Individual assignment - third iteration

IN5480 - Fall 2020

Claudia Sikora

Module 1

Concepts, definition and history of AI and interaction with AI

Alan Turing created a sensation in 1949 writing about the computer entering the fields of human intellect in the *London Times* (Grudin, 2009, p. 49). John McCarthy first used the term "AI" in 1956 for a call to participate in a workshop. J. C. R. Licklider wrote an influential essay about AI exploiting computers in 1960 (Grudin, 2009, p. 50). AI research then rose through the 1960s and the researchers had ambitious visions to the technology. In 1970 AI was perceived as negative and lost all the funding (Grudin, 2009, p. 52). AI as a research field has therefore had either ups with strong funding and optimism, and downs or "winters" with little funding and pessimism.

One definition of AI is *"The ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings"* by B. J. Copeland in 2020. By intelligent beings we mean those that can adapt to changing circumstances. Another definition of 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 human intelligence if executed by a human"* by Russell & Norvig in 2010 (Bratteteig & Verne, 2018, p. 1-2). A third definition of AI is *"The theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages"* by the English Oxford Living Dictionary ("Artificial Intelligence", n.d.). The focus is on AI being a subfield of computer science and imitating humans. My definition of AI is *"A computer system that imitate human intelligence and perform tasks based on that".* The reason for this definition is because I believe the intelligence is not real, it is only a simulation. Furthermore the human is a central part of what an AI is based on.

One contemporary company that works with AI is Google AI. They portray AI as helping people everywhere solve big and small problems; AI is making it easier for people to do things every day and provides new ways to look at existing problems (Google, n.d). They also emphasize that everyone should access it and that´s it built with everyone's benefit in mind; *"Advancing AI for everyone".* Their mission is to make AI universally accessible and useful.

They have a program called “AI for Social Good” that focuses on solving humanitarian and environmental challenges.

In episode 1 season 2 in Black Mirror, interaction with AI is portrayed as talking with a phone and then later a synthetic body. The main character´s boyfriend has died in an accident and she finds out she´s pregnant and therefore decides to buy an AI that replicates her boyfriend to feel better. The AI is simulating the dead’s boyfriend personality and voice, but not being able to replicate the small details and obeying to things the real boyfriend wouldn´t. This causes the main character to distance herself from the AI and becoming frustrated. She fails to get rid of it and ends up keeping the AI in the attic, where her daughter visits the AI from time to time.

Robots and AI systems

The word “robot” origins from the 1920s and means “forced labour” in Czech (“Robot”, n.d.). It was first used in a play called “Rossum´s Universal Robots” by Karel Capek that is about a factory making artificial people. Most robots today work in the industry and just perform the same repetitive tasks forever, like assembly and transportation (Thrun, 2005, p. 9). In the future robots will help people more directly in their homes and workplaces.

One definition of a robot is *“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”* by the Robot Institute of America in 1979 (Thrun, 2005, p. 11). This definition focuses on the robot moving things through motions. Another definition of a robot is *“An automatic device that performs functions normally ascribed to humans or a machine in the form of a human”* by Merriam Webster in 1993. Here we see that the definition also includes comparisons to humans. My definition of a robot is *“A physical object moving in an environment that can sense, compute and act based on the environment”.* This definition focuses more on modern robots and the technical part. The reason for this definition is that I believe that the physical part, the environment and the actions are essentials for a robot.

A robot is different from AI because a robot is a physical object while AI doesn´t have to be. A human can interact with an AI through a device, while a robot interacts with the physical environment, including humans, and can be controlled by a human. A robot doesn’t have to simulate human intelligence, while AI has to. They are similar in that both perform tasks and have some degree of autonomy. Nevertheless I would argue that AI in general has more autonomy than a robot. A robot can be completely controlled by a human, while an AI doesn´t have that possibility.

One contemporary physical robot/commercial robot is the Sony AIBO robotic dog (Thrun, 2004, p. 16). The robot has four legs and mimics the moves of a dog, doing different tricks and gestures besides just walking (Sony, n.d.). A human can interact with AIBO through physical contact and speech. The robot shows feelings to humans through its eyes and reacts to what it sees. Humans uses AIBO mostly at home to keep their company like a real pet where they can shape its personality according to their approach. To teach the robot new movements, you can hold it hands.

Universal design and AI systems

A definition of universal design is *"Designing or accommodating the main solution with regards to physical conditions so that the solution may be used by as many people as possible"* by Digitaliseringsdirektoratet in 2020. I understand universal design as being about including everyone regardless of disability. The goal is to not create new barriers and reduce the existing barriers in systems. Therefore we have to respect all kinds of users and not exclude anyone.

The potential of AI with respect to human perception can make interaction more effective and useful because the AI "understands" what´s happening. For example if a human is lost in a place, the AI can effectively understand the situation without asking and help finding the right way. The potential of AI with respect to human movement is prediction of movement so that human can avoid dangerous situations. For example predicting how pedestrians move across pedestrian crossings (Nvidia, 2019). The potential of AI with respect to human cognition/emotions can also help humans avoiding dangerous situations and helping the human to feel better. For example if a human is driving, the AI can notice that the driver is distracted (Zijderveld, 2019). If the human is feeling a certain way, the AI can recommend a specific activity suited to the emotion.

The potential of AI for including people is that the AI can help those with disabilities, learn about people and behave in a suitable way for those specific people. Speech recognition can help those that are visually impaired and eye tracking/detection can help those that are motor impaired. The potential of AI for excluding people is by discriminating individuals, for example by amplifying and demeaning poverty and automating racial bias. Statistics and numbers challenge uniqueness because they favor the majority.

I make sense of the concept “understand” and “understanding” with the process of sensing, processing, making sense of something and knowing something. In my opinion, machines don´t understand; they only simulate understanding. That means that the users can get the impression that machines understand and the machines can therefore still help users understand because of this. Machines can also “learn”, but not in the same way humans do – this is also a simulation.

Guideline for Human-AI interaction

The guideline I have chosen is *“Make clear what the system can do”.* Besides helping users understand what the AI is capable of doing, you can also help the users understand what the system is not capable of doing. That is, initially showing the limitations of the AI system so that users are informed of what is not possible and what can go wrong when interacting with the AI. For example informing the user that the AI is not capable of predicting your age automatically before giving recommendations, so you have to write it in manually.

The set of HCI guidelines I have chosen is Schneiderman´s Eight Golden Rules. The similarities between the guidelines are that both focus on preventing errors, informative feedback, reducing shot term memory load and support control. We can see this when we compare the HCI guidelines to some of the Human-AI Interaction guidelines: *“Support efficient correction”*, *“Make clear why the system did what it did”*, *“Remember recent actions”* and *“Provide global controls”*.

The differences between the guidelines are that the HCI guidelines include consistency and universal usability, while the Human-AI Interaction guidelines include making clear how well the system can do what it can do, learning from user behavior and mitigating social biases. On the other hand, The Human-AI interaction guidelines include some usability with the guideline *“Matching relevant social norms”*.

Module 2

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, p. 1). Some key characteristics of AI-infused systems are that they have inconsistent and unpredictable behavior. This is because they change by learning, are dynamic and they don´t always understand behaviors when nuances are included. External factors can the systems react different, like for instance noise and lightning. Therefore mistakes are another key characteristic, but also the fact that they are constantly improving.

AI-infused systems have functionalities that are probabilistic and imperfect (Kocielnik et al., 2019, p. 2). Imperfect implies that the accuracy is less than perfect. Other key characteristics include uncertainty of what the system can do as well as how well it can do and output complexity (Yang et al., 2020, p. 6). Black box and opaque are another characteristics of AI-infused systems, because it is difficult to understand and validate output (Liao et al., 2020, p. 1). The systems are also fuelled by large data sets, and they gather even more data from the user for each interaction. This also means that these systems can overlearn and begin to behave undesirable.

Siri is an example of an AI-infused system developed by Apple. She is imperfect in the way that she makes mistakes; she misinterprets words and can´t always understand what the user has said. Background noise and the user having a foreign accent can make her struggle even more with understanding what is being asked. Therefore she can be somewhat inconsistent with her answers. This can be annoying and frustrating for users, making it ineffective for them to use her and can thus lead them to avoid using her at all. Sometimes she answers with a funny phrase, making her a bit unpredictable and more human-like. When using her for the first time, it is uncertain what she can do and how well she can do it. After she has learned a bit about the user she will come up with personalized suggestions, but it is not always very clear what these suggestions are based on.

Human-AI interaction design

Amershi et al. (2019) look at user interface design related to AI and propose 18 design guidelines for human-AI interaction. These guidelines are categorized by when they are likely to be applied when interacting with users: initially, during interaction, when wrong and over time. The guidelines focus on clarity, feedback, relevance, support and learning. There are gaps in our knowledge regarding AI, making usability especially relevant.

Kocielnik et al. (2019) look at users expectations towards AI and the acceptance of an imperfect AI. Expectations affect users perceptions on accuracy and acceptance. They use a Scheduling Assistant for automated email meeting requests as an example. The AI has an accuracy indicator that show the expected accuracy of the system, example-based explanation that help the user understand how the system detects meeting requests and giving the user control over AI decisions. This can prepare the user for imperfections and increase their acceptance for the AI.

One design guideline I will look at in regards to Siri is *G11: Make clear why the system did what it did* (Amershi et al., 2019, p. 3). Siri doesn´t really explain why she is behaving like she is, for example when giving the user suggestions for what to do. This can be improved by giving an explanation for the particular behavior and giving specific examples related to this, like the user´s previous actions. Even when the answer is not based on previous user actions, explaining how she "thinks" to the user will help the user trust the system more and avoid misunderstandings.

Another design guideline I will look at in regards to Siri is *G2: Make clear how well the system can do what it can do* (Amershi et al., 2019, p. 3). Siri doesn´t show how often she may make mistakes. This can be improved by her giving recommendations to the user on how the questions should be asked and what to avoid. Also emphasizing that she is learning will indirectly say that she is not perfect, setting the expectations right. Showing some statistics on how accurate the answer is will also help the user. It can also be helpful to let the users know what she is not able to do or what she struggles with.

Chatbot/conversational user interfaces

A key challenge in design of chatbots or conversational user interfaces include conversation as a design object. Users can ask the same question in different ways and have spelling errors, making it difficult for the chatbot to include all possible scenarios, categorize the data correct and also be consistent. Another challenge is that it is necessary to move from UI design to service design. For example Helsebot is a service bot that helps patients, so here you will also have to take into consideration that you are dealing with customers and that the bot should act in a certain way. A third challenge is that it is necessary to design for networks of humans and bots. This means that the chatbot has to fit different contexts and handle a large amount of user data. The language should thus be designed to be appropriate for and understood by many users, and we should be careful with data gathering from users so that the system doesn´t get gamed.

Adherence to the guideline *G1: Make clear what the system can do* in Amershi et al. (2019, p. 3) can help the user understand what the chatbot can help with, making the service more effective. By giving suggestions for what to ask the chatbot as well as explaining what the chatbot is able to do, the user will quicker understand how to use the service and how to use it right. This also includes making clear what the system can´t do, so that the users don't have to find it out themselves and that they have the correct expectations. Also explaining how the system does what it does will help making it clearer for the user.

Adherence to the guideline *G2: Make clear how well the system can do what it can do* in Amershi et al. (2019, p. 3) can help the user get the right expectations towards the chatbot, making the service better and avoiding frustration. By giving guidance for how questions should be asked and not be so determined with the answers, for example by using words like "may" or "think" instead of "will", the chatbot can make less mistakes and the user can be more forgiving with mistakes. How the answers are formulated will thus affect how the users react on the mistakes. This also makes more sense when the AI is not entirely sure on the answer. Showing the accuracy of the system and the information given will help with clarity and certainty.

Module 3

Human AI collaboration

The “Sheridan-Verplank´s levels of autonomy” is a one-dimensional list from human control to computer autonomy where 1 is low level of autonomy and 10 is high level of autonomy (Shneiderman, 2020, p. 1-2). Therefore a two-dimensional framework was developed, showing that it is possible to achieve both high level of human control and high level of computer autonomy at the same time (Shneiderman, 2020, p. 6). On the horizontal axis we find low to high degree of automation, and on the vertical axis there is computer to human control. According to Phillips et al. animals can be used as inspiration for the design of robots because they promote accurate mental models and trust in collaboration with humans (2016, p. 100). The examples of human-robots collaboration I will be looking into are Big Dog robot, Paro robot and Nano robot.

Big Dog resembles a dog and is designed to replace physical capabilities of a human, more specifically to carry cargo to reduce soldier load and move in rough or uncertain terrain (Phillips et al., 2016, p. 104). The autonomy level of this robot is between 4 and 6 because the human instructs the robot in some degree which way it should go (Shneiderman, 2020, p. 2; Phillips et al., 2016, p. 103). The advantages of decreasing the autonomy level of Big Dog are that the robot becomes more predictable in the way it moves and may make fewer mistakes, and the disadvantage is that the human has to put in more time and effort of instructing the robot. The advantage of increasing the autonomy level of Big Dog is that the humans can save more time and effort because they don´t have to control the robot as much, and the disadvantages are that the robot becomes less predictable and may make more mistakes. I would place Big Dog in the upper right quadrant of the framework – high level of human control and high level of automation (Shneiderman, 2020, p. 6).

The current explainability of Big Dog is that it moves where it is instructed and adapts its moves according to the terrain. Therefore I would argue that the robot has a high degree of transparency and safety (Hagras, 2018, p. 29). A suggestion for the needed explainability of Big Dog can be the opportunity to give feedback to reduce frustration and increase trust and acceptance so that the robot can learn from and adapt to its teammates (Smith-Renner et al., 2020, p. 2).

Paro looks like a small seal and is designed to provide comfort and companionship to elderly people to alleviate depression (Phillips et al., 2016, p. 106). The autonomy level of this robot is between 8 and 9 because the robot is simulating a seal and acting alive, learning to behave in a way that the user prefers over time (Shneiderman, 2020, p. 2; Phillips et al., 2016, p. 105). The advantage of decreasing the autonomy level of Paro is that it will become more predictable and consistent in the way it acts, and the disadvantages are that the robot will probably act less alive and be less personalized which can be less comforting for the users and that humans have to take more control of the robot. The advantage of increasing the autonomy level of Paro is that the humans don´t have to bother controlling the robot in any way and the disadvantage is that the robot will always ignore the human, meaning the human has no control in the way the robot acts. I think Paro would be in the upper right quadrant of the framework – high level of human control and high level of automation (Shneiderman, 2020, p. 6).

The current explainability of Paro is that it reacts on input from the user with movement and sound and adapts its behavior to the user, thus it is has some degree of transparency and the possibility of giving feedback to increase trust (Hagras, 2018, p. 29; Smith-Renner et al., 2020, p. 2). The needed explainability of Paro can be related to bias, because the user can abuse the fact that the robot learns from interaction and get it to act undesirable. To increase transparency, Paro would need to display the learning and data in some way – for instance a model (Hagras, 2018, p. 31).

Nano is inspired by a bird and is designed to replace cognitive capabilities by providing additional sensory information needed for scouting and reconnaissance for soldiers (Phillips et al., 2016, p. 107). The autonomy level of this robot is between 7 and 8 because it follows a series of programmable waypoints to stream video imagery (Shneiderman, 2020, p. 2). The advantage of decreasing the autonomy level of Nano is that it can be easier to trust the information given because the human gets more control over the robot, and the disadvantage is that the human has to put in more work instructing the robot. The advantage of increasing the autonomy level of Nano is that the humans can focus on other tasks and the disadvantage is that the robot may be harder to trust. The current explainability of Nano is the fact that it follows the waypoints, so there is some degree of transparency and safety (Hagras, 2018, p. 29). The needed explainability of Nano can be related to feedback for the same reason as Big Dog (Smith-Renner et al., 2020, p. 2).

References

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., ... & Teevan, J. (2019). Guidelines for human-AI interaction. *In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (paper no. 3). ACM.

Artificial Intelligence. (n.d.). *Lexico.* Retrieved the 9th of September from https://www.lexico.com/definition/artificial_intelligence

Bratteteig, T. & Verne G. (2018). Does AI make PD obsolete? Exploring challenges from Artificial Intelligence to Participatory Design. *Proceedings of PDC, 2018, Belgium, August 2018,* 1-5.

Copeland, B. J. (2020). Artificial Intelligence. *Encyclopedia Britannica.* Retrieved the 9th of September 2020 from https://www.britannica.com/technology/artificial-intelligence

Digitaliseringsdirektoratet. (2020, Jun 29). Om oss. Retrieved the 9th of September from https://uu.difi.no/om-oss/english

Google. (n.d.). Google AI. Retrieved the 9th of September 2020 from https://ai.google/about/

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

Hagras, H. (2018). Toward Human-Understandable, Explainable AI. *Computer*, 51(9), 28-36.

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.

Liao, Q. V., Gruen, D., & Miller, S. (2020). Questioning the AI: Informing Design Practices for Explainable AI User Experiences. *In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (paper no. 463). ACM.

Nvidia. (2019, Feb 22). AI Algorithm for Autonomous Machines Can Predict Human Movement. Retrieved the 10th of September from https://news.developer.nvidia.com/ai-algorithm-for-autonomous-machines-can-predict-human-movement/

Phillips, E., Schaefer, K. E., Billings, D. R., Jentsch, F. & Hancock, P. A. (2016). Human-Animal Teams as an Analog for Future Human-Robot Teams: Influencing Design and Fostering Trust. *Journal of Human-Robot Interaction*, 5, 100-125.

Robot. (n.d.). *Online Etymology Dictionary*. Retrieved the 9th of September from https://www.etymonline.com/word/robot

Shneiderman, B. (2020). Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy, arXiv.org.

Smith-Renner, A., Fan, R., Birchfield, M., Wu, T., Boyd-Graber, J., Weld, D.S., & Findlater. L. (2020). No Explainability without Accountability: An Empirical Study of Explanations and Feedback in Interactive ML. *In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20). Association for Computing Machinery*, New York, NY, USA, 1–13.

Sony (n.d.) Aibo. Retrieved the 10th of September from https://us.aibo.com/feature/feature2.html

Thrun, S. (2005). Toward a Framework for Human-Robot Interaction. *Human-Computer Interaction, 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).

Zijderveld, G. (2019, Feb 4). Our Evolution from Emotion AI to Human Perception AI. Retrieved the 10th of September from https://blog.affectiva.com/our-evolution-from-emotion-ai-to-human-perception-ai

Appendix

Based on the feedback from the first iteration, I have explained why I made that specific definition of both AI and robot, and I have included more history of AI.

Based on the feedback from the second iteration, I have explained Siri in more detail and elaborated more around challenges with the design of chatbots/conversational interfaces. I have also included page numbers in the references.