

Individual assignment – Iteration 1 & 2

Module 1

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

Origin of AI

According to Grudin, AI was already developed during the second world war by mathematician Alan Turing through his codebreaking machine. He famously spoke of how he didn't see a reason for intelligent computers not to cover fields that previously and currently are fueled by human intelligence. This thought, among others, is seen as a main contributor towards the growing interest in the AI-field in the late 1940s (Grudin, 2009).

However, the term “Artificial intelligence” or “AI” was first used in a workshop by another mathematician by the name of John McCarthy in 1956. Furthermore the growing field of AI was and are intertwined with the HCI-field in both research and interest (Grudin, 2009).

Definitions of “AI”

McCarthy is often called the “father of AI”, and according to his webpage his updated definition of AI is: *“It is 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). Here, in 1998, McCarthy explain that AI not only is about developing intelligent programs, but that it also is related to understanding human intelligence. However, the two definitions following build their definition around AI about making systems that mimic human intelligence.

The definition by Merriam Webster as of 2021 *“an area of computer science that deals with giving machines the ability to seem like they have human intelligence.”* (Merriam Webster, 2021) is an example of a short definition that focuses on the fact that AI-systems is developed with the purpose of recreating human intelligence and/or behavior.

The third definition by Tone Bratteteig and Guri Verne goes like this: *“AI is a subfield of computer science aimed at specifying and making computer systems that mimic human intelligence or express rational behavior, in the sense that the task would require intelligence if executed by a human.”* (Bratteteig & Verne, p.1). An interesting part of this definition is the inclusion of the intelligence required to execute a task as if the AI-system was a human, which specifies the need of mimicking human, rational behavior.

When making my own definition of AI I wanted to combine the three mentioned definitions. Based by those definitions, I would define AI *as a field of computer science that focus on developing and engineering systems that is able to mimic human intelligence and behavior in for instance problem-solving or task-execution.*

Review of “Does AI make PD obsolete?” by Tone Bratteteig and Guri Verne

The article “Does AI make PD obsolete” is about how AI challenge PD and the role of PD in a future of AI research. Bratteteig and Verne voiced their concerns that the “goal” with AI – to customize services after their users and to make their life easier - is something that possibly could make PD unnecessary, since PD in many ways has the same purpose. A remark originally made by Holmquist is that since PD follows and includes the user throughout the design process, they likely end up with a finished product already custom made for them. In contrast, AI is always evolving by the data they collect during the use of the AI-system, and this means that a large part of the training falls on the user after purchase. This also make the design and development of the AI hard to predict by the designers, as it is difficult to properly understand how the system works. In their concluding remarks the authors pressure that even though AI indeed challenge PD, the methods used in PD could be useful in the design process of future AI-systems.

Grammarly’s as a contemporary company and their view on AI as a service

The company Grammarly both works with and sells a service involving AI-systems. Grammarly provide their users with what they describe as an “AI-powered writing assistant” (Grammarly, 2021). They further present their services to as a way to help you (the user) express yourself better with the help of their AI-systems, both through spellchecking and by suggesting other wordings and synonyms depending on who the receiver of the writing is. A point they also make it that their service provides more than just regular spellchecking because of their use of AI-systems, and that their service is meant to be seamless so that your life is made considerable easier after installing the service.

AI presented in “The social dilemma”

The Netflix documentary “The social dilemma” mention how AI is being used to gain and build fitted content based on which social media(SM) posts you watch, like or share. One of their main focuses is that we, the users, are the dataset and our continuous input through interaction with the app is in a sense the training set for the AI. They make AI out to be a tool

to keep users exploring and using the SM app – and in some ways brainwashes the users by providing a continuous flow of information – or one-sided information based on what we earlier have taken an interest in, rather than giving us a more elaborate picture of the situation. An example is that if you search for or like a picture of a knitted sweater on Instagram, you soon have your whole explore-feed filled with new pictures of knitwear.

Furthermore, the people in the documentary is called whistleblowers, which sets a rather dark tone, and makes it seem a bit like a warning about what SM has become and they fear it will be in the future.

1.2 Robots and AI systems

Origin of robots

The term “Robot” was first used by Karel Capek in his play R.U.R in 1920, and the word has its origin from the Czech word “robota”, which is translated to “forced labor”. The reason behind the choice of the term is because the robots(manufactured humans) in the play were cruelly exploited by their creators which their name was meant to represent (Britannica, 2005).

Definitions of “robot”

The article about HRI by Sebastian Thrun presents two different definitions of “robot”. The first definition was made by the Robot Institute of America in 1979, and describes robots 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, p.11). In my opinion this definition mostly describes industrial robotics, but given the time in which the definition was given this view on robots fit the scope that robots were used. They focus mainly on the tasks the “manipulator” are able to do and that they are able to manipulate their surroundings through these tasks.

As the second definition was based on the Merriam Webster dictionary, I updated the definition from the 1993 version to the 2021 version, and it says: “*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)*”. This definition covers the form of the robot and describes their abilities more humane than the first definition. Instead of focusing on the tasks they rather focus on the autonomy of the machine.

Again, by making my own definition I used the two definitions above as a starting point. I would define robot as *a machine designed to manipulate its surroundings by different*

levels of autonomy. I found it important to both include the purpose of robots and their abilities, and I wanted to include the term autonomy especially since this is an important factor to distinguish between different kinds of robots.

The relation between AI and Robots

I think that one part of the relation between AI and robots is that robots in many cases is based on, or is a product of AI. As AI is based on human intelligence and how we execute tasks, it unites with the perspective that robots also has the ability to execute tasks like humans. A distinction between the two is that AI does not need to have a physical form, like in the definition by McCarthy, AI is not limited to what can be “biologically observable” by humans. I feel like robots are more limited by a physical form, which both definitions above imply since they specified robots as “manipulators” and “machines” rather than systems or programs.

To me the line between AI and robots isn't all that clear, and I feel like the line between them shifts from which definitions of AI-systems and robots you look at. For one I think that the relation between them is strong because I don't think that a system or machine is either based on AI or is a robot, but could be a mixture of both - like chatbots.

How a robot moves – Robot lawn mowers

The robot lawn mower navigates its route by both GPS-tracking and sensors while mowing. Guri Verne describes her relation to her robot lawn mower by the mutual interaction to better the robots performance in mowing the lawn. A robot lawn mower needs a lined up field of work that the users need to line up for them. Most of the interaction with the robot after the initial setup is to clear out possible hindrances, like garden tools or toys that has been left in the grass (Verne, 2020). As the robot don't have the ability to remove obstacles, this work is left to the owner. With no obstacles to clear, the interaction is mostly non-existent as the robot is able to charge itself and tracks its own route based on the restrictions mentioned earlier.

1.3 Universal design and AI systems

Definition of universal design

When looking for a definition on universal design I went to the pages of UUtilsynet, and here is their definition (in norwegian). “*Universell utforming bygger på tanken om at tenester skal vere tilgjengelege for alle, uavhengig av alder, funksjonsevne og utdanningsnivå.*” (utilsynet, 2021). This definition has a clear focus that all digital or non-digital services should be

available to everyone in the way it is designed. What I think it presents is the thought that instead of including people with disabilities as an afterthought, they should be involved from the beginning of the design-process. Inclusive design is meant to “include” all potential users of a product or a service in a sense that they don’t feel forgotten about or left out.

Also, this isn’t only about those who have permanent disabilities, but also those in a certain contexts aren’t able to use the services. To test and evaluate services in different contexts and with users of different demography and background is very important to ensure that more people are able to participate or use the services that is being developed. This consideration will not only locate possible lack of inclusion of certain people, but also how to better the experiences for all future users.

Potential of AI-systems

I think that AI-systems have the potential to enhance the understanding of both human and possible artificial intelligence. To know how to make AI-systems mimic human behavior you also have to better understand human behavior and psychology. For instance, one could possibly develop AI-systems to better understand and include people with disabilities, which also will entail a better understanding of how they behave and use technology.

Furthermore, I think that AI-systems have the potential to both include and exclude people, like any other services or products have – and one thing I find particularly interesting is the way that AI-systems could help us better provide inclusive services. Subtitles to voice-based content is meant to make the content available to more people. A way that AI-systems could better this service is to create subtitles to live content.

Does machines understand?

To “understand” and “understanding something” is about something being comprehensible for someone, and it goes deeper than only knowing how to operate a product, but rather why this action makes a product do something, and the thought behind it. Like Bratteteig and Verne wrote “*machines cannot reason, only calculate*” (Bratteteig & Verne, 2018, p. 2), and while calculating also is a way of understanding something, reasoning goes to grasp the deeper meaning of some actions or reactions.

In this sense I think that while intelligent machines is able to calculate their actions, like the robot lawn mower, their scope of calculations are mostly based on what their creators wanted them to be able to understand from their calculations. I don’t think that it is easy for us to know what intelligent machines understand either, as we mostly only interact and see

their output and not their design rationale. Like in the article by Verne, her realizations of what limited the robot's job in the garden was made by her alone, as the robot didn't necessarily evaluate or calculate every obstacle and give her a message about removing the hinderances.

1.4 Guidelines for Human-AI interaction

Microsoft's Human-AI Interaction design guidelines

I chose guideline no. 1 (Make clear what the system can do). The example made by Microsoft was that this guideline is needed to make the user understand what the AI-system is capable of doing, but I also find it relevant to make the user understand what the systems can't do as well. As mentioned above about understanding – this guideline makes it very important to understand the constrains and limits about what the system offers, so that the use of the system is optimized.

I chose to look at Nielsen's 10 usability heuristics in comparison to Microsoft's guidelines. In general, many of the heuristics are similar with the guidelines in the sense of focus on usability, feedback and prevention of errors. The first heuristics "*Simple and natural dialogue*" (Nielsen, 1994) targets the same area as "*Show contextually relevant information*" and other guidelines, because of its focus on usability and limiting overload of information – especially technical information that is not relevant for the user in the context of their use.

One thing I feel is slightly different in the two sets of guidelines is that Nielsen specify a heuristic about giving the option of shortcuts to the expert users. Although this could be made possible by Microsoft's 13th guideline about learning from user behavior, it doesn't necessarily give the user this option in the first place.

Module 2

2.1 Characteristics of AI-infused systems

Identify and describe key characteristics

Like Amershi et al. describe AI-infused systems, they are “*systems that have features harnessing AI capabilities that are directly exposed to the end user*”(Amershi et al., 2019, p.1 (footnote)). Some characteristics, also made by Amershi, are that AI-systems are constantly evolving based on interaction with the user of the system. This is a subject that is discussed by designers, because this characteristics makes AI-infused systems difficult to design and predict. Based on this they describe AI-infused systems as uncertain, inconsistent, and that AI-infused systems personalize content behind the scenes - so that some information might get hidden for the users (Amershi, 2019).

Yang et al. further build on this perspective, and they describe that the quality of unpredictability also could lead to unwanted societal impact, because errors could happen. Because of the complexity of AI-infused systems and their unpredictability, it is also difficult to mitigate the possible mistakes or negative consequences – like HCI designers often try to do when developing other systems or interfaces. Yang et al. through Figure 1 give AI-infused systems some characteristics, such as difficult to control, difficult to explain to users, difficult to keep track of (or as they put it “make sure is not creepy”) and difficult to “place the blame” for AI errors (Yang et al, 2020).

One of the main themes in the first lecture, is the focus on capabilities and characteristics of narrow AI- systems. Asbjørn Følstad described characteristics as dynamic learning, makes inevitable mistakes but is improving, is a black box (difficult to understand what happens behind the scenes), and is fueled by large data sets which is further fueled by interaction with users.

On the other hand, in the article by Kocielnik et al., focus more about the expectations of the user of the AI-infused systems (Kocielnik et al., 2019). Characteristics from this article, I think, are more aimed at the errors of AI-infused systems, and that they are not perfect yet (probabilistic). In my opinion, one way they are describing AI-infused systems is that they aren't transparent enough in the way that they work. This is made out to be one of the reasons as to why user satisfaction is bad when a AI-infused system makes mistakes. They also present results based on user earlier actions, and give user generated content (Kocielnik et al., 2019).

Identify one AI-infused system

I chose to look further into Youtube as a AI-infused system. I think that Youtube fit the description as a AI-infused because of its AI-capabilities that have direct contact with users. One of which I know of is the recommendations and how they are a result of AI-capabilities, and this capability is the one I will focus on when looking at Youtube.

Furthermore the recommendations fit more of the characteristics above, such as, little transparency and black box-issue when recommending content, recommending content based on earlier user actions, and probabilistic because all the videos aren't always spot-on.

The implications of these characteristics are that a user would always get recommended videos based on earlier watching's. In Youtube's case, I think that custom selected videos after your preferences are part of their "marketing", and therefore comes as no surprise for users. A downside is that possible good content for the user isn't included because of their most recent searches takes a priority over previous, or not searched yet themes. This falls under what Amershi et al. pointed out about how important information might get left out because of how AI-infused systems operate.

Another way that could affect the user and the recommended content is how a random video could mess up with the recommendations you get for a long time. Because of Youtube's connection to Google one always are logged on with your profile. For instance if you are logged on your Google account in the browser, and you get a message on Facebook where someone sends you a video (which you have to open with Youtube) – many times this video is counted by Youtube. Then the video, that you didn't choose yourself, is taken into account when searching for new content to recommend.

Also like I mentioned, the recommendations are probabilistic – they aren't always on point. Just because you watched one video of how to change a certain lightening bulb, it doesn't mean that you are interested in watching another video on different lightening bulbs or another "how to" videos. The 'recommenders' make errors in recommendations because they don't know the difference in motive behind the videos – is it a one-time interest or is it something the user are really interested in and want more of.

2.2 Human-AI interaction design

Summary: Kocielnik & Amershi

The two articles both discuss the problems surrounding design of AI-infused systems. They have different approaches, and while Kocielnik et al. focus on a framework to regulate a user's expectations, Amershi et al. focus more on developing guidelines to fit the characteristics of AI-infused systems.

The Kocielnik et al. angles their article around how to control and fit the expectations of users of existing AI-system. They think that the scope of AI technologies is broadening especially in end-user applications, and that the expectations of these technologies makes the perception of most AI-systems to be negative. This is because of how much the users are exposed to AI-infused systems, and that when AI-systems are not perfect, it is really noticeable. Because of the uncertainty of AI-infused systems and their error-making the authors find it relevant to “warn” users of it beforehand and limit mistakes in the use of the system. This article in my opinion highlights the reality of how unpredictability also comes with error making, and that although the AI-field have come far, AI-infused systems aren't 100% optimal or error free.

Instead of changing design guidelines like Amershi et al, Kocielnik et al. instead want to supply user friendly design techniques to minimize the negative outcome from False Positives and False Negatives. The experiment that Kocielnik et al. held tested the three techniques they developed to shape user expectations, and it resulted in positive outcome. The outcome supported the claim that you can adjust user expectations, and that the user satisfaction with AI-infused systems improved by including the techniques.

Amershi et al. on the other hand focus more on the aspect of designing AI-infused systems. Their article is about making and evaluating design guidelines for HCI with AI-systems, where they produce a set of 18 guidelines for designers to follow in the future. The reasoning behind the new guidelines is because of the authors opinion that “*AI-infused systems can violate established usability guidelines of traditional user interface design*” (Amershi et al, 2019, s.1).

The example they highlight is the principle of “consistency”, and how the unpredictability of AI-infused systems makes it hard to establish consistency for the users of the system/interface (Amershi et al., 2019). To test the guidelines they involved a heuristic evaluation where the evaluators tested the guidelines to identify them in existing AI-infused systems and how well they could separate the guidelines apart when inspecting the different user interfaces. The other inquiry was a case study where the authors further investigated the

clarity of the guidelines and if they were understandable to use across different types of AI-infused systems. After analyzing the results they presented the current 18 design guidelines (Amershi et al, 2019).

Design guidelines and Youtube

I chose these two guidelines:

G9: Support efficient correction and

G13: Learn from user behavior.

I picked the two different guidelines because, based on own experiences, I think that Youtube support efficient correction in a pretty good way, and that they learn from user behavior to a certain extent.

While Youtube have an administrator page where the user can change what search history Youtube could use, I found this page unnecessarily difficult to navigate to. And while Youtube give the user the option to either give “I am not interested” or “Don’t show me this channel” to specific videos, I think they could find a more efficient way to approach the issue. Youtube does show an interest as to why the video wasn’t interesting, but the feedback options is rather limited – either “I have already seen this video” or “I don’t like this video”. Whether this is kept simple by purpose isn’t clear to me, but by keeping it this simple, it makes it more approachable to novice and children users. To make a bigger impact on the content recommended you have to go to the administrator page – which I am sure most users don’t do.

From the administrator page on what content Youtube use, it also becomes apparent that Youtube use all activity on your Google account as well when finding content to recommend. Personally I feel like this motion is more apparent when selection advertisements, and not the actual videos they recommend.

When looking at G13, I think that this guideline kind of represent what the recommendation-engine is meant to do, and their main purpose. The AI-infused system is meant to learn from its user to customize the content of the webservice. To better learn from the user, maybe Youtube could also become better to correct mistakes or “regrets” made by the users. If I as a user state that I am not interested in more DIY-videos to pop up in my main feed, this behavior should be picked up as well as the videos I want to have shown to me.

Bender and problematic aspects

In the introduction Bender et al. present the question “*What are possible risks associated with technology and what paths are available for mitigating those risks?*” (Bender et al., 2021, p. 610). The arguments that the authors pose are within the scope of large language models (LM). A general argument is surrounding how LMs don’t have the ability to understand the meaning behind words as of yet, and that while this is the case, a large effort must be made to secure a secure and unbiased LM. Also not to mistake the output from technologies as meaningful replies (Bender et al., 2021).

When arguing that environmental and financial costs are one of the problems associated with large LM, the authors point out that in many cases, the ones most likely hurt by the environmental consequences, are also least likely to benefit from LMs. Also, much of the power and energy to uphold the processing of large LMs aren’t necessarily generated through renewable energy sources. Here the authors find it important to document the energy use to use as a trade-off, and then see the full effect of the LM (Bender et al., 2021).

It is also relevant to look at the limitations to large LMs - and that in certain contexts, it will not be a large benefit of having a huge dataset. In many instances large datasets don’t stay without bias – and most can’t, like the authors point out, guarantee diversity because of bias. A difficulty is how to filtrate the data without censoring or leaving out significant data because of offensive meanings – as words also can have more meanings than one. The same goes for happenings that aren’t fully covered in media or through data – which can lead to one-sided and biased information (Bender et al., 2021).

Solutions or recommendations Bender et al. pose are mostly connected to inspecting environmental and financial costs, and make bigger investments to properly and securely documenting the large datasets (and avoid document debt) (Bender et al., 2021).

2.3 Chatbots/ Conversational user interfaces

Challenges

One of the key challenges of designing chatbots or conversational user interfaces, like discussed in the last lecture, is because there is a different focus than other systems. Rather than focusing on designing a usable layout of the system, chatbot design focus more on the conversation between a user and the chatbot. The reason for this is because a chatbot often has a specific purpose, like a health assistant (Helsevenn) or a helper with navigating large companies or municipality pages (KommuneKari). The image of the chatbot here would be necessary, as Helsevenn for instance is meant to reach out to pupils in high school about

difficult subjects. The flow of dialog/information and the choice of words are ,because of this, more important than having a fancy background in the chatbot.

Another issue brought up in the lecture was the presentation of the chatbot to the user. The choice in what the chatbot says or the way it presents itself to the user is also a challenge when designing chatbots.

A last challenge I will mention is the need to change focus from user interface design to a user experience in the sense of service design. Instead of focusing on a single/individual interactions with the interface, the focus needs to broaden to end-to-end experiences.

Chatbots and guidelines

The guidelines in question from Amershi et al.(2019):

Guideline 1: Make clear what the system can do

Guideline 2: Make clear how well the system can do what it can do.

Like mentioned above, one of the challenges when designing chatbots lay with the question of how to present the chatbot to the user. Both guideline 1 and guideline 2 could help the designer here. To make it clear what the system can do, and make clear how well the system can do this is very relevant to shaping the expectations of the user and their experience with the chatbot and dialog. The problem with this, however, is to decide what to include - what would be useful for the user to know beforehand, what will be too technical and complex? How do the designer cover the guidelines without creating an overwhelming wall of text for the user to meet when initiating contact with the chatbot?

Inspired by Kocielnik et al. and their techniques of manipulating expectations I will very shortly explore the question: Could there be regulation of expectations in other ways than the chatbot itself explaining its purpose? One could also use a media-file such as a video that goes more into depth in how the chatbot works. While a video is more time consuming, and could be excluding for some in different contexts, it is a simple way to creatively make the user up to date with chatbots and their functionality and limitations. This also prevents the wall of text you possibly would need to cover the functionality of the chatbot, and still adherence to the guidelines of both what and how surrounding the chatbot.

Appendix 1 – Changes done based on feedback

From the feedback I have done some changes, including changing my definition of “robot” as it was a bit confusing before - and added a little more information about the origin of the AI-field.

References

- Amershi, S., Weld, D., Vorvoreanu, M., Fournery, 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)*.
- 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).
- Bratteteig, T & G, Verne (2018). Does AI make PD obsolete? Exploring challenges from Artificial Intelligence to Participatory Design. *Participatory Design Conference*. Volume 2, p. 1-5
- Britannica encyclopedia (2021) *Robot-technology*. Britannica.no. Available at: <https://www.britannica.com/technology/robot-technology> (Accessed 6.september.2021)
- Grammarly (2021). *About Grammarly*. Grammarly.com. Available at <https://www.grammarly.com/about> (Accessed 7.september.2021)
- Grudin, J (2009). AI and HCI: Two Fields Divided by a Common Focus. *AI magazine*, 30 (4), p. 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).

McCarthy, J. (1998). What is Artificial Intelligence? Available at:

<http://cogprints.org/412/index.html> (Accessed 6. September 2021)

Merriam Webster Dictionary (2021) *Artificial intelligence*. Merriam-webster.com. Available at <https://www.merriam-webster.com/dictionary/artificial%20intelligence> (Accessed 6.september.2021)

Merriam Webster Dictionary (2021). *Robot*. Merriam-webster.com. Available at https://www.merriam-webster.com/dictionary/robots?utm_campaign=sd&utm_medium=serp&utm_source=jsnld#examples (Accessed 6.september.2021)

Nielsen, J. (1994). Enhancing the Explanatory Power of Usability Heuristics
<https://dl.acm.org/doi/pdf/10.1145/191666.191729> (Accessed 7.september.2021)

Thrun, S (2004). Towards a Framework for Human-Robot Interaction

HUMAN-COMPUTER INTERACTION, Volume 19, p. 9–24

UUtilsynet (2021) *Kvifor universell utforming av ikt?* Uutilsynet.no. Available at <https://www.uutilsynet.no/veiledning/kvifor-universell-utforming-av-ikt/240> (Accessed 6.september.2021)

Verne, G (2020) Adapting to a Robot: Adapting gardening and the garden to fit a robot lawn mower. *HRI '20 Companion*. Available at: <https://dl.acm.org/doi/pdf/10.1145/3371382.3380738> (Accessed 7.september.2021)

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).