IN5480 - Individual assignment fall 2020

Linda Østerberg (Lindaeo)

Table of content

1 First module

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

Between philosophical attempts to define intelligence and early evolution of computing, is the cradle of Artificial Intelligence (AI) and the emerging of a new research field. In 1949, New York Times magazine published the following controversial words written by Alan Turing, a mathematician, logician, and at that time - leading codebreaker (Grudin, 2009).

“I do not see why [the computer] should not enter any one of the fields normally covered by the human intellect, and eventually compete on equal terms. I do not think you can even draw the line about sonnets, though the comparison is perhaps a little bit unfair because a sonnet written by a machine will be better appreciated by another machine.”

The term AI was introduced in 1956 by an American mathematician and logician named John McCarthy after a workshop at Dartmouth College, Hanover. The road from there has been winding with its fair share of ups and downs. There's been eras of grand visions and generous funding altering periods with crushed expectations (Grudin, 2009).

In the 1960s, AI grew in the spotlight of the academical world as well as ordinary people and support and fundings rising substantially led to a period of financial independence

(Grudin, 2009). Periods were interest as well as fundings where low has been referred to as AI-winters (Hendler, 2008). One famous AI-winter started in 1970s subsequently to an article criticizing the state and lack of progress in the field of AI in UK (Lighthill, 1973).

Definitions of AI

By referring to the following three, amongst the vast variety of definitions of AI, I wish to highlight the pattern related to expectations and perception of the word *intelligence,* starting with John McCarty who coined the term AI. “[...] the science and engineering of making intelligent machines” ... “[where] intelligence is the computational part of the ability to achieve goals in the world” (John McCarthy, 1955). A more recent definition uses *mimic*

human intelligence, which is further from proclaiming that an AI machine possesses human intelligence than McCarthy's definition. Even if that is not clearly outwritten, due to the, at that time contemporary perception of opportunities related to the intelligence of machines, it's easier to read more into it. “AI is a subfield of computer science aimed at specifying and making computer systems that mimic human intelligence or express rational behaviour, in the sense that the task would require intelligence if executed by a human.” (Russell & Norvig 2010)

The last definition is from AI100, an initiative from Stanford University where leading thinkers have been invited to study and investigate influences of AI on people and society. The long-term project includes a wide span of faculties to give a more nuanced perspective. “Artificial intelligence is that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment.” (Stone et.al., 2016)

For now, I chose to focus on that definition of intelligence still debated; that a machine, even when possessing intelligence likeworthy a human, is still not a human and thereby not automatically or maybe even possibly fully equipped with attributes associated with what is commonly perceived as human intelligence.

Artificial Intelligence is the aim to develop a technological based ability to make non-living organisms able to independently act or make rational decisions as a response to input or interaction.

Facebook and the use of AI

To get insight into Facebook's use of AI you need an active investigating approach and it is not necessarily something ordinary users are presented to or aware of in their everyday use. More easily accessed, at the webpages engineering.fb.com and ai.facebook.com, they do however present their research in the field (2020). “Facebook Artificial Intelligence Research (FAIR) seeks to understand and develop systems with human-level intelligence by advancing the longer-term academic problems surrounding AI. Our research covers theory, algorithms, applications, software infrastructure, and hardware infrastructure across deep learning, computer vision, natural language processing, speech, and reasoning. (Facebook engineering, 2020)”. Facebook lifts their contribution and what FAIR brings to the field, while their own

gain from implementing AI is not as equally clear. One could argue that, for Facebook, it's also of essential economical value to understand the needs and patterns of their users.

AI in contemporary movies

The umbrella academy is a Netflix series about seven siblings with different superpowers and their strict adoptive father, who when present, is mostly concerned with preparing them for saving the world. The caretaking and loving part of their updrawn is handled by an AI android robot the children call “mom”. She is embodied as a beautiful woman with a stereotypic housewife look and a kind voice. Her moving pattern is human-like, as well as her ability to express reactions to common emotions by facial expressions. Though it is clear something is missing, and the notion that she is programmed gets present when something unexpected happens. The series explores the inner conflict experienced by the children dealing with emotionally affection for the woman who raised them acting as a loving mother, and their growing notion that she in fact is a robot and thereby not capable of doing more than merely mimicking this kind of human emotions.

1.2 Robots and AI systems

Etymology: The word *robot* originates from the Slavic from *robota* for compulsory labour. The modern use of it can be traced back to the 1920s when the Czech author Karel Čapek used it in a play called *Rossumovi Univerzální Roboti - Rossum's Universal Robots* (“Robot”, n.d).

Definitions of Robot

As mentioned in Sebastian Thrun's paper (Thrun, 2004), the following is the Robot Institute of America’s definition of a robot: “[...] 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”

The Merry webster dictionary states definition of a robot as “[...] a machine that resembles a living creature in being capable of moving independently (as by walking or rolling on

wheels) and performing complex actions (such as grasping and moving objects)" ("Robot", n.d.).

The definitions outlined above describe two very different perspectives on robots, which actually is very descriptive for the subject since robots can come in many forms. The first are focusing on a more industrial type of robot, obviously with the primarily purpose of assisting in physical tasks. The second definition by Merry Webster exemplifies another perspective, a robot that *resembles a living creature in being capable of moving independently.* The second example emphasises a different level of autonomy and complexity and might be closer to a more common contemporary perception of what a robot is.

Based on previously stated definitions, my definition that could be seen as a bit more general is: *A physical embodied technical device that is able to perform tasks based on its capability to compute, sense, and actuate.*

The relation between AI and robots

Even though they are somewhat connected, AI and robots do not define the same thing. In practice, AI is a program, often without a physical embodiment which often is a criterion for an artifact to be defined as a robot. Robots with embedded artificial intelligence is a bridge connecting the two fields. The functionality of embedded AI is however just one a part of a complex robotic system constituting a complete robot.

Contemporary physical robots

Milo, a humanoid robot released in 2013 is an example of robots with embedded AI. It is used for helping children within the autism spectrum to practice recognizing emotions and expressing empathy. He can walk, talk, and even model human facial expressions. There is a touchscreen on his chest displaying icons as he speaks to help the children better understand what he is saying (robots4autism, 2019).

1.3 Universal Design and AI systems

«Universal design» means designing or accommodating the main solution with respect to the physical conditions, including information and communications technology (ICT), such that the general functions of the undertaking can be used by as many people as possible, regardless of disability. (Equality and Anti-Discrimination Act nr 18).

Meaning designing accessible and understandable products for all people regardless of their individual needs. This could specifically entail the inclusion of users with special needs such as cognitive limitations, visual impairment, color blindness, shaking hands or other physical challenges.

One important thought on this subject though, is that most users will sometime during their life be in need of facilitation due to special needs, this might be as simple as that they are using glasses or have a broken arm. Equally important, universal design isn't only about facilitation for people with special needs or challenges. Securing that products live up to a certain standard of usability, makes it more usable for all users.

The potential of AI

AI holds a great potential to contribute in the terms of Universal Design. As of today there are already multiple devices out there helping people with different disabilities. Some good examples of this are: text to speech for people who are visually impaired, advanced spelling program helping with dyslexia and how people with aphasia through Speech synthesis,

There are multiple areas where AI discreetly are supporting the works of humans making their work easier performing working tasks such as sorting mails and files, prioritizing information and thereby helping in reducing cognitive overflow and saving time.

A lot of research has been done on AI with respect to human perception, human movement and human cognition/emotions. E.g. there are robots like earlier mentioned *Milo*, helping children with autism practise recognizing and expressing emotions or AI that support people with other cognitive challenges such as memory loss.

There is also research being done on how making robots move more like humans, making them less alien and easier to approach, by using principles of animation, can help in giving users a better and more genuine experience while interacting with robots (Schulz et al, 2018).

The potential of AI for including and excluding people

Recent years there have been debates on AI and exclusion-related topics, such as racism. Since a machine does not possess the capability to by itself judge *right from wrong* as humans interpret *right from wrong* by its own. This means that a lot of responsibility is put on those designing, developing and training it.

The Human AI-Interaction guidelines in WCAG 2.1 uses the concept understanding, meaning being able to make sense of given information. When talking about AI and machines, I would say that they in a logical aspect are able to understand. The word *understand* could on the other hand also include a more human emphatic perspective which a machine can't have.

1.4 Guidelines for Human-AI interaction

Mitigate social biases is an example of Microsoft guidelines for design interaction with AI. This means making sure that the system does not reinforce some undesirable stereotypes or biases. This is referring to the *during interaction* phase and could in practice mean e.g. not giving an AI artefact a dialect or use of language that work against desirable perception of it.

One famous set of design guidelines for HCI is Donald Norman’s six design principles: visibility, feedback, affordance, mapping, constraints and consistency (Norman, 2013).
I would say Microsoft's AI guidelines are more direct and divided into different phases which Normans more abstract guidelines are not. They are very similar in the way both are handling themes such as feedback, visibility and that the main focus is design for a user friendly product.

2 Second module

2.1 Characteristics of AI-infused systems

AI-infused systems of today are much more common than most people are aware of and exist all around the society, assisting in a large variety of tasks spanning multiple sectors and areas of use. Often the work is performed so smooth and quietly that we do not reflect on how and when AI is infusing systems in our environment. This is particularly hard to define since the definition of AI itself is somewhat floating.

To use a recent definition of AI infused systems Amershi et al. (2019) defines it as "*systems that have features harnessing AI capabilities that are directly exposed to the end user*". The article identifies several characteristics typical for an AI-infused system such as:

- Learning over time
- Changing based on learning
- The reason behind change might be unpredictable and hard to analyze
- Unreliable and inconsistent
- Vary in interaction and capability

As stated by Amershi et al. (2019) these characteristics might cause AI-infused systems to "*demonstrate unpredictable behaviours that can be disruptive, confusing, offensive and even dangerous*". Examples of this could be how sensitive AI systems can respond to new input in the environment, sometimes not even noticeable for limited human senses, or how big, complex collections of data can generate results that were not predicted by humans and thereby not planned for. Also since the logic of AI isn't naturally based on the same rules and norms as human logic, it doesn't possess certain empathic and ethical ground rules that we might take for granted. This results in the fact that AI will not automatically follow human norms for how to behave and interact with its surroundings, this is something that needs to be learned, trained and tested.

As described by Kochelnik et al. in the article: Will You Accept an Imperfect AI? Exploring Designs for Adjusting End-user Expectations of AI Systems; users expectations do affect

their perception and acceptance of a system (2019). Combining that with the challenge in predicting the behaviour of AI-infused systems and as mentioned by Yang et al. (2020) designers struggle to even envision and prototype AI systems. Thereby making it hard to live up to users expectations, when the system cant be properly tested or predicted.

On this background the emerging field of XAI, explainable AI, is addressing the issue of how to make AI and algorithms understandable for users and the need for tools, methodologies and frameworks handling this issue is argued (Liao et al., 2020).

Spotify and their use of AI

Spotify, as a world leading music provider, uses AI and machine learning to adjust their services, driving decisions and acting on data that are collected on users' behaviour. AI has been so smoothly implemented in the system the past few years, that functionality deriving from it, now might be perceived as an obvious part of the system.

Relating to earlier mentioned characteristics typical for AI-infused systems, it is clear that the users now have the opportunity to enjoy more advanced functions enabled by AI, for example the ability to learn and get to know the users preferences has made the use of spotify much more personalized. The recommendations of artists and songs you get from Spotify, are based on what the AI-systems has learned and playlists like “discover weekly”, giving recommendations based on data collected on previous use.

This might all be appreciated functions, making the user experience more personalized, but to problematize it, is it possible to overdo? Is it important that we are aware that the content we are exposed to are personalized and not the same for everyone else? What happens with our ability to search and find our own path, make our own choices? Can an AI really, based on limited data, predict how we would respond to options that has not yet been demonstrated through previous use? Maybe we don't even know it ourselves yet before we have tried it, and now don't get the option to try. This issue is somewhat mentioned by Amershi et al. (2019) lifting differing effects when the users are exposed to so-called false positives or negatives. Some might argue that AI could help us to do just that, try new things. Along with its subtle testing of our preferences, sending out hooks for us to grab onto if we find them interesting, it

gets to know us and with that combined knowledge being able to give us recommendations based on connections we could posible make with our limited human mind.

2.2 Human-AI interaction design

Both Kocielnik et al. and Amershi et al. are arguing a need for extended knowledge on the complexity of designing for AI-infused solutions that are operating directly with end-users, also suggesting techniques and guidelines as strategies from two different approaches.

Kocielnik et al. (2019) refers to previous research regarding negative impacts on user experience as a result of bloated expectations, arguing that it is necessary to explore how to best adjust the users expectations to create better interaction with AI. The authors state a lack of studies exploring methods for setting appropriate expectations before initial use of AI-based systems, and aim to contribute to this area by testing several different expectation setting techniques. They are studying the impact on user acceptance, also designing three techniques for shaping expectations prior to use. These are based on findings showing that "focus on High Precision rather than High Recall of a system performing at the same level of accuracy can lead to much lower perceptions of accuracy and decreased acceptance."

Amershi et al. (2019) stresses the need for advancement research and new clearer guidelines developed for AI-infused interaction design. The present 18 validated human-AI interaction design guidelines, arguing for their relevance e.g. through a user study conducted with 49 participants testing AI-infused products according to the guidelines.

Exemplifying two guidelines from Amershi et al.

- ***G8: Support efficient correction***
  Make it easy to edit, refine, or recover when the AI system is wrong.

- ***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.

So, how do spotify's use of AI adhere to these two guidelines for AI-infused interaction design and could they inspire further improvement? Setting them up against Spotify's

AI-infused interactions design i would say it’s obvious that they already have put some thought into this. They do for example inform that this is recommendations based on previous listened music, often mentioning names and when you don't like a song or style you can correct the AIs impression of that preference by asking not to get recommended that genre or artist again. In general you can do a lot of conscious customizing during the whole useexperience.

2.3 Chatbots / conversational user interfaces

Chatbots are one type of AI-infused systems, and they as well come with their individual challenges. The interaction with a chatbot is based on predefined instructions on how the bot is to respond to unpredictable input typed into the interface by the user. This means that there is a vast variation of alternative outcomes of this interaction. And there is a limited interface to adhere to guidelines, such these two, also from the set of 18 guidelines by Amershi et al. (2019).

- ***G1: Make clear what the system can do***
  Help the user understand what the AI system is capable of doing.

- ***G:2 Make clear how well the system can do what it can do***
  Help the user understand how often the AI system may make mistakes.

It is often not clear what the system can do or why it does what it does, when interacting with conversational interfaces. Examples on how this could be improved is to secure a language that clearly expresses information such as “ have a look at these suggestions based on the destination you asked for that you *might* enjoy.” Making it clear that it is merely suggestions that the user *might or might not* agree with, and that the AI is not perfect. One other example could be to inform about limitations by telling and at the same time asking for appropriate input as in “to give you more precise recommendations within this region I will also need to know which route you will be taking from London ”.

3 Third module

3.1 Robots and animals as team members

The increasing interest in Human Robot Interaction (HRI) has been pushing recent research in new directions, thus new domains are being explored. One example of this is Philips et al. (2016), discussing HRI related to how humans interact with animals serving as a team member, for example in the military or health care services. Since animals historically have been used to support humans in their work and everyday life, their contribution and role as a team member is something most people are familiar with. Therefore, we can say that there are established mental models based on previous knowledge about what different kinds of animals could contribute when performing a certain kind of task. Philip at al. proposes that taking use of these models and presumptions could help demonstrate the possibilities and challenges related to a robot team member (2016). They argue that this analogy could help explain future-robot team relationships with a more realistic view of human robot collaboration than contemporary assumptions related to robots.

Equally to an animal, the robot needs to learn and thus, it first needs to be trained. This is something that might be easier to emphasize when relating the human-robot collaboration to human-animal collaboration. Also when collaborating with other humans, or animals trust between the involved collaborators needs to be established. Philip et al. also lifts aspects such as how humans, when working with animals, even when trust is established, has a certain understanding for how the animal might act *like an animal* and demonstrate unpredictable behaviour (2016, p. 109). This could be helpful in explaining that even if the robot can be trusted to perform the tasks it was trained for at a satisfactory level, there might need to be a certain understanding for how it sometimes could behave unpredictably.

3.2 Robots collaborating with humans

In their article, Philips et al. (2016) presents a taxonomy of different robots that are collaborating and interacting with humans in various ways. Following are two examples coming from that taxonomy.

The BigDog robot

The BigDog robot is a military robot developed in 2005 by Boston dynamics to assist humans, serving as a robotic pack mule. It has four legs, is dynamically stable and has been designed to handle rough terrain not accessible by vehicles. Thanks to its four legs that are equipped with low-friction hydraulic cylinders it has varied movement patterns that include walking, sitting and crawling.

The terapeutic PARO robot

The PARO robot is a caregiving robot whose purpose is to give therapeutic results similar to those from patients interacting with animals. For example it has shown to reduce stress levels and improve motivation. It has five different types of sensors that makes it able to perceive touch, light, temperature and movement. It can learn to act on different desired behavior and responds to different voice input, for example its name or tactile input such as being stroked.

3.3 The levels of autonomy and explainability

Schneiderman (2020) describes different levels of autonomy in his two dimensional model, see figure 1. Historically autonomy has been discussed according to a one dimensional model presented by Sheridan-Verplank. The Sheridan-Verplank model on the contrary to Schneidermans, that puts human control and computer control as opposites and the model thereby spans between them as to polars and Schneidermans model might view a more nuanced realistic perspective of human computer automation.

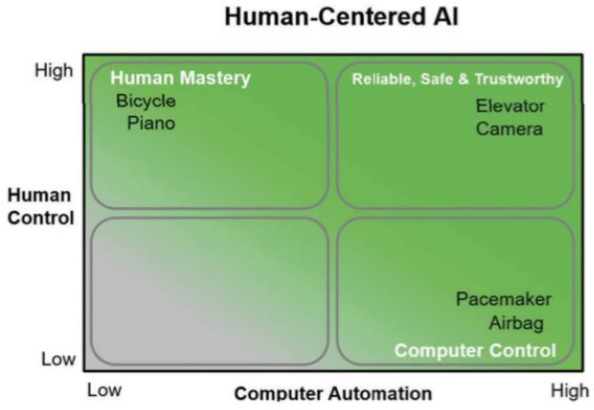

Figure 1. Schneidermans model of automation (2020).

The BigDog robot

In one way the bigDog robot could be described to have a high level of autonomy because of its many sensors that makes it capable of moving and interacting with its environment without constant assistance for every detail from the human controlling it. I would place BigDog in the upper right corner, since it has this capability of moving around but still is controlled and managed by a human. Advantages of making the BigDog robot more autonomous would be that it perhaps could perform its working tasks faster and more efficiently without demanding time capacity from a human telling it where to go.
Making it more autonomous would create even higher expectations on security since it is a big machine and potentially dangerous for humans. An even higher level of explainability would thereby also be needed for making it safe and trustworthy for humans controlling it and operating in its environment.

The PARO robot

Thes sensory information and input that the PARO seal robot receives makes it able to interact with the user and not every step needs to be controlled. For the users to perceive the PARO, most possible, as a *living* animal it needs to have a high level of autonomy. The interaction needs to be smooth and close to reality, otherwise i think it would lose some of its purpose. If it would have an even higher level of autonomy that could of course make the user experience better, making it even more similar to a real thinking animal, though the positive results from use indicates that that might not be necessary to get the wanted therapeutic effect. When it comes to explainability and the PARO seal, I would say the case is a bit special partly because of how, as mentioned by (Smith-Renner, 2018) “Complex machine learning (ML) models can be incomprehensible for end users who are not ML experts”, and maybe not necessary for enhancing this user experience. There would need to be a balance between improving the insight for one group of users, for example the cre takers that assist and the patients that might not need to be reminded of that it is an AI system and not a seal.

References

Module 1

Equality and Anti-Discrimination Act. (2017). Act relating to equality and a prohibition against discrimination (LOV-2019-06-21-57). Retrieved from https://lovdata.no/dokument/NLE/lov/2017-06-16-51

Facebook AI. (2020). Homepage. Retrieved from https://ai.facebook.com/

Facebook Engineering. (2020). Homepage. Retrieved from https://engineering.fb.com/

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

Hendler, J. (2008). Avoiding Another AI Winter. *IEE Intelligent Systems,* 23(2), 2-4. https://doi.org/10.1109/MIS.2008.20

Kerstin, D. (2018). Some Brief Thoughts on the Past and Future of Human-Robot Interaction. *ACM Transactions on Human-Robot Interaction,* 7(1). https://doi.org/10.1145/3209769

Lighthill, J. (1973). Artificial intelligence: a general survey. *Artificial intelligence: a paper symposium*, Science Research Council. 1-21.

Norman, D. (2013). The Design of Everyday Things. Basic Books. Robot. (n.d.). In *Merriam-Webster.com dictionary*. Retrieved from https://www.merriam-webster.com/dictionary/robot

Robots4autism. (2020). Meet Milo. Retrieved from https://robots4autism.com/milo/

Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., Hirschberg, J., Kalyanakrishnan, S., Kamar, E., Kraus, S., Leyton-Brown, K., Parkes, D., Press, W., Saxenian, A., Shah, J., Tambe, M., & Teller, A. (2016). Artificial Intelligence and Life in 2030: One Hundred Year Study on Artificial Intelligence. Stanford University. http://ai100.stanford.edu/2016-report

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

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

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 ACM.

Liao, Q. V., Gruen, D., & Miller, S. (2020, April). 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.

Yang, Q., Steinfeld, A., Rosé, C., & Zimmerman, J. (2020, April). Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. In Proceedings of the2020 chi conference on human factors in computing systems

Module 3

Phillips, E., Ososky, S., Swigert, B. and Jentsch, F. Human-animal teams as an analog for future human-robot teams, Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Vol 56, Issue 1, (2016) pp. 1553 – 1557 DOI: https://doi.org/10.1177/1071181312561309

Shneiderman, B., Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy, arXiv.org (February 23, 2020). https://arxiv.org/abs/2002.04087v1 (Extract from forthcoming book by the same title)

Smith-Renner, A., Fan, R., Birchfield, M., Wu, T., Boyd-Graber, J., Weld, D.S., and 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. DOI: https://doi.org/10.1145/3313831.3376624

Appendix 1: peer-review adjustments

After module 1

Based on advice from my pair-reviewer I added some explanation of the robot definitions and made it a bit more comparative, focusing more on the differences between them. I also added some thoughts on universal design to demonstrate my understanding of its role in design and how AI could be a helpful tool in including different types of users.

After module 2

I extended the explanation of why AI can be unreliable and inconsequent and gave some more examples of this. I related the section about Spotify more to the characteristics and discussed how use of the guidelines could inspire improvement of Spotify's services,