Individual assignment 1 - module 1

Martwo

1. 1 Concepts, definition and history of AI and interaction with AI

<u>The history of AI</u>

AI had its origin in the 1950s. In 1949 the potential of computing was established where the mathematician Alan Turing had a role as a code breaker. Here it started and they saw the potential of computing and AI. Alan continued to promote and talk about the future of computing as a competition or an equal to human intellect. " i do not see why [the computer] should not enter any one of the fields normally covered by humans intellect and eventually compete on equal levels" (Grudin, 2009, s.49)

The mathematician and logician John McCarthy was the first to use the term artificial intelligence in a workshop in 1956 (Grudin, 2009, s.49) Later between the mid 60s to the mid 70s AI started to become more known. They got funds for AI laboratories, which gave AI researchers financial independence as well as the opportunity to establish AI as a field.

<u>Definitions of AI</u>

John McCarty defined AI in 1956 as: " 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."(IBM Cloud Education, 2020)

Verner and Bratteteig defined AI as: "*AI is a subfield of computer science aimed at specifying and making computer systems that mimic human intelligence or express rational behaviour, in the sense that the task would require intelligence if executed by a human*" (Verner & Bratteteig s.1-2, 2018)

Britannica dictionary has another definition on AI which is: *"the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings"*(Copeland, 2020)

Verner and Bratteteig define AI as a science and a subfield of computer science, with some added functions to do tasks that require the intelligence of humans. McCarty also talks about AI more as a research field science than a system. Britannica talks about what abilities the AI has.

I would say AI is more of a technology with specific abilities instead of a subfield. I will define AI as: "intelligent technology that makes it possible for system and machines to mimic humans intelligence and actions both in physical and non physical products"

Short review of the article Does AI make PD obsolete? Exploring challenges from Artificial Intelligence to Participatory Design

This article is about how AI challenges Participatory design and how new technology poses a new challenge to design

Bratteteig and Verner describe the difference between PD and AI goals and their role. They talk about how their goal is the same, but their approach is different. One of the main discussion points in this article is how AI, since it is programmed to give the best result, can take over the role of PD. Do we need PD if an AI solution is the best possible outcome?

Since AI systems change a lot over time it is difficult to foresee or understand how AI may affect the interaction with users and how it then threatens the role of PD.

Bratteteig and Verner conclude that PD still is important, but argues that there are three challenges that AI poses to PD. One of these is how it is challenging for AI to differentiate between use and learning since AI is trained when being used.

AI in companies

Tesla is a company that uses a lot of AI technology in their cars. On their web page Tesla talks about using "an approach based on an advanced AI for vision and planning, supported by efficient use of interference hardware, is the only way to achieve a general solution for full self-driving and beyond."(Tesla, n.d) They talk about AI as a framework or a service that is a part of a product e.g technology that is a part of a car that can make it self-driven. They have applied new cutting-edge research to train deep neural networks on problems like perception to control e.g bird-eye-view network takes video from all cameras to output the road layout. (Tesla, n.d)

<u>Interaction between people and AI in movies</u>

In Black mirror season 2 episode 1: "Be right back" we meet Martha and Ash. Ash dies early and after his death, Martha gets the opportunity to talk to and interact with a software that mimics her dead boyfriend. The software finds all the information that he has posted on the internet and can say things that he would say. She starts talking to the software by using her phone, because it sounds like him.

In the beginning the interaction takes place through her phone just like a regular phone call between two humans. Later she gets a robot that looks and acts exactly like Ash. The interaction then becomes more and more real like he actually is a human. The robot learns and changes after what she says e.g closing his eyes and breathing when he is "sleeping" to make him more human. Even though he is very like Ash there are small differences she notices, like when he doesn't argue back like a human, but obey an order as a system given a code.

In this episode her interaction with AI is very similar to interaction between two humans because the AI software has so much information and the ability to portray Ash in a very similar way as if he was alive.

1.2 Robots and AI systems

<u>The history of the term robot</u>

The term robot came from the Czech word *"robota",* which means *"forced labour"* or *"serf".* The term was used in Karel Čapek's play called R.U.R from 1920 (Moravec, 2021).

<u>Definitions of robot</u>

Robot Institute of America defined a robot as: *"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"(Sebastian Thrun, 2004, s.11)*

Another definition came in 1993 from the Merriam Webster's collegiate dictionary: *"An automatic device that performs functions normally ascribed to humans or a machine in the form of a human."(Sebastian Thrun, 2004, s.11)*

The first definition talks about a robot as a designed product that is programmed to move and perform a variety of tasks. It is a definition of what a robot is when it comes to the technical part of it. However the second definition talks about a robot as a more electronic type of human where it both looks and acts like a human.

I would describe a robot as: " a physical electronic and autonomic product , programmed with similar abilities as humans" . In my opinion a robot has to be something visible and a physical product that has technology that makes it act like humans. What I think of with similar abilities is the ability to have a conversation with another human including some of the body language so the conversation becomes more fluent. Including the ability to adapt its movement and language based on the interaction with someone else.

The relation between AI and a Robot

I would say that a robot consists of AI technology or is "an AI". But "an AI" technology doesn't necessarily have to be a robot, it can be as part of a bigger system, a phone or a different device. Both a robot and an AI is designed/created to do specific tasks that are similar to humans and mimics humans actions.

Based on my own definition of robot, I would say that a robot needs to be a physical and visible product and since an AI doesn't have to be a visible product I would say that is the biggest difference between them.

An example of a contemporary physical robot

"The Paro seal robot has been used to look at how elderly and people with dementia react to a robot in a nursing home context"(Shultz, Herstad, Torresen, 2018, s.1) This robot looks like a stuffed animal that elderly can interact with by holding and petting. The robot makes sound and small movements that are stimulated by touch as well as the head moves to make it more real.

1.3 Universal Design and AI systems

Definition of Universal design

The UN convention on the rights of persons with disabilities(CRPD) defines that:
"Universal design" means the design of products, environments, programmes and services to

be usable by all people, to the greatest extent possible, without the need for adaptation or specialized design. "Universal design" shall not exclude assistive devices for particular groups of persons with disabilities where this is needed.(UN, n.d)

This definition talks about how we design products so they are usable by all people without any added specialized design. I find this definition to cover the important element with universal design and is defined in a thorough way.

How AI including and excluding peoples

AI is a technology that can be used to include more users by a system more diverse and available for users. This can be by making a system in all of the languages in the world so it is possible for everyone to use the system.

Another example of how to include more users is text to speech is an example of including more users. This technology includes blind people so they still can navigate on web pages and apps by listening instead of reading.

It can be difficult for AI technology to understand and relate to human cognition and emotion and it can be misunderstood. This can lead to exclusion of people in difficult situations in life. Humans have different perceptions of things based on knowledge and earlier personal experience. It is important for best possible interaction between humans and AI systems that the system is based on human perception, emotion and movement. If we look at an AI system as a robot it is easier to apply these human abilities. It is then easier for the human to understand the system since it resembles humans abilities and actions, as well as easier for the human to bond or form a type of relationship with the system.

Examples of AI product that including and excluding people

AV1 robots are made for helping children that aren't able to attend school physically. The robot is the children's eyes and ears so that children still can follow the school course from home. This keeps children from falling behind and losing social contacts. This is an example of how AI robots can include people that in other ways would have been left behind and excluded.

In the video "Weapons of Math Destruction | Cathy O'Neil | Talks at Google" from "Talks at Google" Cathy o`Neil talks about her book Weapon of math destructuction and one example

of how AI technology excludes people. She talks about how a man with bipolar disorder failed a personality test when looking for a job and also found out that he was redlighet from that article. Many companies have the same algorithm that excludes him from getting a job because of his mental health. This shows how AI can exclude people.

Explain "understand" and "understanding".

The concept of understanding is about having the knowledge of something so it makes things easier. e.g if you understand english it is easy to see what an english sign says.

Understanding is also about designing something that is well known, so it is understandable for people and makes it easy to interact. If you understand something, it doesn't necessarily mean that you are understanding everything. To be understanding also is about to understand things in different settings and to use what you have understood in on setting in another

Do machines understand?

Machine understands what it is taught, but it is up to the programmer to teach the machine so it can understand. It is all about the data. Does the machine have enough data it can understand the following actions, but not like us humans can. Robots and machines only learn from data, not from the situation itself. So even though an AI system understands something it doesn't mean it can use that to understand something different.

1.4 Guideline for Human-AI interaction

 Example of one of Microsoft`s guidelines

I chose the first Microsoft guideline "Make clear what the system can do". It is a guideline for before the interaction, an initial guideline that helps the user see what the AI system can do. An example of this can be a text, like a placeholder in text fields that describe what the user can write and the system possibilities.

Difference between HCI design guidelines and Human-AI interaction guidelines

I chose Norman´s HCI guidelines, he calls them the fundamental principles of interaction. Here the focus is on both the emotion users feel when interacting and cognition (Norman, 2013, s10) Normans guidelines HCI is more focused on the the users interaction with

something and a more specific context of use compared to Microsoft's guidelines that has a more broader focus on the user and the society around the user and the context of use. Microsoft has guidelines that not only focuses on the interaction but also what the system should or shouldn't do, error occurring and the effect on the society on a bigger scale.

Feedback

Based on the feedback from assignment 1 wrote more about what "similar abilities as humans" means for me connected to my robot definition. I also tried to include more on how human perception, human perception and human cognition/emotion is related to inclusion and exclusion with AI systems.

Individual assignment 2 - module 2

2.1 Characteristics of AI-infused systems

Key characteristics of AI-infused systems.

In Amershi`s article, an AI-infused system is described as a "system that has features harnessing AI capabilities that are directly exposed to the end user" (Amershi et al, 2019, s.1.)

In Asbjørn følstads lectures he talks about Learning, improving, black-box and fueled by large data sets as key characteristics of AI-infused systems.

Learning: An AI- infused system can learn by doing and will learn based on user inputs. AI-infused often have a lot of errors, but with the ability to learn it can also learn to correct the errors.

Improving: Based on the fact that AI-infused systems can learn from users input and actions, they also have the capability to improve themself, by undoing errors and become a better system based on what it learns from earlier experiences with user interaction.

Black-box: You don't see what is happening inside the system. You just see the result and don't see how it happens. You don't see how the system the system learns and improves or what is going on, you just see the result of it and based on input and output you understand that it has learned and improved.

Fuelled by large data sets: An AI-infused system is fueled by large data sets. The system obtains and stores all of the data from users' input and interaction. Based on all of the data the system can learn and improve.

AI-infused system: advertisement on Facebook

Facebook uses a lot of AI technology and is considered an AI-infused system. They use AI technology to personalize advertisements based on what the user is googling and looking at on the internet. The key characteristics used in facebook related to ads is both learning and improving. The system learns what the user may be interested in based on earlier search history and Facebook will improve the ads by personalizing them and viewing them on the platform. The system is also learning and improving the advertisements when a user presses *"hide advertisement"*(fig. 2) and tells the reason for it(fig 3). The system learns that the user doesn't want this advertisement and will show something else based on the user's action.
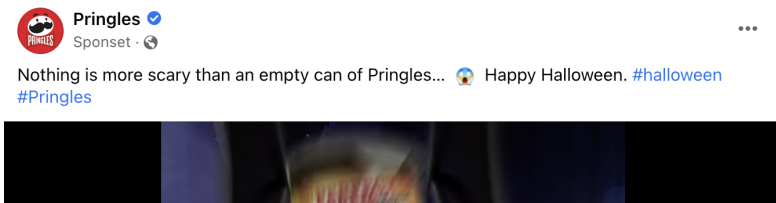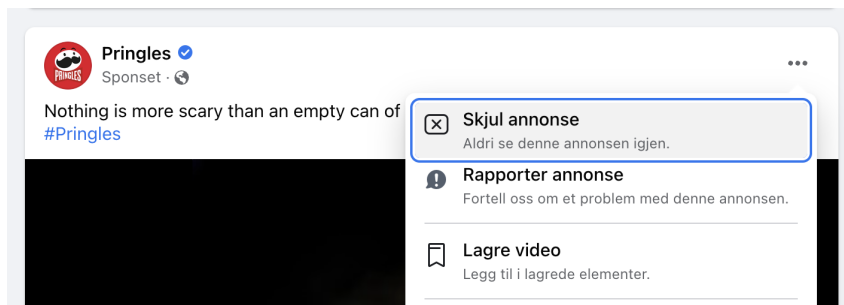

fig. 1 An ad on facebook showing up in newsfeed

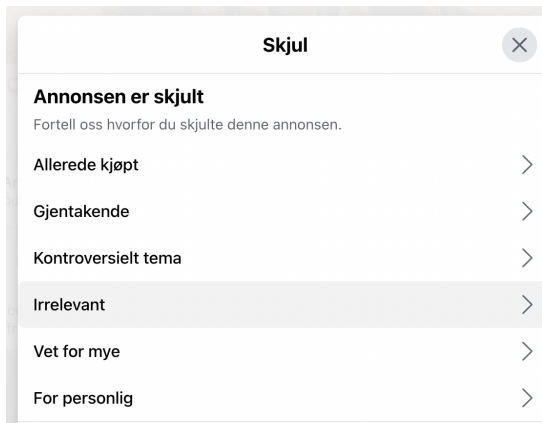Fig. 2: The possibility for the user to hide the ad.

Fig. 3 User can tell the reason for wanting to hide the ad.

2.2 Human-AI interaction design

Amershi et al.(2019):

In this article they create a set of guidelines for user interface design of AI-infused systems. They are testing and evaluating different guidelines and are ending up with x guidelines. The result of their tests shows how relevant the guideline is to offer a spectrum of interaction scenarios. It also shows missing knowledge, highlights and opportunities for further research.

The guidelines was made to be a resource to both designers working on the designing AI infused systems and researchers interested in further development of the guidelines for human-AI interaction design The goal of making these guideline was to get a better and more human centric AI-systems as well as for their research to be a starting point on more research. As AI-technology gets more and more included in the technological landscape they point out the importance of these types of guidelines for the possibility of further develop and refine design guidelines for human-AI interaction.

Kocielnik et al(2019):

This article is about how we can design to adjust the end users expectation of an AI-infused system. It is about designing expectation adjustment techniques so that the user is prepared for imperfect AI-infused systems.

The article presents an experiment that investigates if a user accepts an imperfect system and how to make it easier for users to accept the system.

They explored the scheduling Assistant, which is an AI system for automated meeting request detection in free-text email to look at how to design an AI-infused system so that the end users accept the system.

They designed three expectation adjustment techniques and with experiments they showed their effectiveness in improving user satisfaction and acceptance of an imperfect AI system, the email scheduling assistant. As a result of the experiments that focus on high perceptions of accuracy and decreased acceptance. This opens the possibility to shape expectation as an effective way of improving users' acceptance of AI technologies.

Facebook relation to guidelines

I will look at how facebook ads are related to guideline 13 and 16 from Amershi et al 2019.

G13: Learn from user behavior: the users search history will affect the ads shown in facebook`s newsfeed and will also learn from users behavior if they want to hide the ad.(this is also explained when talking about the system as a learning and improving system). This is important so that the system doesn't make the same mistakes over and over again. This can also weaken the interaction between system and users, including the users likelihood of wanting to use a system.

G16 Convey the consequences of user actions: Related to the possibility for the user to hide ads, the user will not receive the same advertisement. This is something the user gets told so it is visible what the consequences of the user's action are. By making it clear for the user that their actions will have consequences can make the interaction better. This is because the user will acknowledge that the system will change based on user input as well as a confirmation on something that may be important to change for the user. This can be seen on facebook, where the user gets a confirmation that based on the action "hide ad", the ad will not be shown again.

Large language models

One of the key arguments in Bender et al(2021) is that the big language models have a large negative impact on the environment, as well as high financial cost when it comes to model training and development. It is important to keep in mind how the risk and benefits are distributed, when performing risk/benefit analysis of language technology. This is because

both risk and benefits are not distributed to the same place and people. This means that *"negative effects of climate change are reaching and impacting the world's most marginalized communities first"* (Bender et al, 2021, s. 612)

The authors give recommendations to report training time to minimize the carbon footprint. As well as urged the government to invest in compute cloud that provide equitable access to research. Other recommendations are *"run experiments in carbon friendly regions, consistently report energy and carbon metrics, and consider energy- performance trade-offs before deploying energy hungry models."(Bender et al, 2021, S.612)*

The large language model consists of a lot of data that has been proven to contain a lot of problematic characteristics that has resulted in "models that encode stereotypical and derogatory associations such as gender, race, ethnicity, and disability status". (Bender et al, 2021, s 613) This is something bender talks about as a problem due to the large language models. The problematic characteristics in the data have proven to exhibit various kinds of bias, including stereotypical associations or negative sentiment towards specific groups. Human opinion and bias affect the training data, which result in data that differentiate between humans. One of the examples from Bender et al 2021 is that when one of the language model systems are referencing a person with disabilities more negative words are used.

2.3 Chatbots / conversational user interfaces

<u>Key challenges in the design of chatbots / conversational user interfaces.</u>

Some of the challenges connected to the design of chatbots can be to know what the system is capable of doing based on data. This is also related to the first two guidelines in the Amershi et al. article. In the same article they talk about how Ai-infused systems can demonstrate unpredictable behaviors that can be disruptive, confusing, dangerous and offensive. It is also systems that are performed under uncertainty and are often producing false positives and false negatives. This can make it difficult to predict what's right or wrong for the user and you may think all of the outputs from the system are correct even though it may not be.

In one of the lectures Asbjørn følstads talked about the importance of designing a system that makes it easy for the user to understand how to interact with the system so you can facilitate good interaction between the user and the system. It is also important to look at it in a broader perspective and look at the whole network with systems that are related. Ofen it is not only one way of retrieving the information. e.g "kommune-kari" chatbot who gives you information that is already possible to find on the webpage, but there is only less interaction for the user to do.

Guideline G1 and G2 in Amershi et al. (2019)

The first guideline*: "Make clear what the system can do"* is about how to show users what the system is capable of doing so it is easier for the user to understand how to interact. This is something that is important when interacting with a chatbot or similar system because you don't know what the system is capable of to get the right response. e.g if a chatbot only understands numbers, but users only respond with text. The chatbot has to show and make the user understand that you need to use numbers as input to get a more accurate result.

The second guideline *"Make clear how well the system can do what it can do " is* related to the first guideline, but is more about how to help the user to know that the system may make mistakes. This is based on what the system is capable of doing and how well the system is as well as how often the system makes mistakes. This is also important when interacting with a chatbot or similar system because it makes the users aware that the system doesn't always provide the user with the best answer or information. It is then important to show the user that the output from the system may or may not be right.

Feedback

Based on the feedback from assignment 2 I included more on the possible solutions and recommendations the authors presented in the article related to the problematic areas of use of Large Language models.

References:

Amersh, S., Vorvoreanu, M., & Horvitz, E. (2019). *Guidelines for human-AI interaction design*. Microsoft.
https://www.microsoft.com/en-us/research/blog/guidelines-for-human-ai-interaction-design/

Bratteteig, T., & Verne, G. (2018). *Does AI make PD obsolete? exploring challenges from artificial intelligence to participatory design* Proceedings of the 15th Participatory Design Conference: Short Papers, Situated Actions, Workshops and Tutorial - Volume 2, Hasselt and Genk, Belgium. https://doi.org/10.1145/3210604.3210646

Copeland, B. (2020, August 11). *artificial intelligence. Encyclopedia Britannica*.
https://www.britannica.com/technology/artificial-intelligence

Education, I. C. (2020, 3 june 2020). *Artificial Intelligence (AI)*. IBM.
https://www.ibm.com/sa-en/cloud/learn/what-is-artificial-intelligence

Grudin, J. (2009). AI and HCI: Two Fields Divided by a Common Focus. *AI Magazine*, *30*(4), 48. https://doi.org/10.1609/aimag.v30i4.2271

Google [Talks at google]. (2016, November 2). *Weapons of Math Destruction | Cathy O'Neil|*

Talks at Google [Video]. Youtube. https://www.youtube.com/watch?v=TQHs8SA1qpk

Moravec, H. Peter (2021, February 4). *Robot. Encyclopedia Britannica*.
https://www.britannica.com/technology/robot-technology

Norman, D. A. (2013). *The design of everyday things* (Rev. and exp. ed. ed.).Basic Books.

Schulz, T. W., Herstad, J., & Tørresen, J. (2018). Classifying Human and Robot Movement at Home and Implementing Robot Movement Using the Slow In, Slow Out Animation Principle. *International Journal on Advances in Intelligent Systems*, *11*(3&4), 234-244.

Tesla. (n.d). *Artificial Intelligence & Autopilot*. https://www.tesla.com/AI

Thrun, S. (2004). Toward a Framework for Human-Robot Interaction. *Human–Computer Interaction*, *19*(1-2), 9-24. https://doi.org/10.1080/07370024.2004.9667338

UN. (n.d). *Convention on the Rights of Persons with Disabilities and Optional Protocol*.
https://www.un.org/disabilities/documents/convention/convoptprot-e.pdf

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., ... & Teevan, J. (2019). *Guidelines for human-AI interaction.* In Proceedings of the 2019 CHI Conference on

Human Factors in Computing Systems (paper no. 3). ACM.
(https://www.microsoft.com/en-us/research/uploads/prod/2019/01/Guidelines-for-Human-AI-Interaction-camera-ready.pdf)

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 (pp. 610-623). ACM.
(https://dl.acm.org/doi/pdf/10.1145/3442188.3445922)

Kocielnik, R., Amershi, S., & Bennett, P. N. (2019). *Will You Accept an Imperfect AI?: Exploring Designs for Adjusting End-user Expectations of AI Systems.* In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (paper no. 411). ACM.
(https://www.microsoft.com/en-us/research/uploads/prod/2019/01/chi19_kocielnik_et_al.pdf)