INDIVIDUAL ASSIGNMENT

Module 1

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

History of AI

As Grudin highlights, the term artificial intelligence (AI) was first used in 1956 in relation to a workshop written by John McCarthy (Grudin, 2009, p.48). However, as discussed in the article, artificial intelligence was around before it was first conceptualized as artificial intelligence. Already during the second world war, we saw examples of artificial intelligence with the codebreaking machine of Alan Turing (Grudin, 2009, p. 49). I believe that this is an example of how AI came from making models based on biology, and human behavior.

Definition of AI

As mentioned above, John McCarthy was the first to use the term artificial intelligence and defines AI 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" (McCarthy, 1998, p.2).

McCarthy distinguishes AI from not only being related to human intelligence, but rather something that involves studying the problems presented to intelligence by the world. In contrast Tone Bratteteig and Guri Verne defines AI as following:

> "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).

With this definition, we understand AI and human intelligence as something more correlated, rather than McCarthy. A similar definition is by Oracle which states that

"artificial intelligence refers to systems or machines that mimic human intelligence to perform tasks and can iteratively improve themselves based on the information they collect (Oracle, 2022). Based on this I would define AI as the science to develop machines that can learn on their own to become more intelligent over time, in many cases simulating human intelligence.

Similarities and differences between human autonomy and machine autonomy

Thrun talks about robotic autonomy as "a robot's ability to accommodate variations on its environment" (2004, p.14). Personally, I would say that robotic autonomy is more manipulated by the user. In contrast with human autonomy, where we are freer to decide our own ethical values and self-determination, to an extent of course.

When talking about something being automatic, I imagine a process that is working by itself, at least as much as possible, without human help. Something being autonomous, I would define as more capable of dealing with its surroundings and working by itself for an extended period. For example, you could push a tell robot to start walking and it would start moving its legs and move across the floor automatically, however, if it also understands that there is an object in front of me and shifts direction it would be understood as mot autonomous.

Article: Ironies of Automation

The article by Bainbridge (1982) is looking into how automation of industrial processes in some cases can lead to expanding problems rather than eliminating them. I chose this article because I found it quite interesting reading about how we are trying to replace human labor with AI but are ending up with humans having to supervise the automated systems and therefore creating a paradox (Bainbridge, 1982, p. 129).

In the second half of the article, several solutions to how we can solve the problems regarding the paradox are presented. Because as Bainbridge states:

> "There will always be a substantial human involvement with automated systems because criteria other than efficiency are involved, e.g. when the cost of automating some modes of operation is not justified by the value of the product, or because the

> public will not accept high-risk systems with no human component" (Bainbridge, 1982, p. 133).

I think this article shows that AI is not always the cheapest and most efficient solution, at least not like it is today. A lot of the automated processes are not automated enough and is still requiring humans to monitor.

AI company

I chose to investigate the Norwegian AI company Simplifai. Their value proposition is "We offer AI solutions in the form of Digital Employees" (Simplifai, 2022). The plug-in-play solution automates the task of translating many unstructured emails and documents between businesses and industries. Further, they advertise this solution to reduce operating costs by 50% and increase efficiency by six times.

Documentary about AI

For this part, I have chosen the documentary *Peter 2.0: the human cyborg* (NRK, 2020). Peter Scott-Morgan had a Ph.D. in robotics, wrote books about robotics, and was a scientist. He believed that any problem could be solved if you were smart enough and had access to cool technology. Therefore, when he got diagnosed with motor neuron disease (MND) he decided to start working towards a solution that could keep him alive despite his illness. He turned himself into the first human guinea pig to see how far we can turn humans into a science fiction world. Slowly he was becoming part human and part computer.

This documentary challenges the idea of what it means to be human – and with his project and new technology, it is a part of redefining the term human and its meaning. He was increasingly turning into a full cyborg. And by adding more and more AI he would turn into a full cyborg. However, Peter said himself that he is never going to stop being a human, but rather challenge the meaning of what it means to be a human.

What I also found quite interesting with this documentary, was how the term interface was now tubes that connect to computers. For example, he got one tube that fed him intravenously, one that got rid of waste, and a bunch of other tubes reading his mind so this could be communicated to others around him. They worked on making an artificial

voice, with the goal of it sounding exactly like Peter. In this way, his personality would still be preserved by the time he would lose his ability to talk. Even though Peter died in June (2022), his way of thinking is interesting to reflect on because this might be a contemporary way to fight motor neuron disease or other diseases.

1.2 Robots and AI systems

The origin of robots

There is some disagreement about when the word robot was first used. Some believe that it originates from the role play RUR in 1920 by the Czech playwright Karel Capek (Adelaide Robotics Academy, 2022). However, there are some that believe it was Karel's older brother, Josef Capek, who first used the word in his short story Opilec in 1917. But the Capek brothers Society in Prague, claims that this is not correct, the word used was "automat" (Adelaide Robotics Academy, 2022). Despite the debate about which brother used the term first, I think we can conclude with it descended from the Capek family around 1920 in central Europe.

Definition of a Robot

Sebastian Thrun presents different definitions of a robot. Firstly, there is the definition from 1979 by the Robot Institute of America that defines a robot as:

> "A programmable, 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, p.11).

I think this definition is describing a specific type of robot, but not robots in general. It reminds me of industrial robots designed to move materials, parts, and tools. I think a robot could be much more beyond that.

Secondly, Thrun presents the definition from 1993 by Merriam Webster that says robots are "An automatic device that performs functions normally ascribed to humans or a machine in the form of a human" (Thrun, 2004, p.11). In contrast to the first definition, this one is more general in what a robot does. However, I also think this only describes some specific robots and not robots in general. For example, I don't think robots

necessarily need to be like humans, just take the robotic lawnmower as an example of this. Based on the two definitions presented I would define robots as physical object that is programmed to interact with their physical surroundings. I don't think it's necessary to define its looks or specific tasks, since there are so many different types of robots.

The relation between AI and Robots

Personally, I find it difficult to distinguish AI and robots a lot of the time. This might be because I often think of robots as something based on AI. However, this is not always the case. For me, one of the elements distinguishing AI from robots is that it does not come in a physical form. If it appears as physical, it is often through robots. Therefore, you can have a robot without AI and AI without the robot.

Beyond their appearance, I also believe that intelligence is different from robots and AI. Robots like Paro that are programmed to do something, in this case, make cuddling sounds and movements do not necessarily understand why it does it. However, a chatbot might have a better understanding that the user wants to know more about this therefore I will provide information about it.

Paro – how a robot moves

Paro is a therapeutic robot that is most used in elderly care. The robot proven to lower stress, anxiety, and pain and help against depression for senior citizens with dementia (Paroseal, 2022). Later the robot has shown results among other "patient" groups, for example, children and adolescents with developmental disorders. Paro is designed to look like a lifelike seal and responds with emotions and gestures. The robot is most known for being cuddled by dementia patients. You can have the robot on your lap, pet it, and the robot will react by moving its head, tail, and eyes and making sounds.

1.3 Universal design and AI systems

Universal design

The Norwegian ministry of children and equality wrote in their action plan on how to have Norway universally designed by 2025 that: "Universal design is an expression of a value put on equality by society" (2009). For me, universal design could be accomplished

by facilitating something to be usable, accessible, and inclusive. Universal design with respect for inclusion would therefore make something accessible for everyone despite the situation.

Potential of AI and its ways of being inclusive and exclusive

It can be difficult to understand the exact goal behind some technology. I think Siri could be an example of trying to include more users by making text input "easier" and more efficient, but in many cases end up excluding users based on language or pronunciation. Another example could be chatbots that don't really understand if you have untraditional or complex questions. The user's struggles to describe could potentially exclude them from getting help. Therefore, I would say that AI's goal is often to make something more efficient and simpler. However, in some cases, this is not always the case. Sometimes, by using AI we extend the gap between expectations and user input.

The concepts understand and understanding

I think that robots can understand their tasks, for example in a way that if X = A the robot understands that Y = B. However, I don't think the robot understands why it's like that – they just know it must be like that. In one way this could be understood as a way of calculating in the sense Bratteteig and Verne describe that "machines cannot reason, only calculate" (2019, p.2).

1.4 Guideline for Human-AI interaction

I chose to look at guideline 11 from Microsoft "Make clear why the system did what it did". This could be seen in relation to one of Schneiderman's eight rules "offer informative feedback" (Microsoft, 2022). Even though they both focus on feedback, they could be different from each other in the way that the guideline from Microsoft focuses more on the technical aspect and explaining the system, rather than focusing on the user's interaction like the one of Schneiderman. Like what Grudin discusses with how AI focused on the algorithms and improving this, we focus on the interaction and user experience when talking about HCI (Grudin, 2009, p.48).

Module 2

2.1 Characteristics of AI-infused systems

Identify and describe key characteristics

Amershi et al. define AI-infused systems as «systems that have features harnessing AI capabilities that are directly exposed to the end user» (2019, p.1). Further, Amershi et al. characterize AI-infused systems as systems that might demonstrate behavior that's unpredictable which can be confusing, disruptive, offensive, or even dangerous (Amershi et al., 2019, p.1). In addition, based on the interaction between the AI-infused system and the user, the AI system is constantly evolving which makes characterizes the AI-infused systems as difficult to design and predict. As a result, Amershi et al. characterize AI-infused systems as uncertain and inconsistent (2019, p.1). This is because the personalization of the AI-infused systems happens behind the scenes and information might be invisible to the users, which is another characteristic of AI-infused systems (Amershi et al., 2019, p.1).

In contrast, Kocielnik et al. focus more on what the users expect from AI-infused systems; therefore, the characteristics will be more based on errors of AI-infused systems (2019). Users' expectations are impacted by the degree of external information, knowledge, understanding, and first-hand experience (Kocielnik et al., 2019, p.2). Therefore AI-infused systems are characterized by having transparency issues, as the users get dissatisfied when the AI-infused system makes mistake, and they do not understand why (p.2). This also relates to another characteristic defined in the article, that AI-infused systems are probabilistic, resulting in the user being disappointed and abandoning the technology (p.2). Based on this, the user's action will affect the user experience, since the systems are dynamic and react to users' input.

Yang et al. also describe AI-infused systems' unpredictability and how should be leveraged, and at the same time limited to minimize their unintended consequences (2020, p.8). Moreover, AI-infused systems are difficult to control, difficult to keep track of, and difficult to understand whom to blame when errors occur (p.2). Which makes it more complex and challenging to figure out how to leverage these weaknesses to mitigate negative consequences, which often is a goal within the HCI field.

In addition to the three articles, the first lecture also highlights some of the characteristics mentioned. One example is how the AI system is a black box, which makes it difficult to understand because you don't understand what is happening behind the scenes (Følstad, 2022, p.39). This can result in users being disappointed by the performance of the system. The lecture also highlights the complexity of these systems, how they are fueled by large data sets, and over time how this evolves through the interaction with users (Følstad, 2022, p.43).

Identify one AI-infused system

For this part, I chose to focus on Instagram as an AI-infused system. I believe that Instagram is a good example of an AI-infused system since it has direct contact with users. Further, the interaction with the user will affect how the AI behaves, regarding recommendations with how posts and accounts are rated in the feed. The way the AI works to adjust Instagram to your interests and persona happens often in the black box, in a way that users don't notice or understand that it happens. Which makes it relevant to what has been mentioned earlier when characterizing AI-infused systems. However, if the AI makes mistake in customizing the feed, the user could be disappointed and frustrated because it does understand why.

Moreover, the chances of persuading the AI are, in some cases, small by watching one reel, profile, or post and it could affect what you get in your feed for some time ahead. Regardless of whether you liked the content or not, the AI only works based on time spent on different content which does mean it understands the users' interests. This also makes it relevant to another characteristic, the AI-infused system is probabilistic. The system could make mistakes in recommending content because it does not understand the motivation for watching it, it could be a one-time thing, curiosity, or a mistake.

2.2 Human-AI interaction design

Summary Amershi et al. & Kocielnik et al.

Kocielnik et al. highlight how there has been conducted a lot of research on machine learning algorithms, but there has not been conducted research regarding methods for facilitating appropriate expectations before the use of an AI-infused system (2019, p.1).

Therefore, they conduct research on different methods of expectations settings for a Scheduling Assistant. The research leads to findings concluding with how different focuses can lead to different perceptions when it comes to accuracy and acceptance. Based on their findings, they design three techniques for expectation adjustment that help the users to prepare for AI imperfections (2019, p.12). The goal is for these techniques to help in increasing users' acceptance of AI-infused systems, which they also prove to some extent in the article with the case of Scheduling Assistant (p.12).

Further, Amershi et al. also discuss how there is a need for more research in the field of AI-infused systems in HCI. Therefore, they propose 18 design guidelines for human-AI interaction (Amershi et al., 2019, p.1). The guidelines have been validated and evaluated through a study where 49 design practitioners tested the guidelines with 20 AI-infused systems. With the findings from the study, Amershi et al. argue that using the guidelines as a guide when designing AI-infused systems, can result in the systems becoming better and more human-centric (2019, p.12).

Based on this, it can be understood that both articles discuss problems with designing AI-infused systems. However, through different approaches, while Amershi et al. focus on developing guidelines according to the characteristics of AI-infused systems, Kocielnik et al. focus on developing techniques to prepare users for imperfections with AI-infused systems.

Design guidelines and Instagram

For this I chose the following guidelines:

G13: Learn from user behavior

G15: Encourage granular feedback

Guideline nr. 13 “Learn from user behavior”, which covers how the users' experience is personalized by learning from the users' actions over time (Amershi et al., 2019, p.3). Guideline nr. 15: “Encourage granular feedback”, covers how the users should be able to provide feedback that indicates their preferences when interacting with the AI systems (Amershi et al., 2019, p.3). I chose both guidelines because I believe that Instagram is an example of adhering to both.

As mentioned earlier, Instagram as an AI-infused system customizes the content of users' feeds based on their behavior on the platform. If you click on content with similarities, then you will get more of this type of content in your feed. Whit this, Instagram is learning from user behavior and adapting its content based on this (Guideline 13). I believe that this is something that increases the user experience because you get a platform that is more interesting based on your interests. However, there is also a chance for it to create an echo chamber for the users, which is a disadvantage.

Further, Instagram has now a feature where the user can give on whether they want to see more of the content or not. If you get a lot of cooking content in your feed because you watched a couple of reels, you can now mark it as something you don't want to see anymore. In this way, the users can give feedback that indicates their preferences when interacting with the AI system. Therefore, I believe that this is an example of how Instagram is adhering to guideline 15.

Summary Bender et al.

Bender et al. (2021) conduct a critical discussion of AI-infused systems based on large language models. They are interested in investigating language models (LMs) and "how big is too big?" (Bender et al, 2021, p. 610). Through their research, they discover several costs and risks with the constant want to enlarge LMs (p. 619). These costs are both environmental and financial costs.

An environmental cost could be the amount of power and energy that is needed to maintain the processing of LMs (Bender et al, 2021, p. 613). In addition, Bender et al. highlight that those affected by these costs are least likely to benefit from the LMs. In addition, a lot of LMs don't understand the meaning behind words and it will require some effort to achieve secure and unbiased LMs (Bender et al, 2021, p. 613). This effort will also be very costly.

Other limitations that are highlighted in the article are, in some cases, how the huge data set of LMs is not beneficial (Bender et al, 2021, p. 616). One of the reasons is that the bigger the data set gets, the more difficult is it to keep it free from bias which will affect the diversity. One example of this is how to filtrate without censoring something that should not be censored. This is relevant to what's mentioned above, how the LMs don't

understand the meaning behind every word. In this case, it could be a word that makes a post censored, but the context of the post is not offensive and should therefore not be censored.

Bender et al. discuss some solutions, one of them being investing more in documenting the large dataset so this is done in a more proper and secure way (Bender et al., 2021, p.617).

2.3 Chatbots/conversational user interfaces

Key challenges when designing chatbots

Even though chatbots are one type of AI-infused system, it requires a different mindset when designing. This is because we experience a shift from focusing on designing the layout to designing the actual conversation that occurs between the user and the chatbot. Without the conversation, a chatbot does not really have a function anymore, therefor this shift is crucial, but it also makes the design process more challenging than designing other systems. Since this conversation is dynamic, we can't predict the whole interaction. Hence, more pressure on designers to design an intuitive chatbot.

Moreover, users often expect more from a chatbot because their perception is that it is smarter than it is. This could be, as discussed by Kocielnik et al., because there is a lack of transparency (2019, p.2). However, technology gets better and better with time, so this might not be that big of a challenge over time. Still, there is an increase in the datasets which also brings in challenges as mentioned above.

I believe that when designing complex systems such as chatbots, there will always be challenges regarding balancing the users' perceptions and technology.

How to solve the challenges?

The guidelines are:

G1: Make clear what the system can do

G2: Make clear how well the system can do what it can do

I believe that one of the major challenges with designing chatbots is for the users to understand what the chatbot can help with. Personally, I get often disappointed when interacting with chatbots for different services because I perceive it as stupid. Most of the time I experience that the chatbot can only help me with easy questions, but not the more complex ones that I need help with. I believe that this misunderstanding happens because it is not clear to me what the chatbot can do (G1).

Moreover, I don't believe that it is clear to users how well the system can do what is can do (G2).

To conclude, how I am familiar with chatbots today, they don't adhere to the two guidelines as much as they should, and adhering to both could help designers in how they should design the chatbot for the best interaction possible. If it is clear what the system can do and how well it can do it, I believe that the user experience would increase because the chatbot's actions will be more compatible with the users' expectations. However, this might be easier said than done, because if the chatbot gets too technical this could have a negative effect on the experience.

3 Appendix

Based on the feedback I got on my previous assignment, I have done the adjustment I got recommended. In addition, I have corrected a couple of typos I found.

4 References

Adelaide Robotics Academy. (2022). *Who did invent the word "robot" and what does it mean?* Roboticsacademy. Retrieved from 12. September 2022 https://www.roboticsacademy.com.au/who-invented-the-word-robot-and-what-does-it-mean/

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)

Bratteteig, T & G, Verne (2018). Does AI make PD obsolete? Exploring challenges from Artificial Intelligence to Participatory Design. Participatory Design Conference. Volume 2, 1-5

Bainbridge, L. (1982*).* Ironies of Automation. IFAC Proceedings Volumes,15(6), 129-135

Følstad, A. (2022). INTERACTION WITH AI AND AUTONOMOUS SYSTEMS – MODULE 2. Retrieved from 7. November 2022 https://www.uio.no/studier/emner/matnat/ifi/IN5620/h22/interacting-with-ai-2022---module-2---session-1---handout.pdf

Grudin, J (2009). AI and HCI: Two Fields Divided by a Common Focus. AI magazine, 30 (4), 48-57.

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)

McCarthy, J. (1998). What is Artificial Intelligence? Available at: http://cogprints.org/412/index.html (Retrieved 11. September 2022)

Microsoft. (2022). *Guidelines for human-AI interaction design.* Microsoft Research Blog. Retrieved from 12. September 2022 https://www.microsoft.com/en-us/research/blog/guidelines-for-human-ai-interaction-design/

Norwegian Ministry of Children and Equality. (2009). *Norway universally designed by 2025.* https://www.regjeringen.no/globalassets/upload/bld/nedsatt-funksjonsevne/norway-universally-designed-by-2025-web.pdf

NRK. (2020). *Forsøksmennesket.* NRK TV. Retrieved from 11. September 2022 https://tv.nrk.no/program/KOID28002718

Oracle. (2022). *What is AI? Learn about Artificial Intelligence.* Oracle. Retrieved from 13. September 2022. https://www.oracle.com/artificial-intelligence/what-is-ai/

Paroseal. (2022). *Home.* Paro. Retrieved 12. September 2022 https://www.paroseal.co.uk/

Simplifai. (2022). *About us.* Simplifai. Retrieved from 11. September 2022 https://www.simplifai.ai/

Thrun, S (2004). Towards a Framework for Human-Robot Interaction HUMAN-COMPUTER INTERACTION, Volume 19, 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). (https://dl.acm.org/doi/abs/10.1145/3313831.3376301)