Individual assignment

Module 1 **1**

1.1 Concepts, definition and history of interaction with AI and autonomous systems 1

1.2 Robots and AI systems 4

1.3. Universal Design and AI systems 5

1.4 Guideline for Human-AI Interaction 7

Sources - Module 1 7

Module 2 **9**

2.1 Characteristics of AI-infused systems 9

2.2 Human-AI interaction design 11

2.3 Chatbots / conversational user interfaces 12

Sources - Module 2 13

Module 1

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

The history of AI

During World War II computers were successfully used to break codes. As a result of this success, after the war, several governments started to fund the building of expensive computers at universities(Grudin, 2009). Consensus seemed to be that these computers could be used for other purposes as well, with Alan Turing even stating that: “I do not see why [the computer] should not enter any one of the fields normally covered by the human intellect (…)” (Grudin, 2009). Several books were later published on the topic of computers and intelligence (Grudin, 2009). Interdisciplinary conferences were also held to discuss these topics related to cybernetics (Grudin, 2009). These were some of the events that eventually would lead to the first appearance of *artificial intelligence* as a term (Grudin, 2009). The term was first used by the American mathematician and logician John McCarthy in a 1956 workshop (Grudin, 2009).

Three different definitions of AI

“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” (Russell et al., 2010 in Bratteteig & Verne, 2018).

The above definition of artificial intelligence was defined by Russell et al in 2010 and referred to in Bratteteig and Verne’s article from 2018. In their article, Bratteteig and Verne use this definition of AI to explore the challenges from AI to participatory design with the research question: “Does AI make PD obsolete?” (Bratteteig & Verne, 2018).

In an article for Britannica Encyclopedia in 2022 B.J. Copeland defines AI as:

"(…) the ability of a computer or a robot controlled by a computer to do tasks that are usually done by humans because they require human intelligence and discernment. Although there are no AIs that can perform the wide variety of tasks an ordinary human can do, some AIs can match humans in specific tasks."

In this article Copeland describes several aspects of artificial intelligence and related terms and concepts, such as 'what is intelligence?', the Turing test etc. (Copeland, 2022).

In return, Merriam-Webster dictionary defines artificial intelligence simply as: "the capability of a machine to imitate intelligent human behavior". This definition of AI seems to be more general, and does not focus as much on the specific type of machine involved for instance.

My own definition of artificial intelligence would be something like: "the ability of a machine to act as though it is just as intelligent as a human being".

The similarities and differences between human autonomy and machine autonomy.

As humans we are able to control ourselves, our own thoughts and actions - at least to a high extent. We have complete freedom to act or function independently. These are some qualities that define human autonomy.

Machine autonomy on the other hand, could be quite different. I would argue that a machine to a much lesser extent is able to control itself (maybe AI is an exception?). Machines are not as dynamic as humans, and do not have the same complete freedom to act as they please.

A brief review of an article

"AI and HCI: Two Fields Divided by a Common Focus" is an article written by Jonathan Grudin in 2009. The article deals with how the histories of AI and HCI intertwines, as the two fields have struggled to coexist. As suggested in the article the differences between the two fields have resulted in periods of great success for one or

the other, but limited signs of joint success throughout their history. While AI has thrived, HCI has struggled, and opposite (Grudin, 2009). The article suggests that one of the reasons for their unsuccessful coexistence is that while HCI is looking for more short-term foreseeable futures, AI is very much about the long-term. It seems to me that when the 'AI forecast' is optimistic, there seems to be no point in focusing on short term HCI problems and solutions - since, this will all be solved by artificial intelligence in the end. The article also suggests that "CHI focuses mainly on technologies when they are first coming into widespread use" (Grudin, 2009, p. 55). Previously, AI was not considered to be of 'widespread use', but in recent years the technology is becoming more apparent (Grudin, 2009, p. 55). This is an interesting observation, as one has to remember that this article was written in 2009. Though the article seems to agree and conclude that the two fields are getting closer together in recent years (Grudin, 2009, p. 55), I would argue that a conclusion written today would suggest that the two fields are even closer - maybe so close that they are starting to overlap.

A contemporary company that work with AI and how this company present AI on their web pages

Intel does more than just the building of microprocessors. According to their website they are also taking on artificial intelligence, stating: "Data-driven insights are transforming our world in wonderful ways. At intel, we're on a mission to open artificial intelligence and data science to more use cases, in more places, for more innovators." Furthermore, they say that they are "clearing the path forward to scale AI everywhere." Intel is clearly very optimistic about increased use of AI, and are talking about it as a service or idea that they want to put forward and into mainstream use. They say that they are on 'a mission' to get this done, implying that this is something that is very important to the company moving forward. (Intel, n.d.)

The Social Dilemma: the use and interaction between people and AI.

The Social Dilemma is a documentary about the use of social media and the creation of related applications. It discusses the statement: "Never before have a handful of tech designers had such control over the way billions of us think, act, and live our lives." (Social Dilemma, n.d.). According to the documentary, our interaction with AI

in relation to social media is consuming our lives - possibly damaging them. Among the damaging effects of AI on social media is the creation of echo chambers.

1.2 Robots and AI systems

How did the word 'Robot' come about?

The word 'robot' first appeared in a play written by Czech writer Karel Capek (Love, 2020). Capek always made it a point to credit his brother, Josef, who suggested the word "robota", meaning 'forced serf labor' (Love, 2020).

Definitions of "robot".

Cambridge Dictionary refers to a robot as a: *"machine controlled by a computer that is used to perform jobs automatically"*.

On the other hand, in 1979, the 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"* (Thrun, 2004).

What I found interesting between these two definitions is that the Robot Institute doesn't refer to a robot as a machine. In addition their definition does not even specify that it has to be controlled by a computer, unlike in the Cambridge Dictionary definition.

I would define a robot as a "machine that performs tasks given by a computer".

The relation between AI and Robots.

One way of differentiating AI from robots is to look at the physical aspects of a robot. The previous definitions of robot listed above refer to it as a 'machine' or 'manipulator'. In some ways it seems a robot is something more physical or at least a thing that has the ability to get different tasks done. On the other hand, AI is referred to as an 'ability' or 'capability', like in the Merriam-Webster definition: "[AI is] the

capability of a machine to imitate intelligent human behavior". Thus, one could argue that AI is a capability that can be bestowed upon a robot. Or in other words, a robot could be capable of performing artificial intelligence.

One contemporary physical robot

The iRobot Roomba 980 is making it much easier to clean up your apartment. This robot vacuum cleaner can connect to an app on your phone. With the app functionality you can order the robot to clean up while you're not even home. The robot completely cleans on its own. In just a push of a button it begins cleaning the desired area. Say you are in a meeting and your colleagues suggest to eat dinner at your place after work, but you know your apartment is a dusty mess. No worries, just go into the iRobot app and turn on your vacuum cleaner. The place will be pristine by the time you arrive home. (Elkjøp, n.d.)

1.3. Universal Design and AI systems

A definition of Universal design

The Norwegian Equality and Discrimination Act section 17 defines universal design as:

"utforming eller tilrettelegging av hovedløsningen i de fysiske forholdene, inkludert informasjons- og kommunikasjonsteknologi (IKT), slik at virksomhetens alminnelige funksjoner kan benyttes av flest mulig, uavhengig av funksjonsnedsettelse." (Bufdir, n.d.)

The facilitation to every user need is addressed in this definition, and shows how Universal Design seeks inclusion.

The potential of AI with respect to human perception, human movement and human cognition/emotions.

AI has the potential of including more users by facilitating use of digital products. Examples of this could be 'text to speech' functionality, where AI technology can turn text into speech to accommodate for blind people. This could also be switched around to include 'speech to text' functionality that accommodates for deaf or hard of

hearing people. In addition AI could be used for image interpretation, to explain images to blind or visually impaired people.

The potential of AI for both including and excluding people

As demonstrated above AI has the potential to be inclusive, but there are also examples where AI has proven to be excluding towards some groups of people. One example is face recognition excluding female, Black and 18-30 year olds. "A growing body of research exposes divergent error rates across demographic groups, with the poorest accuracy consistently found in subjects who are female, Black, and 18-30 years old." (Najibi, 2020).

The concept "understand" and "understanding"

In my opinion the concept "understand" refers to the ability to comprehend knowledge and use it in fitting contexts or situations. I think that in a way machines are only taking direct commands, performing tasks accordingly. I think machines lack the ability to comprehend the whole context of the situation, they are perhaps not as dynamic as humans are. Thus, I think machines would find it difficult to apply something they have learned in other situations than what was intended. However, I do think that more modern and intelligent machines could possibly be able to understand, if not now, in the very near future.

1.4 Guideline for Human-AI Interaction

One of the 18 guidelines for human-AI interaction from Microsoft

One of the guidelines for human-AI interaction from Microsoft states: "Make clear what the system can do" (Amershi, 2019). By this they mean how one should make clear what AI is capable of doing, so that the users are more aware. An example would be letting users know that a chatbot only takes text input and no speech input for instance.

Similarities and differences between design principles and the Human-AI interaction guidelines.

Don Norman's design principles focus on six different concepts (Preece et al, 2002, p. 21). Some of them are similar to the Human-AI interaction guidelines (Amershi, 2019). For example the design principle **visibility** (focuses on the need for functions to be visible to the user) is similar to the Human-AI guideline: "01 - Make clear what the system can do" and "02 - Make clear how well the system can do what it can do". On the other hand the Human-AI interaction guidelines have some guidelines that Don Norman's design principles do not cover. An example of this would be guidelines such as "05 - Match relevant social norms" and "06 - Mitigate social biases".

Sources - Module 1

Amershi, S. et al. (2019, February 1). *Guidelines for human-AI interaction design*. Microsoft. https://www.microsoft.com/en-us/research/blog/guidelines-for-human-ai-interaction-design/

Bratteteig, & Verne, G. (2018). Does AI make PD obsolete? *ACM International Conference Proceeding Series, 2,* 1–5. https://doi.org/10.1145/3210604.3210646

Bufdir (n.d.). *Universell utforming og tilgjengelighet - Universell utforming A-B-C*. Bufdir. Retrieved 14th of september 2022 from: https://ny.bufdir.no/likestilling/universell-utforming/

Cambridge Dictionary (n.d). *Robot*. Cambridge Dictionary. Retrieved 14th of september 2022 from: https://dictionary.cambridge.org/dictionary/english/robot

Copeland, B. (2022, August 24). *artificial intelligence*. Encyclopedia Britannica. https://www.britannica.com/technology/artificial-intelligence

Elkjøp (n.d.). *iRobot Roomba 980 robotstøvsuger*. Elkjøp. Retrieved 14th of september 2022 from: https://www.elkjop.no/product/hjem-rengjoring-og-kjokkenutstyr/stovsugere-og-re

ngjoring/robotstovsuger/irobot-roomba-980-robotstovsuger/ROOMBA980?scid=SEM6279&gclid=Cj0KCQjw2_OWBhDqARIsAAUNTTFwHy1fgZmcMcA7Rhf4kdV_5aUloEGCISWzz4vJxRPxlZETYWh7IMYaAq5GEALw_wcB&gclsrc=aw.ds#features

Grudin, J. (2009). AI and HCI: Two fields divided by a common focus. *The AI Magazine, 30*(4), 48–57. https://doi.org/10.1609/aimag.v30i4.2271

Intel (n.d.). *AI and Data Science*. Intel. Retrieved 14th of september 2022 from: https://www.intel.com/content/www/us/en/artificial-intelligence/overview.html

Love, D. (2020, 03rd May). Where does the word 'robot' come from? *BBC Science Focus*. https://www.sciencefocus.com/future-technology/where-does-the-word-robot-come-from/

Merriam-Webster (n.d.). *Artificial Intelligence*. Merriam-Webster. Retrieved 14th of september 2022: https://www.merriam-webster.com/dictionary/artificial%20intelligence

Najibi, A. (2020, 24th October). Racial Discrimination in Face Recognition Technology. *Harvard University*. https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/

Preece, J., Rogers, Y., Sharp, H. (2002). Interaction Design: Beyond Human-Computer Interaction. *New York: Wiley*

The Social Dilemma (n.d.). *The Dilemma*. The Social Dilemma. Retrieved from 14th of september 2022 from: https://www.thesocialdilemma.com

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

Module 2

2.1 Characteristics of AI-infused systems

Identify and describe key characteristics of AI-infused systems.

Amershi et al. (2019) defines AI infused systems as "systems that have features harnessing AI capabilities that are directly exposed to the end user". Through the lectures on artificial intelligence (module 2) we have learnt that some of these capabilities/characteristics include: *Learning*, *Improving*, *Black box* and *Fuelled by large data sets*.

In short, an AI infused system can be described as a system that is dynamic (learning) and gathers data through interaction (fuelled by large data sets). An AI-infused system is always improving, and in this process mistakes are inevitable. In addition, output from an AI-infused system can be difficult to explain (referring to the Black box characteristic). In the article by Yang et al. (2020) they refer to this 'Black box'-characteristic as:

> "*capability uncertainty* (uncertainties surrounding what the system can do and how well it performs) and *output complexity* (complexity of the outputs that the system might generate)."

Kocielnik et al. (2019) also reflect on some AI imperfections such as False Positives and False Negatives, which again relates back to the statement that AI-infused systems are always improving and mistakes are inevitable.

Identify one AI-infused system which you know well, that exemplifies some of the above key characteristics.

TikTok is an app that is based on artificial intelligence, specifically algorithms that make suggestions based on large data sets gathered from you and all the other users of the platform. The algorithm learns from what you have previously liked and what you seem interested in. It is always improving by suggesting new videos for you to watch, and learning from your response to that suggestion.

Discuss the implications of these characteristics for the example system

The implications of TikToks characteristics include *curated content for the user*; the user gets more and more relevant content based on their previous interaction with videos. I would argue that, as a result of this, users are getting only content the app thinks they will enjoy, disregarding everything else. Potentially this could result in echo chambers and a narrower world view (only what you want to see, not what you maybe should see..).

2.2 Human-AI interaction design

Amershi et al. (2019) and Kocielnik et al. (2019) discuss interaction design for AI-infused systems. Summarize main take-aways from the two papers.

The article *"Will You Accept an Imperfect AI? Exploring Designs for Adjusting End-user Expectations of AI Systems"* by Kocielnik et al. (2019) discusses how different error focuses can lead to very different perceptions of accuracy and acceptance of artificial intelligence by the user. The article also suggests some expectation adjustment techniques that could help the user prepare for AI imperfections, which could then lead to more acceptance.

The article *"Guidelines for Human-AI Interaction"* by Amershi et al. (2019) elaborates on principles for human-AI interaction. In regards to that, they propose 18 design guidelines that are applicable to human-AI interactions. Through many rounds of evaluation Amershi et al. have verified the relevance of these guidelines, revealing gaps in our understanding as well as opportunities for further research.

The main point is that the guidelines could prove useful for designers working with and developing AI-infused systems.

Discuss how the AI-infused system you used as example in the previous task adheres to, or deviates from, two design guidelines.

The AI-infused system, TikTok, adheres to the 13th design guideline: *Learn from user behavior*. By improving video suggestions based on the user's interaction with the content, the system is able to learn and personalize the user experience over time. On the other hand, I would argue that TikTok to some extent deviates from the 11th guideline: *Make clear why the system did what it did*. It is not always easy to get a grasp of how the TikTok algorithm works and why it does what it does. It might not be that clear to the user if it is, for instance, only *likes* that video suggestions are based on. (Amershi, 2019)

Briefly discuss whether/how these two design guidelines could inspire improvements in the example system.

To some extent I believe that the two design guidelines identified above could inspire improvements in TikTok. One could suggest that they inspire improvements in *transparency* towards the user. In other words, the system could try to make it more clear how it learns from user behavior and display this to the user.

Summarize Bender et al. (2021) argument concerning problematic aspects of textual content and solutions based on large language models.

In the article “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?” by Bender et al. (2021) the authors discuss important questions such as: “Are ever larger LMs inevitable or necessary? What costs are associated with this research direction and what should we consider before pursuing it? Do the field of NLP or the public that it serves in fact need larger LMs?”. Furthermore they note some risks with the development of larger language models. These include *environmental costs*

(especially for those who do not directly benefit from the technology), *financial costs*, *opportunity cost* and *risk of substantial harm* (e.g. stereotyping, wrongful arrest etc.). One example of a solution they suggest is to encourage cautiousness - that the NLP researchers should carefully weigh the risks before pursuing this direction.

2.3 Chatbots / conversational user interfaces

Discuss key challenges in the design of chatbots

I have noted several challenges related to the design of chatbots. Among the ones I have discovered is the *complexity of the interaction*. For a chatbot to work smoothly and effectively it has to be dynamic in its conversations. This requires the design and implementation to include all possible answers to questions/responses from the user. In that sense the interaction is complex, as many different outcomes must be considered. In addition, a successful chatbot should (in my opinion) be *intelligent* such as to not disengage the user. This is difficult to achieve, and will require a lot from the design of the chatbot. A key question here would be: *What does it mean to be intelligent?*

Revisit Guidelines G1 and G2 in Amershi et al. (2019). Discuss how adherence to these could possibly resolve some of the challenges in chatbots

In relation to the first guideline: "*Make clear what the system can do"* - several challenges could be addressed. If one makes clear what the system is capable of, the user will lower their expectations and thus not expect the system to be as intelligent as previously thought. In addition the second guideline: "*Make clear how well the system can do what it can*" could be useful in further lowering the user's expectations and specify the system's capability. All in all, adherence to these guidelines would likely improve the user experience and reduce frustration with the system, because the user now knows what to expect.

Sources - Module 2

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)

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). (https://dl.acm.org/doi/abs/10.1145/3313831.3376301)