Individual assignment IN5620

Module 1

1.1 Concepts, definition and history of interaction with AI and autonomous systems

History of AI

AI can be traced all the way back to the second world war, when Alan Turing wrote about the possibilities of the computer doing human tasks. His development of the code breaker machine and other contributions are viewed as central in the history of AI. "Artificial intelligence" as a term was first used by John McCarthy in 1956 in a call for participation regarding a workshop (Grudin, 2009).

Definitions

First, Wikipedia defines AI as following: *"Artificial intelligence (AI) is intelligence demonstrated by machines, as opposed to the natural intelligence displayed by animals including humans".* (Wikipedia, 2022).

This definition focuses on the technical aspect of AI, where the machines are described as intelligent regardless of human aspects or behavior and rather explained as the opposite of natural intelligence.

As a computer scientist and mathematician, it is no surprise that John McCarthy also left the human aspects out of his definition: *"The science and engineering of making intelligent machines".* (McCarthy, 1956).

Moving on to the third definition, Bratteteig and Verne from the Department of Informatics at University of Oslo, has a different take on the matter. With a background within design, the two defines AI as this: *"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 and Verne, 1-2, 2018).

I find it interesting that this definition not only includes both rational and human behavior, but also explains them as a basis for AI. As I understand it, human and natural intelligence is required for artificial intelligence to exist and be developed because it reflects the same actions.

My own definition is based on reflections regarding the topic, and includes both the technical and human aspects:

"Artificial intelligence is the science of independent machines learning from and acting with natural behavior."

It states that the machines are independent and intelligent by themselves, but that they act based on natural principals. Further they can learn themselves to act and take actions based on their own "brain", through machine learning. This makes them independent.

Human autonomy and machine autonomy

Human autonomy and machine autonomy are similar in the sense of being able to make decisions in a certain context. Given a situation or a problem, both humans and machines have the capacity to choose A or B based something on their own. A difference is what those decisions are based on. For humans it can be feelings or thoughts, but for machines it can be data and algorithms. The human mind and the machine "mind" can both be viewed as a database for further developing and learning, but the difference is that the machine has to be created and programmed at some point to achieve this. This differs from natural human learning processes which we already are born with.

Automatic and autonomous

We talk about automatic and autonomous systems, but is there actually a difference? I believe the concept of autonomic can be explained by something or someone doing something on its own. When we talk about the concept of autonomous, that something or someone are able to continue doing something on its own without any input or interaction with other things or people. The clock can be seen as an autonomous system since it sets

and adjusts itself (Herstad, lecture-notes). Things can have automatic features, like an automatic car, but it doesn't necessarily make them autonomous.

Select one of the articles from the curriculum. Write a brief review from this article. If there is something that is unclear, or you disagree with, or that you question with the article – write about this.

Article

The article I chose was Bratteteig and Verner's *"Does AI make PD obsolete? Exploring challenges from Artificial Intelligence to Participatory Design".* This paper explores the relationship between AI and PD (participatory design), and explains some issues regarding the development and testing of AI systems through involving users in the processes. It touches on challenges such as AI being unpredictable and constantly evolving, which makes it hard to evaluate and test. The way AI systems are somewhat black-box technology for the users will also make it difficult to decide its correctness since the basis for the actions are hidden.

I think this article has many valid points, and I found it quite interesting. I think PD is important for developing high quality AI systems which provides interaction between user and system in a natural and intuitive way, but I still wonder how we are able to do valid tests on the same basis when these systems are completely heterogenous and never acts the same? Maybe the PD process could be adapted and modernized to fit the development of AI technology to ensure user-centered design through participation.

Company

24SevenOffice is a provider of cloud-based ERP- and accounting-systems. They present an AI-assistant as a part of their solutions, where it is described as something that gives the system "superpowers". I would say they market it as a service or a functionality that magically does all the manual tasks for you. (24sevenoffice, 2022).

Film

Ex Machina is a film directed by Alex Garland from 2014, where a programmer, Caleb, wins a contest where he gets to evaluate an AI robot called Ava through the Turing test. The interaction between Caleb and Ava is quite human in the sense that Caleb ends up falling in love with her, and eventually being left alone and trapped for Ava to escape and be free. The interactions are based on the Turing test where Avas consciousness is to be proved. When she shows emotions and human behavior, she is able to use Caleb to escape, but it seems like empathy is lacking since she leaves him in the end without a doubt. The question remaining is if Ava actually is an AI if she doesn't have all human factors. (Garland, 2014).

1.2 Robots and AI systems

Robot - the word

"Robot" comes from the Czech word "robota", which means "labour". It was first used in Karel Capek's play R.U.R, where the plot centerers around the creation of artificial people. (Interesting Literature, 2016).

Robot – definitions

Britannica defines robot as: "*any automatically operated machine that replaces human effort, though it may not resemble human beings in appearance or perform functions in a humanlike manner".* (Encyclopedia Britannica, 2022).

This definition focuses on the fact that robots can be the replacement of human tasks, therefore based on natural intelligence.

IEEE defines robot as: "*an autonomous machine capable of sensing its environment, carrying out computations to make decisions, and performing actions in the real world".* (IEEE, 2020).

This definition on the other hand, leaves out the human aspect. It explains robots more as an individual, highlighting its ability to analyze, compute, decide and act on its own.

My definition would be something like this: “an automatic machine with the ability to do something by itself”.

With a more general definition like this, I think both human behavior and technical abilities can be applied.

AI and robots

Robots and AI are similar in many ways because they are both programmed machines which can calculate, perform and give feedback. They are also both created based on natural behavior to solve natural tasks. AI differ from robots in the sense of self-learning, where AI is more autonomous and doesn’t necessarily need to be updated or reprogrammed. An AI should be able to learn and further develop itself, like humans. Yet, the concept of robots is constantly changing, and we don’t know how these will differ in 20 years. (Dautenhahn, 2018). Perhaps all robots will be AIs in the future?

Contemporary physical robot

The robot-dog Raia is a commercial robot designed to comfort and act as a companion for elderly. It has sensors which registers touch and movement, and gives feedback through sound and movement. When the user talks to her, she will turn her head and bark. When she is touched, she will wiggle her tail and cuddle the user. She also has a heartbeat at all times. (VilMer, 2022).

1.3 Universal Design and AI systems

The UN defines Universal Design as this:

“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”. (United Nations, 2022).

This definition includes the whole society and world that we live in, not only products or functionality. It explains how Universal Design can help minimize the gap between people and the demands of society. I understand it as a way to develop and create open and accessible products with as many groups included as possible.

Potential

AI creates endless possibilities when it comes human extensions such as welfare technology. For example, robots have been tested at Kampen+ care center to assist elderly users with daily tasks (Herstad et al., 2018). Furthermore, AI has the potential to include more users by providing functionality such as voice control and object recognition in the sense of including blind users. Voice control can for example be used to turn the heat up in a smart home, and object recognition can be used as a tool for orientation.

On the other hand, AI can be excluding by discrimination when it comes to minorities. AI makes choices based on what's most likely. For example, a trans woman may be misgendered by a face-recognition functionality based on her features. Non-binary people may also feel excluded if they are placed in a box of either male or female by such technologies.

Understanding

"Understand" and "understanding" in the context of the WCAG 2.1 principles and the Human AI-Interaction guidelines can be used when describing the interaction between AI and people. How an AI system understands a given context or situation can for example be related to language recognition and understanding of speech. Understanding on the other hand, may be related to how it further can evaluate and act on the matter. Understand and understanding can also apply the other way around from the user or designer's perspective, regarding how we understand the AI and its behavior.

1.4 Guideline for Human-AI interaction

I selected guideline 9: **Support efficient correction.**

Make it easy to edit, refine or recover when the AI system is wrong.

An example of this guideline in use could be a pop-up asking the user if the suggested word in a speech-to-text system is correct, especially if the assumption has a higher chance of being wrong. Also, visible and easy-to-find tools for editing and deleting such as symbols and text available at all times would also be a good way to follow this.

Donald Norman, an important researcher within the HCI field, has created a set of design principles (Norman, 1990): visibility, feedback, affordance, mapping, constraints and consistency.

These principles are in the same way as the AI guidelines supposed to help designers and developers create usable and intuitive systems which are easy to use. Keywords such as feedback and visibility can be found in both. When it comes to differences, the HCI principles aim to help designing the system from the very start, while the AI guidelines focuses more on the "afterwork" of displaying and explaining the functions to the user.

Module 2

2.1 Characteristics of AI-infused systems

AI-infused systems are ' systems that have features harnessing AI capabilities that are directly exposed to the end user' (Amershi et al., 2019). Identify and describe key characteristics of AI-infused systems. Draw on the first lecture of Module 2 and three of the mandatory articles (Amershi et al. (2019), Kocielnik et al. (2019), Yang et al., (2020)).

AI-infused systems are diverse and evolving, but there are some key characteristics which usually applies: learning, improving, the black-box concept and the fact that they're driven by big data sets. As presented by Asbjørn Følstad during the first lecture of module 2, these characteristics are to be found in the work of Amershi, Kocielnik and Yang et al.

Firstly, Amershi describes the learning-aspect by pointing on inconsistency due to learning and therefore changing over time (Amershi et al., 2019, p. 2). Yang also mentions the difficulty of designing and building ideas regarding such systems; "The performance of a

deployed AI system can constantly fluctuate and diverge when it gains new data to improve its learning" (Yang et al., 2020, p. 6).

When it comes to improving, AI-infused systems will improve by learning, but are never free of errors. As they learn and change over time, the error-rates will do the same. Yang writes; "The system's performance and error modes are likely to change as it learns from new images and tags" (Yang et al., 2020, p. 9).

The concept of black-box is a well-known analogy for not understanding how the output is generated. A user of an AI-infused system will most likely feel some sort of distance to it due to this lack of transparency. Kocielnik writes about system expectations and how the actual experience usually differs from these. A user's expectations are formed either through external information, understanding of the system or first-hand experience. (Kocielnik et al., 2019, p. 3).

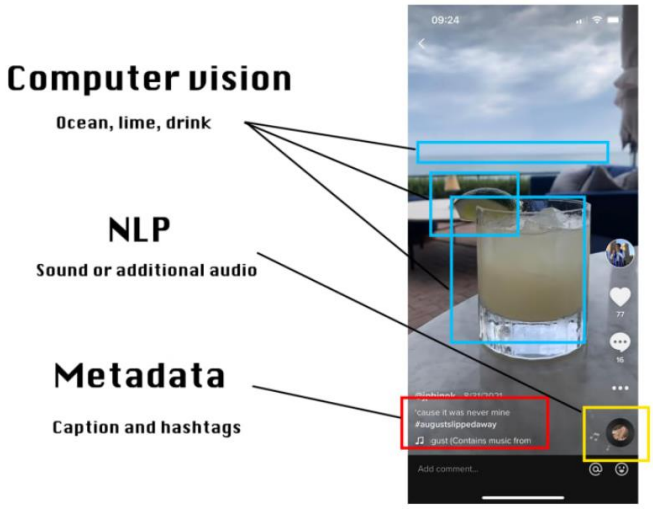
Big data sets fuels AI-infused systems by giving them what they need to work, and further learn and improve. The amount and type of user-generated data determines the capabilities of AI. Yang describes that these systems work "better" for more frequent users and situations, and "worse" for the less consistent ones due to lack of data. "... the system's performance improves for users and use contexts that have produced rich data" (Yang et al., 2020, p. 6).

AI-infused system: TikTok

An AI-infused system that is in my daily life is TikTok. This is an app for entertainment, but is also used for information-spreading and promotion. Once you have created a user, you can begin to scroll through videos with text and sounds in the for-you stream.

The concept of learning is present through how the system uses data about the user to generate a personalized view. Data such as how long you watch certain videos or what sort of creators you click on will be used in this learning process. Then, by analysing how well the personalized stream is fitting the user, the algorithm will improve and become more and more precise. TikTok is therefore fuelled by data, and its performance will highly depend on the amount of data from the user. This is all generated in the "black-box", as users are never

implicitly asked about interests or preferences. As described, all of these characteristics affect the user even though they don't necessarily see it. It can potentially cultivate certain thoughts or ideas that the user already has a slight interest in, both in positive or negative directions, simply because the user is hearing and seeing it so often.



2.2 Human-AI interaction design

Amershi et al. discusses interaction design for AI-infused systems through proposing guidelines validated through evaluation. They point out that AI challenges traditional design principles like "consistency", which needs to be handled differently since AI is quite unpredictable. The guidelines are divided into different usages like "initially", "during interaction", "when wrong" and "over time". (Amershi et al., 2019, p. 1-3).

Kocielnik et al. writes about the end-user's expectations and acceptance of AI-infused systems, and presents designs and techniques for different types of expectation-setting through a study. For example, an accuracy indicator is used to set the user's expectations regarding the system. (Kocielnik et al., 2019, p. 5). The ratio between false-positives and false-negatives is also a subject of discussion in this paper, where the same study concludes that this ratio is dependent on the type of domain and system design and can therefore not be generalized. (Kocielnik et al., 2019, p. 12).

Guidelines and TikTok

G11: Make clear why the system did what it did. Enable the user to access an explanation of why the AI system behaved as it did.

TikTok informs their users that the system provides you a personalized stream of videos based on what you watch, like and share. They explain what it does in general terms, but not really on a more detailed level. The G11 guideline is something that TikTok could take more into consideration, as you don't have access to information regarding your personal algorithm and why you are seeing those specific videos. The user can therefore be left with questions regarding why the system has made these assumptions, but is never given an explanation.

G15: Encourage granular feedback. Enable the user to provide feedback indicating their preferences during regular interaction with the AI system.

As a user of TikTok you are able to provide feedback for each video that pops up in your feed, but only "negative" feedback such as "I don't like this video". From there TikTok will adjust the upcoming videos and filter out similar ones. This is easy and accessible, and gives the user a way to define their preferences. For improvement, "positive" feedback could also be an option.

Problematic aspects of textual content and solutions

Bender et al. discusses problematic aspects of LMs such as stereotypical and discriminating assumptions/biases, as well as environmental and financial costs. They connect the LM's size to these effects, and points out that size doesn't guarantee diversity. (Bender et al., 2021, p. 612-613). They also explore the risks and potential harm which these aspects can cause, like hate speech and dehumanization, and emphasises the need for documentation and accountability regarding LM's training data. (Bender et al., 2021, p. 615-617).

2.3 Chatbots / conversational user interfaces

There are several challenges when it comes to design of chatbots/conversational user interfaces, such as privacy and social aspects. Skjuve explains some of these issues through her project regarding Replika, where different users experienced isolation, “heartbreak” and other personal challenges in addition to many positive effects. (Skjuve, 2022, lecture).

Følstad and Brandtzæg points out ethics and privacy implications, and discusses different challenges within the field. They write about the importance of conversation as the object for design, and also proposes that chatbots should be designed with a service-design approach rather than through the classical user-interaction framework. Chatbot interaction within a big network of both users and other bots is also a challenge, where machine-learning easily can go wrong. Biases and discrimination are also mentioned as challenges of chatbots as they are set up today. (Følstad, Brandtzæg, 2017, p. 40-42).

The guidelines G1 and G2 in Amershi et al. (2019) proposes that chatbots and conversational user interfaces should communicate clearly *what* they can do, and *how* well they can do it to the user. This will possibly decrease the gap between user expectations vs. actual results, which is also discussed by Luger & Sellen (2016, p. 5293). This can improve the user experience and adjust the user's judgement of the system. I believe that this can also minimize the impact of biases and wrong assumptions, since the user will be aware of these possibilities beforehand.

References

Amershi et al. (2019). Guidelines for Human-AI interaction.

Bender et al. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?

Bratteteig, Tone. Verne, Guri. (2018). Does AI make PD obsolete?: exploring challenges from artificial intelligence to participatory design.

Dautenhahn, Kerstin. (2018). Some Brief Thoughts on the Past and Future of Human-Robot Interaction.

Encyclopedia Britannica. (2022). Found at: https://www.britannica.com/technology/robot-technology/The-future

Følstad, Asbjørn. Brandtzæg, Petter Bae. (2017). Chatbots and the New World of HCI.

Garland, Alex. (2014). Ex Machina.

Grudin, Jonathan. (2009). AI and HCI: Two fields divided by a common focus.

Herstad, Jo. Torresen, Jim. Schulz, Trenton. (2018). Classifying Human and Robot Movement at Home and Implementing Robot Movement Using the Slow In, Slow Out Animation Principle.

Interesting Literature. (2016). Found at: https://interestingliterature.com/2016/03/the-curious-origin-of-the-word-robot/

IEEE. (2020). Found at: https://robots.ieee.org/learn/what-is-a-robot/

Kocielnik et al. (2019). Will you accept an imperfect AI? Exploring designs for adjusting end-user expectations of AI systems.

Norman, D. A. (1990). The Design of Everyday Things.

United Nations. (2022). Found at: https://www.fn.no/om-fn/avtaler/menneskerettigheter/konvensjon-om-rettighetene-til-personer-med-nedsatt-funksjonsevne

VilMer. (2022). Found at: https://nettbutikk.vilmer.no/products/digital-terapihund-raia?variant=42271703335074&currency=NOK&utm_medium=product_sync&utm_source=google&utm_content=sag_organic&utm_campaign=sag_organic&gclid=CjwKCAjw1ICZBhAzEiwAFfvFhBJceDvyAVrL78htpmf0pGUvPtUOXOu6HDYo-SVvZ9l-nhjQ0DhUMhoCK08QAvD_BwE

Wikipedia. (2022). Found at: https://en.wikipedia.org/wiki/Artificial_intelligence

24sevenoffice. (2022). Found at: https://24sevenoffice.com/no/erp/ai-bedrift/

Appendix

Working with feedback

I got some feedback that I could ellaborate and discuss some more in module 1, and also use litterature more actively. Due to lack of space, I found it a bit difficult to go more in depth on each task, but I still managed to do some minor changes, e.g in the section where I discuss the article regarding PD and AI.

After module 2 I recived good feedback on my discussion of the papers, and also on the challenges regarding chatbots. Furthermore, I could also further discuss how users of TikTok might be affected by the different AI characteristics, which I wrote a bit on in this third iteration.

In general, I've used this feedback to view my text in a different perspective. It has been a great way to review my own writing, but also to get ideas for new points of discussion.