IN5480 - Individual assignment fall 2021

Module 1

Appendix

After handing in the first iteration, I received some feedback on how to improve task 1.3.4 on what it means to understand. I've now taken the feedback into account, and tried to articulate a little more precisely how I understand the difference between 'understand' and 'understanding'.

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

1.1.1 A brief history of AI

The term "artificial intelligence" was officially coined by John McCarthy, an american mathematician and computer scientist at Stanford university in 1956. However, British mathematician and leading code breaker Alan Turing speculated in why computers shouldn't be able to enter fields that were originally designed for humans. (Gruding 2009, p. 49). On the same note, in his article "Computing Machinery and Intelligence" from 1950, Turing presents methods for testing a machine's intelligence, famously known as the *Turing test*. It would not surprise me if the thought of intelligent machines have existed as long as machines themselves.

1.1.2 Three definitions of AI

1st definition

John McCarthy (2004), researcher at Stanford University, defines artificial intelligence as "*the science and engineering of making intelligent machines, especially intelligent computer programs*". He further specifies that "*...but AI does not have to confine itself to methods that are biologically observable.*" McCarthy's definition clearly puts that artificial intelligence is first and foremost an engineering feat; an invention created by humans in which potential

we are yet to fully understand. It's clear that McCarthy seeks to distinguish learning in humans and animals from machine learning..

2nd definition

Bratteteig and Verne (2018), researchers at Institute of informatics at UiO define AI as "*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.*" In this rather modern depiction of AI, there is a heavy emphasis on human-like behavior within computer systems, and how they should be able to perform task that 'ordinary' machines wouldn't be able to do. I think these are tough requirements for any AI, but I think the description fits pretty good with my own imagination of AI.

3rd definition

Wikipedia, the free encyclopedia, defines artificial intelligence as "*intelligence demonstrated by machines, as opposed to the natural intelligence displayed by humans or animals.*" (Wikipedia, 2021). The article was written by an anonymous contributor. In this definition, there is a clear boundary between natural and artificial intelligence, in the way that it is displayed by different entities. This definition lacks any further description of what actually separates different types of intelligence, but it's clear that machines are not able to display the same kind of intelligence as humans or animals.

My definition

I would define AI as: "a computer system that is able to improve its performance and adapt to changes in its environment through machine learning processes." I think it's important to specify what separates AI from traditional software, and I believe this lies in its dynamic nature, compared to 'normal', static algorithms. Improving its performance (e.g. success rate, speed, or effectiveness) through 'learning' is something that I find of key importance for a computer system to be named artificial intelligence.

1.1.4 Brief review of an article

Grudin's article "*AI and HCI: Two fields divided by a Common Focus*" takes on the two research fields of AI and HCI, and how they share a lot of common ground, both through history and in their focus on improving existing and future algorithms. The main part of the article discusses how the different fields have evolved since the 1950s. Finally, the author shares a piece on how technology trends point in the right direction, that AI and HCI research benefit one another, and that the potential for usable AI and "*synergy between the fields are outstanding*" (Grudin, 2009, p. 55).

To the article, I would like to ask the author how he imagines a hybrid field between AI and HCI could actually look like?

1.1.5 Contemporary company that works with AI

IBM presents their AI-package for business called *Watson* with a focus on making a business "smarter", stating that their solutions are "*designed to reduce the costs and hurdles of AI adoption while optimizing outcomes and responsible use of AI*." There is a clear focus on the AI being *responsible*, stating the fact that the processes of AI-led decisions are "transparent".

1.1.6 Human interaction with AI in film

In Alex Garland's film *Ex Machina* (2014), young programmer Caleb is selected to participate in an experiment with "synthetic" intelligence by evaluating the human qualities of a highly advanced humanoid AI (IMDB). A big portion of the film consists of Caleb interacting face-to-face with the humanoid AI, Ava. Interaction is shown through human-like conversations, including switching eye contact, emotions, learning and joking around. The AI is super-convincing, and as Caleb grows fond of the humanoid, he is fooled into helping her escape, while she leaves him to die.

1.2 Robots and AI systems

1.2.1 Robot origin

The word “Robot” was introduced by Karel Čapek in 1920, a Czech writer, in his play *Rossum’s Universal Robots.* The word itself stems derived from the Czech word “robota”, which loosely translates to “forced labour” (Love, 2020).

1.2.2 Two definitions of “robot”

1st definition

In 1993, Merriam Webster’s collegiate dictionary defined robot as “*an automatic device that performs functions normally ascribed to humans or a machine in the form of a human.*” (Struhn, 2004, p. 11). This definition implies that a robot is *either* an automatic device that performs a task originally performed by a human *or* that it’s a machine which takes the shape of a human being, which perhaps hints to a divide in robot discussions at the time?

2nd definition

In 1979, the Robot Institute of America defined a robot as “*a reprogrammable, multifunctional manipulator designed to move materials, parts, tools, or specialized devices through various programmed motions for the performance of a variety of tasks.*” (Struhn, 2004, p.11). Although quite a lot more aged definition, this one is more focused on a robot’s traits and which specific tasks they are designed to solve. Reprogrammability and multiple functions are defined almost as requirements.

My definition

I would define a robot as “a sensory machine that performs pre-programmed tasks and responds to user input.” I think importance aspects of a robot, is that it can be programmed, but at the same time takes feedback from users. I also think it’s essential for a robot to be able to use sensors, in order to ‘see’ or navigate through its environment.

1.2.3 The relation between AI and robots

I believe that a robot can utilize AI in order to perform more generalized or complicated tasks, as AI's don't necessarily need a physical manifestation of a robot. At the same time, according to my own definitions, a robot is dependent on some sensory input, and how it's programmed to interact with its environment might be dependent on some AI under the hood in order to function as intended.

1.2.4 Description of a robot

Robotic lawnmower's have perhaps become one of the most adopted piece of robotics to enter the global market. Their entire functioning consists of autonomous actions, from driving, cutting, dodging obstacles parking in the charging port. It requires next to no human interaction, other than requiring a preset area for cutting.

1.3 Universal Design and AI systems

1.3.1 Definition of Universal Design

Uutilsynet.no (2021) defines universal design as "*the thought that services should be accessible to all, regardless of age, functional ability and level of education.*" I believe the most important aspect of universal design is accessibility, whether it being a service, a product or a building: All people should be able to use it, equally.

1.3.2 The potential of AI with respect to human reception

Self-driving vehicles have amazing potential for people with reduced eyesight, or other physical handicaps who normally might have to resort to expensive, specially designed vehicles, or who might not even be able to drive at all. In digital education, AI-based speech-to-text systems might be a fantastic option for people with reduced hearing.

1.3.3 The potential of AI for both including and excluding people

In her TED Talk, Joy Buolamwini shares her personal experience with algorithmic bias. Globally used facial recognition software was not able to detect her face, because the people behind the algorithm didn't account for it to learn a broad range of skin tones. (Buolamwini,

2016). I think developers have a responsibility to make sure that their algorithm is trained properly, as it can potentially cause unintended exclusion.

1.3.4 'Understand' and 'understanding'

To me, understanding is being able to collect information, process, reflect on it, use it, and learn from it. However, I figure that to understand is a simpler process that doesn't necessarily involve any reflection or learning, and that it's simply about being able to collect, process and use information for a purpose.

1.3.5 Do machines 'understand'?

Whether machines are able to "understand" is a very complex question, but according to my own definition of 'understand', I'd say sure! Machines are undoubtedly able to process and store information, and sometimes even connect relevant information together. However, being able to reflect on the given information and connect it to a broader context in order to act, is something I think that machines aren't capable of doing, yet.

1.4 Guideline for Human-AI interaction

1.4.1 Microsoft guideline for Human-AI interaction

Guideline 13 "Learn from user behavior" is about "*personalizing the user's experience by learning from their actions over time.*" Using for example an AI-based web browser, it would be useful if the user was provided relevant web pages based on previous browsing.

1.4.2 HCI design guidelines vs. Microsoft guidelines

Based on Shneiderman's eight golden rules, the HCI design guidelines and Microsoft guidelines for Human-AI interaction share many similarities, especially when it comes to system feedback, ease of use and user efficiency. I do, however, notice that the Human-AI guidelines are more focused on relating to the user's context and social norms, as well as gradually adapting to the user, rather than being designed purely based on a set of static rules.

Module 2

2.1 Characteristics of AI-infused systems

2.1.1 Identify and describe three key characteristics of AI-infused systems.

According to our lectures on design of interaction with AI, the shared characteristics between AI-infused systems is that they are learning, improving, that they involve some kind of 'black box' logic, and that they are driven by big data sets.

AI-infused systems can be describes as *learning* as they over time will recognize use patterns, and dynamically change its look, feel, content, and so on, based on how the user interacts with the system.

They can also be described as *improving*, an attribute that builds on the learning aspect. As the system is learning, it should also improve its service, accuracy or functionality over time. However, it's often not so easy to understand how an AI-infused system learns (Yang et al. 2020, p. 2).

'Black box' implies that the system is driven by massive data sets and changing algorithms which makes it difficult to actually understand what is going on 'under the hood' and what causes the system's behaviour. This can cause problems not only for users, but also for designers and implementors of AI-infused systems, as it can be hard to predict and measure how the system will behave.

2.1.2 Identify one AI-infused system and exemplify some of the above key characteristics

For this task, I've decided to go through with Spotify and its recommendation algorithms, in terms of its playlist generator 'Discover Weekly' and the playlist 'radio' function which automatically keeps playing music that matches the current listening session.

Spotify's recommendation algorithms can be described as *learning*, as they keep track of your listening habits by collecting data about the genres, artists, moods and albums you're

listening through during the day. It uses this data to create a fresh playlist every week with suggested songs that *should* match your taste.

The algorithms can probably also be defined as improving, though I'm not able to verify this claim as of right now. However, Spotify's feeding huge amounts of data into these algorithms, and I would assume that your listening pattern is constantly compared to their huge musical library, so that if you were to e.g. switch from listening to hip-hop to country over the course of a week or two, I'm sure your 'Discover Weekly' playlist would change as well.

The black box concept most certainly applies to Spotify, in the sense that these different AI-infused systems are based on massive amounts of data, but exactly how this data is used, in which manner and what it actually affects is uncertain.

2.2 Human-AI interaction design

2.2.1 Main take-aways from the two articles

Kocielnik et al. (2019) explores how expectation adjustment can affect users impressions of AI systems. Through a set of experiments on a scheduling assistant, it was found that the users' expectations for a system impacts how well-functioning they perceive it to be, and that if users are informed early on what to expect from the system, the user experience was improved. It is therefore advantageous for both users and designers/developers to explicitly describe what a system can (and cannot) do in order for users to accept an imperfect AI-system.

Amershi et al. (2019) looked at over 150 AI-related design recommendations and generated 18 generally applicable design guidelines for human-AI interaction (2019, p. 12). As AI technologies have become more mainstream and are often integrated into both existing and brand-new systems, the usability and user experience of such systems has not necessarily become any better. Users are unsure about what to expect from AI-infused

systems, and the guidelines will hopefully enable designers and developers to create systems that are better and more user friendly.

2.2.2 How the two design guidelines could inspire improvements

Amershi et al. (2019) identifies 18 different AI Design Guidelines, where two of them is defined as:

- G11: Make clear why the system did what it did
- G13: Learn from user behaviour

The eleventh guideline is about enabling opportunities for the user to get an explanation of *why* the system behaved as it did, especially when the system is in the wrong. The thirteenth guideline describes how an AI-infused system should learn from user behavior over time, by personalizing the experience.

According to Whitehouse (2021), Spotify uses AI in the shape of collaborative modelling, natural language processing and convolutional neural networks. As far as I'm concerned, Spotify never explicitly tells the user which parts of their system is AI-infused, which fails to follow the first guideline. However, according to the article, most users will experience AI interaction through "Discovery Weekly" playlists, which recommends music that reflects the user's preferences based on artists, albums and listening patterns. If the Discover Weekly fails to match the user's preferences, there's never an explanation from the system about why it made the choices it did, which can confuse the user. Based on my own experience, the generated playlists actually does a pretty good job at recognizing my listening patterns, and it will often recommend artists I've never heard that matches my preferences surprisingly well. Still, by following the thirteenth guideline, the system could be improved in the instances where it doesn't do that good of a job.

2.2.3 Summarized argument

In Bender et al. (2021) the researchers question whether ever larger language models are necessary compared to the associated costs and risks (2021, p. 619). The authors have examined costs surrounding environment, economy and opportunities, and encourage NLP

researchers to tread carefully when exploring huge language models, that could potentially require massive amounts of energy while producing unethical results – especially when working with 'human-like' applications. Feeding such language models with huge quantities of data pose the risk of creating an application that can spread misinformation, political propaganda or promote hateful content when prompted with the right questions. These types of systems can have a dangerous impact on individuals and social groups, if not society as a whole. It might therefore be a better idea to avoid, or at least carefully consider the risks up against the potential gains from working on such language models.

2.3 Chatbots / conversational user interfaces

2.3.1 Key challenges in design of chatbots

In a general sense, based on our lecture on interaction with AI, designing chatbots is about designing conversations as well as understanding the chatbot's role in a complete (and sometimes complicated) user journey. As a designer, one cannot design a chatbot framework while skipping out on the actual conversational content, as this is what actually describes the chatbot's look-and-feel and lays a lot of the foundation for the user experience.

On another note, when designing a chatbot, one should consider the risk of a mismatch between user expectations and the chatbots actual area of use. Kocielnik et al. (2019) and Amaersi et al. (2019) both discuss how expectations toward an AI-infused system will influence the overall user experience, and I can't see how this doesn't apply for chatbots as well. If it's, for example, not clear what a given chatbot can help the user with, what types of sentences it responds best to, or if it's enable to 'remember' earlier messages, what can the user actually expect from the interaction? Will it only cause frustration? Or can the chatbot provide meaningful assistance without the need for a human colleague to step in?

2.3.2 Resolving some challenges with chatbots / conversational UI

In Amershis et al. (2019), guidelines G1 and G2 are written as follows:

- G1: Make clear what the system can do.

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

Considering the key challenges I've highlighted in the previous task, both of these guidelines could improve or even resolve the addressed problems. Making clear what a system can do, sounds, to me, as a great way of introducing a chatbot to a (new) user. As most chatbots are designed for a particular purpose, it will probably be helpful for users to understand the scope and limitations of the system they are interacting with. A chatbot at e.g. Skatteetaten can probably help users navigate to their tax return, but it most likely can't help anyone figure out the closest bus stop. However, to (perhaps especially inexperienced) users, this might not be obvious at all, which is why I think explicitly stating the capabilities of the chatbot will greatly improve the user experience.

I think the guideline of making clear how well a system can do what it can do is a bit more of an abstract concept, that might not be feasible for *all* chatbots. However, I think that designing chatbots that informs the user when it gives an answer it's unsure about, can improve the user experience somewhat. Rather than not giving an answer at all or being confident in an answer it's uncertain of, I think designing a system that somehow is able to visualize its 'certainty level' can create more trust in the system over time. It might even feel more 'human like' when a machine conveys some degree of uncertainty, rather than giving hard yes-or-no answers.

References

Module 1

Bratteteig T., Verne, G. (2018). Does AI make PD obsolete? Exploring challenges from Artificial Intelligence to Participatory Design. *PDC'18, August 20-24, 2018, Hasselt and Genk, Belgium*.

Buolamwini, J. (2016). How I'm fighting bias in algorithms. *TEDxBeaconStreet.* Created November 16th 2016. Watched September 7th 2021. https://www.ted.com/talks/joy_buolamwini_how_i_m_fighting_bias_in_algorithms

Chen, J. (2020). Neural Network. Investopedia. Read September 7th, 2021: *https://www.investopedia.com/terms/n/neuralnetwork.asp*

Grudin, J. (2009). AI and HCI: Two Fields Divided by a Common Focus. *AI magazine 30, no 4.*

IBM. Read September 7th 2021. *https://www.ibm.com/watson*

IMDB (2021). Ex Machina. *https://www.imdb.com/title/tt0470752/?ref_=nv_sr_srsg_0*

Love, D. (2020). Where does the word 'robot' come from? Published May 3rd 2020. Read September 7th 2021. *https://www.sciencefocus.com/future-technology/where-does-the-word-robot-come-from/*

McCarthy, J. (2004). What is artificial intelligence? *Stanford University.*

Schulz, T., Herstad, J., Torresen, J., (2018). Classifying Human and Robot Movement at Home and Implementing Robot Movement Using the Slow In, Slow Out Animation Principle. *International Journal on Advances in Intelligent Systems, vol 11 no 3 & 4, year 2018, http://www.iariajournals.org/intelligent_systems/*

Struhn, (2004). Toward a Framework for Human–Robot Interaction. *HUMAN-COMPUTER INTERACTION, 2004, Volume 19, pp. 9–24, Lawrence Erlbaum Associates, Inc.*

Uutilsynet. Kvifor universell utforming av ikt? Read September 7th 2021. https://www.uutilsynet.no/veiledning/kvifor-universell-utforming-av-ikt/240

Wikipedia (2021). Artificial intelligence. *Downloaded September 7th 2021 from:* https://en.wikipedia.org/wiki/Artificial_intelligence

Module 2

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P. & Teevan, J. (2019). *Guidelines for human-AI interaction*. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (paper no. 3). ACM. https://www.microsoft.com/en-us/research/uploads/prod/2019/01/Guidelines-for-Human-AI-Interaction-camera-ready.pdf

Bender, E. M., Gebru, T., McMillan-Major, A., & Mitchell, M. (2021). *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?.* In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 610-623). ACM. https://dl.acm.org/doi/pdf/10.1145/3442188.3445922

Kocielnik, R., Amershi, S., & Bennett, P. N. (2019). *Will You Accept an Imperfect AI?: Exploring Designs for Adjusting End-user Expectations of AI Systems.* In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (paper no. 411). ACM. (https://www.microsoft.com/en-us/research/uploads/prod/2019/01/chi19_kocielnik_et_al.pdf)

Whitehouse, K. (2021). *How Spotify Uses Artificial Intelligence, Big Data, and Machine Learning.* Data Science Central. https://www.datasciencecentral.com/profiles/blogs/6448529:BlogPost:1041799

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

Følstad, A., & Brandtzæg, P. B. (2017). *Chatbots and the new world of HCI.* interactions, 24(4), 38-42. https://dl.acm.org/citation.cfm?id=3085558

Luger, E., & Sellen, A. (2016). *Like having a really bad PA: the gulf between user expectation and experience of conversational agents*. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (pp. 5286-5297). ACM. https://www.microsoft.com/en-us/research/wp-content/uploads/2016/08/p5286-luger.pdf