009, p. 49).

1.1.2 Definitions of AI

Definition 1: ***"****It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable."* - McCarthy (2007, p. 2). This is the same McCarthy mentioned above who is behind the term AI. He defines AI with the main focus on the science of the technology, rather than the more human and social aspects of it when it comes to behavior of the technology.

Definition 2: *"AI is a subfield of computer science aimed at specifying and making computer systems that mimic human intelligence or express rational behavior, in the sense that the task would require intelligence if executed by a human."* - Bratteteig & Verne (2018, p. 1-2). This term is used in the author's article that questions if AI makes participatory design obsolete. Compared to McCarthy's definition this one is a lot more recent and focuses more on the human aspects of the technologies behavior.

Definition 3: *"Artificial intelligence leverages computers and machines to mimic the problem-solving and decision-making capabilities of the human mind."* - IBM (2020). This definition from IBM webpages is short and specific. It focuses on the social and human aspects of the technologies behavior, but doesn't explain so much else.

My definition: *"Artificial Intelligence (AI) is the science around computers and machines where the technology interacts with humans and mimics human intelligence, with continually learning from this interaction."* I think AI is a broad spectrum of different important aspects, but want to point out that both the social (human) and technical aspects are two important aspects to include. My definition builds on what I have learned so far in this course with inspiration from the three different definitions above.

1.1.3 Article from the curriculum

I have chosen Bratteteig & Verne's (2018) article *Does AI make PD obsolete? Exploring challenges from Artificial Intelligence to Participatory Design*. The article is, as the name states, about different challenges when it comes to AI and PD. In the article, we learn more about what both Artificial Intelligence and Participatory Design are and the different challenges these two meet, especially when it comes to designing. Some of these challenges can be difficult because of the unpredictable change in AI technologies. Before I read this article I hadn't given much thought about it and wasn't sure what the connection between AI and PD was. After reading it I see that it is an important discussion that is still very relevant, maybe even more for the future designers and how they can use the PD methods in their design process.

1.1.4 IBMs work with AI

IBM is a contemporary company that works with AI and offers a broad range of information about AI on its web pages. This includes everything from definitions, history, and the different types of AI. In this way they talk about AI from different angles, before they present their own product/service; IBM Cloud also called IBM Watson. IBM themselves explains that this product *"gives enterprises the AI tools they need to transform their business systems and workflows, while significantly improving automation and efficiency"* (IBM, 2020).

1.1.5 Interaction between people and AI

Black Mirror - Be Right Back (s2e1)

In this episode of Black Mirror, we get to know a couple, where the guy named Ash dies suddenly in an accident. His girlfriend Martha is grieving and gets to hear about a new service that lets you communicate with the people that are no longer with us with a newly established system. This system is based on AI that uses all the data that is available online or you give it. At first, she's very skeptical but gets signed up anyway. It all starts with a chat, and then the technology evolves to a phone call with the voice-based on videos you share

with the system. Then it hits the next level when the company delivers a human-like robot that is programmed with the AI system to behave exactly like Ash. The AI is very good at listening to and learning from Martha about what Ash was like and not, but there are flaws based on ironic data that Ash posted online.

1.2. Robots and AI systems

1.2.1 The word Robot

The word *Robot* originates from the old Slavonic word “robota” which translates to “forced labor” or “servitude” (Science Friday, 2011). The word *Robota* is connected to the central European system of serfdom where the rent of a tenant got paid in forced service of labor.

1.2.2 Definitions of Robot

Both of the two definitions of the word “robot” I will explain are taken from Thrun’s (2004) article toward *a Framework for Human-Robot Interaction*.

Definition 1: *“a reprogrammable, multifunctional manipulator designed to move materials, parts, tools, or specialized devices through various programmed motions for the performance of a variety of tasks” (Russell & Norvig, 1995)*. This definition specifies what a robot is made to do, which in this case is to “move” different things when performing different tasks. It doesn’t say much about uses in any different settings.

Definition 2: *“An automatic device that performs functions normally ascribed to humans or a machine in the form of a human.” (Merriam Webster’s collegiate dictionary 1993)*. This definition is shorter than the first definition and in contrast, doesn't say anything about a robot being designed to move things. This is a more “open” definition, but an important part of it is that the use of robots replaces tasks humans normally do.

My definition: *“A physical form of a machine that independently can perform different assigned tasks”.* I think both of the definitions I used are a bit outdated when it comes to the technological leap since the 90s. My definition is broader and doesn’t define specific tasks a robot does or that it’s replacing human work, which I think is important. The part about robots being *independent* or with some degree of autonomy is also very relevant when it comes to the robots that exist today.

1.2.3 Relation between AI and Robots

When it comes to the different definitions I have described, we can see that there are some big differences between AI and robots. First, we have intelligence which is very important when it comes to AI. Robots are programmed and can be reprogrammed, but don't have the same intelligence and ability to learn by themselves like AI. Then we have the human perspective which is also different. AI aims to mimic human behavior, whereas robots often perform human tasks. Although these differences, and maybe because of them, there are many examples where AI and robots work together combined. An example of this is the human-like robot from the episode of Black Mirror *(answered in task 1.5)* which uses AI in the way it interacts with other humans.

1.2.4 The Robot lawnmower and human interaction

Robot lawnmowers are a type of robot that has become popular and replaced the “old manually” lawnmower in many settings. These robots move within sections on a lawn defined by wires under the ground. Human interaction can vary, but in theory, these robots are automatic and don’t need a lot of interaction after the installation. They have their own charging stations where they go before they run out of power and automatic timers. But this technology is not flawless and in many settings, it may take more human interaction than what was intended - maybe the robot gets stuck if something is left on the lawn, they run out of power before they get back to charging or you need to adjust the wires.

1.3. Universal Design and AI systems

1.3.1 Definition of Universal Design.

“Universal Design is the design and composition of an environment so that it can be accessed, understood and used to the greatest extent possible by all people regardless of their age, size, ability or disability.” (CEUD, n.d.).

I think this is a good definition when it comes to Universal Design because it is short and right to the point. It is about designing for, and including everyone. Making the design accessible, whatever that means specifically to you. This is especially important when it comes to minorities in society and including everyone by focusing on removing obstacles in everyday life.

1.3.2 The potential of AI

AI has a huge potential for including people in society when it comes to helping with disabilities. An example of this is Siri or other voice assistants powered by AI. This can be a great tool when it comes to challenges with movement and sight. Instead of needing to click precisely on your phone or see the screen, you can use your voice to make a call or send a text message. Self-driving cars are another example where AI can include people with physical or cognitive disabilities that might not make it possible to drive a car.

1.3.3 Including and excluding people with AI

AI has a great potential for including people, like with the examples listed above. Although it is also important to be aware of the dark side of AI with the potential of exclusion. An example of this is software with Face Recognition not working on everyone, especially women with darker skin tones. This was a result of the data AI used to learn was mainly based on the faces of white males (SITNBoston, 2020). Issues like this, which leads to limiting and excluding huge parts of society are very important to address and also a reason why Universal Design is important in AI.

1.3.4 The use of “understand” and “understanding” in the WCAG 2.1 principles and the Human AI-Interaction guidelines

What I make sense of the use of “understand” and “understanding” is that it is critical to *understand* something to make it accessible and including. Both the designer and the user need to have some form of understanding. The designer needs to understand the different types of users and the users need to understand the design. This also includes AI. To the question “*do machines understand?”*, I guess the answer depends. It depends on the setting and on the machine. Maybe AI has some degree of understanding, but can it understand something it has not yet learned?

1.4. Guideline for Human-AI interaction

1.4.1 Microsoft guidelines

Guideline: ***01 Make clear what the system can do.*** *Help the user understand what the AI system is capable of doing.*

An example of this is to in the relevant format show the user what the system does. If it's a Voice Assistant powered by AI, start the first conversation by telling what the system is/does and how to speak so the system starts listening.

1.4.2 HCI design guidelines vs. Human-AI interaction guidelines

The HCI design guidelines I want to discuss are Donald Norman's 7 principles (HCI-06129, n.d.). The core of these is to simplify difficult tasks. When we look at the Human-AI interaction guidelines it is also important to make things *simple* and *understandable*, however, the focus seems to be more on the *system* compared to Norman's principles where the focus is more on the *use* and *user.*

Feedback from the first iteration:

From the first iteration I got feedback on one thing I could improve; to elaborate the answer on HCI design guidelines vs. Human-AI interaction guidelines. Therefore I have explained a bit more in-depth how they focus on the systems and users.

Module 2

2.1. Characteristics of AI-infused systems

In the first lecture of Module 2 with Asbjørn Følstad, we were introduced to some different key characteristics of AI-infused systems. These characteristics include **learning;** that they are dynamic and adaptive to learning from the users' behaviors. **improving**; this is connected to learning and mistakes are inevitable, but the systems can continually learn and improve from the mistakes. **black box**; the systems are opaque - we don't know how the processing happens, which can make them different to understand. **fuelled by large data sets**; the systems learn and improve from different data gathered through interaction with the users, this can be either active or passive.

Amershi et al. (2019) define AI-infused systems as *"systems that have features harnessing AI capabilities that are directly exposed to the end-user"*. I want to present three key characteristics from this article; first, we have uncertainty, then it is inconsistency and behind-the-scenes personalization. Kocielnik et al. (2019) also present some key characteristics of AI-infused systems. The different characteristics he describes are transparency issues (which we can connect with black box), impact by user actions (learning, improving, and fuelled by large data sets), and that they also often are linked to probabilistic. We can see a clear link here with both Amershi et al. (2019) and Kocielnik et al. (2019) characteristics and the characteristics explained in the first lecture. Yang et al. (2020) also explain an important cause to be aware of when it comes to difficulties of designing for user experience; this is because of the characteristics of the AI systems, there is a lot of uncertainty of what the system actually can and will do.

YouTube is an example of one AI-infused system. This is a system that is based on different algorithms which will give you recommendations of what to watch based on previous watch history and preferences. From the interaction the user has with the system, it will use this data (and probably other data that it can get online) to make different recommendations. Here we can see the characteristics of *learning, improving,* and *fueled by large datasets.* The users are affected by these characteristics by (hopefully) getting a lot of relevant videos recommended, but we can also connect this to *black box.* The users don't know *how* the systems do the processing, *when*, or even *what* particular data the system uses in the different algorithms.

2.2. Human-AI interaction design

Interaction design for AI-infused systems

In Amershi's et al. (2019) paper we are presented with 18 guidelines for Human-AI interaction as well as with different examples for applications of these. The guidelines were evaluated and validated in three rounds and have their origin from around 150 different AI-related design recommendations (Amershi et al., 2019). The goal for the guidelines is to design better AI-infused systems that are more human-centric, as well as facilitate more research.

In Kocielink's et al. (2019) paper the focus is on exploring different designs for adjusting the end-users expectations of AI-infused systems. The result of the work is three different techniques of expectation adjustment. The paper also presents for us the effectiveness these have when it comes to acceptance of imperfect AI-infused systems (here an email scheduling assistant) and the improvement in user satisfaction (Kocielink et al., 2019). These findings are meant as a beginning for how user acceptance can be improved by shaping the expectations of the AI systems.

YouTube and Design Guidelines

G12 "Remember recent interactions" and *G13 "Learn from user behavior"*

Both of these design guidelines adhere to YouTube as an AI-infused system. YouTube has both a search history and a watch history. Especially the search history we can compare to a short-term memory like the guideline explained: *"allow the user to make efficient references to that memory"* (Amershi et al., 2019). This we can see in YouTube's search bar, where the most recent search will come up as a suggestion first. In the watch history, we can see similarities with the guideline about learning from user behaviors.YouTube does exactly like the guideline states; *"personalize the user's experience by learning from their actions over time"* (Amershi et al., 2019). This is both from actions that involve watching history, interaction with the system (videos liked, saved, skipped, etc.), and possibly other data YouTube has access to from the user's behavior online.

Can Language models be too big?

Bender et al. (2021) present a critical discussion concerning problematic aspects of textual content and solutions based on large language models. Their arguments are built upon identifying a wide variety of different risks and costs associated with the rush for huge language models (Bender et al., 2021, p. 619). These include for example environmental, financial, and opportunity costs and risk of substantial harms; which can include stereotyping and extremist ideology. Based on this, the paper encourages NLP researchers to take these different risks into consideration in further research, with a focus on the efficiency of tasks rather than only the focus on a huge amount of data (Bender et al., 2021, p. 619).

2.3. Chatbots / conversational user interfaces

Følstad & Brandtzæg (2017) explain several different challenges for HCI when it comes to the design of chatbots, which include; *conversations as the object of design*, *move from user interface design to service design*, and *the need to design for interaction in networks of human and chatbot (AI) actors*. When we see these challenges in the light of the different characteristics of AI-infused systems described in 2.1 we can see the need for new thinking when it comes to designing. In the human conversation interaction with a chatbot, the design focus will now move away from the graphics and more with other usability concerns and the actual conversation flow (Følstad & Brandtzæg, 2017, p. 41).

G1 “Make clear what the system can do” and *G2 “Make clear how well the system can do what it can do”*

When we look at the challenges of chatbots and *conversations as the object of design* (Følstad & Brandtzæg, 2017, p. 40), Amershi et. al (2019) first and s guideline can help. By following these the chatbot can make it clear at the beginning of the conversation what it is able to answer, how well it may be able to answer the user’s questions/engage in a conversation, as well as how the users should respond and talk with the bot to get the most out of the conversation.

References

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.

Bender, E. M., Gebru, T., McMillan-Major, A., & Mitchell, M. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?. *In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency.* 610-623.

Bratteteig, T. & Verne, G. (2018). Does AI make PD obsolete? Exploring challenges from Artificial Intelligence to Participatory Design. *In Proceedings of the 15th Participatory Design Conference: Short Papers, Situated Actions, Workshops and Tutorial. 2(8), 1–5.*

Følstad, A., & Brandtzæg, P. B. (2017). Chatbots and the new world of HCI. *Interactions.* 24(4), 38-42.

Grudin, J. (2009). AI and HCI: Two Fields Divided by a Common Focus. *AI Magazine*. 30(4). 48.

HCI-06129. (n.d.) *Norman's 7 Principles.* Retrieved September 07, 2021, from https://sites.google.com/a/nu.edu.pk/hci-060129/lectures-1/norman-s-7-principles

IBM. (2020). *Artificial Intelligence.* Retrieved September 07, 2021, from *https://www.ibm.com/cloud/learn/what-is-artificial-intelligence*

Kocielnik, R., Amershi, S., & Bennett, P. N. (2019). Will You Accept an Imperfect AI?: Exploring Designs for Adjusting End-user Expectations of AI Systems. I*n Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (paper no. 411). ACM.*

McCarthy, J. (2007). What is Artificial Intelligence? *Computer Science Department, Stanford University.*

Science Friday. (2011, April 04). *The Origin Of The Word 'Robot'.* Retrieved September 07, 2021, from https://www.sciencefriday.com/segments/the-origin-of-the-word-robot/

SITNBoston (2020, October 24). *Racial Discrimination in Face Recognition Technology.* Retrieved September 07, 2021, from https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology

The Centre for Excellence in Universal Design (CEUD). (n.d.) *What is Universal Design*. Retrieved September 07, 2021, from http://universaldesign.ie/What-is-Universal-Design/

Thrun, S. (2004). Toward a Framework for Human-Robot Interaction. *Human–Computer Interaction*. 19(1-2). 9-24.

Yang, Q., Steinfeld, A., Rosé, C., & Zimmerman, J. (2020). Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. *In Proceedings of the 2020 CHI conference on human factors in computing systems (Paper no. 164).*