3rd iteration of individual assignment - IN5480

Module 1, 2 and 3 - Saralok

Table of Contents

1.1 Artificial Intelligence

The history of AI

The history of artificial intelligence (AI) began with World War II, with the English mathematician Alan Turing in the front lines. Turing was hired by the British government to help encode the strategic messages sent by the Germans to their allies. This was a very difficult task, as there could be millions of different combinations to the solution and the code changed every 24 hours. To boost the effectiveness of this, Turing created a machine that would try all possible combinations automatically. After the war, Turing published the seminal paper "Computing Machinery and Intelligence" which introduces us to the famous "Turing Test". His research formed the field of computer science as we know it today.

Although Turing was the first to introduce this concept, the actual term "artificial intelligence" was first used by an American computer scientist named John McCarthy in 1956 in a proposal he wrote for the famous Dartmouth conference. This conference was where AI was first stated as a field (Grudin, 2009).

Definitions of AI

As it was John McCarty who first used the term, it is only natural to begin with his definition of AI:

"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)

McCarthy's definition is focused on explaining AI as a field of science, more than an actual computer system. Another definition worth mentioning is from the English Oxford Living Dictionary:

*"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" (*Lexico 2020)

This Oxford definition defines AI as a way to adapt human intelligence into an automated system, giving it human-like abilities. What is interesting about this definition is that it is a bit more modern than what the previous was. It is more theoretical and takes modern technology into account. AI is still a field in great development and is in many ways still being defined.

The Britannica Dictionary had another similar, yet different definition:

"*Artificial intelligence (AI), the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings."* (Copeland, 2020, section 1)

My definition of AI:

"*Automated systems created to simulate human behavior, used to simplify and bring efficiency to tasks usually conducted by humans."*

My definition of AI comes with the reasoning that AI-systems do not have emotional behavior and human reasoning. They simulate these behaviors, but they are all pre- learned patterns or a result of pattern recognition or learning.

AI in business (GHOST)

Ghost is one of the top 50 businesses on Forbes Most Promising Artificial Intelligence Companies, and they have a goal of delivering perfectly safe, self-driving cars, by the end of 2020. They advertise AI as a product with a software you can install in the car you already own, and market this by saying you do not have to pay attention when driving, as the computer will do all the thinking for you.

"Real self-driving means you can fully turn your attention elsewhere and leave control of your car, to a computer." (Ghost, 2020)

In addition to this, they say their product is safer than driving yourself, as the computer is better at expecting unpredictable situations and will avoid other dangerous drivers. *(Ghost, 2020)*

AI in entertainment (HER)

Artificial intelligence has always been an exciting subject and is regularly portrayed in literature, tv-series and movies. It usually plays with the futuristic aspect of society, lets us predict the future both rationally and irrationally, and gives us an idea of what society might look like years from now.

One of the more prominent movies out there with a focus on artificial intelligence, is the American science-fiction movie *"Her"* from 2013. The plot takes place in the near future, and revolves around a newly divorced, lonely writer who earns his living by writing love letters for other people. Struggling with the split from his wife, the writer gets introduced to a new operating system portraying the traits of a human being. He installs this in his apartment and gives it the name of *"Samantha"*. Samantha is able to evolve and learn new things as time passes, and he eventually starts a relationship with *"Her"*.

In this movie, AI is portrayed as a way of filling the void of human contact. The AI-system works as a replacement for actual social interactions, and it is tailor-made to fit perfectly for the end user by adapting to his pattern and lifestyle. Throughout the movie you get to see the AI evolve and learn to a point where you barely can separate between it, and an actual human being.

1.2 Robots and AI systems

The history of the word "Robot"

The word "Robot" was first used by the Czech writer Karel Capek in 1921. He wrote a play called "Rossum's Universal Robots, and in these stories the robots were usually portrayed as servants for the humans. The word Robot comes from the Czech word *Robota*, which means

"laborer". In this play the robots eventually rebel against the humans, which leads to an extinction of the human race.

Definitions of "Robot"

The first definition of the word "Robot" is from The English Oxford Dictionary. They define it like this:

*"A machine resembling a human being and able to replicate certain human movements and functions automatically" (*Lexico 2020)

What is interesting in this definition from Oxford is that it has to function automatically for it to be a robot. It does not include the concept on remote-controlled devices. The second definition comes from Webster, an online dictionary who explains the term like this:

"An automatic device that performs functions normally ascribed to humans or a machine in the form of a human." (Merriam Webster, 2020)

Both of these definitions describe a robot as something physical that replicates both human behavior and form. It does not take into account that a robot might be formed like another biological creature, or not like a creature at all.

My definition of Robot:

"A machine that resembles or takes inspiration from biology in the way it can physically move parts of itself, and that can operate periodically or indefinitely without physical human interaction."

I have created my own definition of the word, based on the two definitions above. I made some alterations on the parts of the definitions I did not agree with above.

The relation between robots and AI

By taking the previous definitions in account, there is a clear difference between artificial intelligence and a robot. While AI is a system or a digital program, robots have more of a physical appearance. They are both able to help humans and fulfill certain human-like tasks, and we are able to interact with them in different ways. An AI is more of an intelligent system that is able to learn, adapt and analyze. It can be installed or programmed into a robot to give it these traits, but a robot without an AI system will in my opinion be less autonomous. It will run more on automatic movements and predefined features, rather than actual governing independently.

Human-robot interaction (PEPPER)

Pepper is a commercial robot created by Softbank Robotics, and is the first social- humanoid robot who is able to recognize faces and basic human emotions. He is currently used by over 2000 companies worldwide to welcome, inform and guide guests when they enter stores and buildings. He can communicate both through conversation (15 different languages) and through a touch-screen placed on his chest. The robot is programmed to mimic and imitate basic human movements, with both cameras and sonars around him for autonomous navigation.

Pepper is mostly used in retail to enhance the user experience when shopping by engaging in conversations with the customers, but also to gather comprehensive data and generate client insight. (Softbank Robotics, 2020)

1.3 Universal Design and AI systems

Definition of universal design

"Universal Design is the design and composition of an environment so that it can be accessed, understood and used to the greatest extent possible by all people regardless of their age, size, ability or disability." (Difi, 2017)

Universal Design is about ensuring that everyone regardless of disabilities are able to partake in society. It is about designing products and solutions in a way that doesn't exclude people from participating in activities, without the use of external aids. In Norway we take universal design very seriously, and it is mandatory by law to follow certain design principles (Lovdata, 2013). These principles are enforced by the Norwegian Directorate for Management and ICT.

The potential of AI

AI can be used to extend human capabilities in a lot of ways. One example is self- driving cars. An AI will be able to interpret situations and minimize the risk of accidents, since it often reacts much faster than what a human is capable of. It can also be of help for people with different disabilities. People with visual impairment might have problems reading but will be able to have access to information with an AI reading the text out loud for them.

AI and exclusion

Even though AI can be of great use when it comes to including people in our society, it can also touch upon exclusion. One of the groups who might fall behind with this technological progress, is the older generation. Technology can at times be too advanced for them to keep up on, as they are not used to learning technologies like our generation today. It is also important to have people with different disabilities in mind. For example, people who are mute will also have trouble using voice-recognition and voice-controlled devices, as people.

1.4 Guideline for Human-AI Interaction

Learn from user behavior

Guideline number 18 in the Microsoft Guidelines for Human-AI interaction is to" Learn from user behavior" (Microsoft, 2019). This example states that the AI should personalize the user's experience by learning from their actions over time. By tailoring the user's actions, the human-AI-interaction will be more efficient and fluent. An example of this could be a search-

engine who memorizes your last searches, or an auto-generated playlist with all your favorite songs. This guideline is similar to number seven of Jakob Nielsen's design heuristics:

"Accelerators — unseen by the novice user — may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions." (Nielsen, 1994)

HCI guidelines and Human-AI guidelines

Looking at Nielsen's Design Heuristics, there are several other guidelines who are comparable to the Microsoft Guidelines for Human-AI-Interaction.

They are both invested in being clear on what the system can do. It should be easy for the user to understand and make connections just by looking at them. This also applies to showing clear and relevant information to the user. Both of the guidelines include error prevention, and relevant feedback when an error occurs as well as. (Nielsen, 1994)

2.1 Characteristics of AI-infused systems

Key characteristics of AI-infused systems

An AI-infused system is described as a *"system that has features harnessing AI capabilities that are directly exposed to the end user"*. (Amershi et al., 2019).

During this module, we have learned that AI-infused systems are recognized by four different key characteristics: learning, improving, "blackbox" and that they are fueled by large data sets.

By **learning** we refer to the AI system being able to gather information over time. To be able to adapt and evolve with the data it is given. By **improving** we refer to using the data gathered to better itself. Through interactions the AI will learn what works, what does not work, and ways of responding properly to users.
The term **black box,** refers to viewing the AI-system as a black box. The Oxford Living Dictionary defines a "Black Box" as a complex system or device whose internal workings are

hidden or not readily understood *(*Lexico 2020). It is not obvious to the user, why the AI presents the data it does. There is a lot going on under the surface of what the user cannot see. This can be a challenge as it is hard to connect the input to the received output, and will in some ways make the interaction harder. When saying the AI system needs to be **fueled by large data sets,** we refer to the collection of data. Systems need a continuous flow of data to be able to learn, improve and adapt over time. This data is collected from the users in several different ways. Chatbot, as an example, will be able to adjust their output based on all previous user inputs. The more input, the more precise output.

Spotify as an AI-infused system

Spotify is the one of the largest, if not the largest service for streaming music today (Spotify, 2020). They use artificial intelligence in several different ways to enhance user experience and tailor the program to fit each user individually. For example, by collecting data through the user's search history pattern they are able to recommend songs and playlists to the user based on their preferences and previous song choices. The more data they are able to collect, the better will the recommendations be. The system is therefore both learning and improving over time.

2.2 Human-AI interaction design

Guidelines for Human-AI Interaction by Amershi et al. (2019)

The field of AI is evolving rapidly, and with this evolution comes new ways of integrating AI capabilities into user-facing systems. The research paper "Guidelines for Human-AI Interaction" is addressing these advances and are calling out the need for common guidelines for AI-Human Interaction, as they will generate both new challenges and opportunities for the HCI community. Throughout the last two decades, many guidelines have been made, but none which is covering all aspects of AI-Human Interaction. A team of Microsoft researchers have been working together to establish 18 guidelines that covers the field as much as possible. During the first phase of the research collected already developed guidelines, both from industry resources as well public articles and scholarly papers. They found 168 relevant guidelines, which were further clustered into 20 different concepts. In the next phase, they

conducted a heuristic evaluation on these concepts using them on already existing AI-solutions, before conducting a user study where 49 practitioners all, with HCI background gave feedback on the guidelines' clarity and applicability on a variety of different products (Amershi et. al., 2019).

Will you Accept an Imperfect AI? Exploring Designs for Adjusting End-user Expectations of AI Systems by Kocielnik et al. (2019)

"Inflated expectations about usability and ease of use have been shown to decrease user satisfaction and willingness to use products when those expectations are not met". (Kocielnik at al, 2019)

Kocielnik at al. is about exploring the end-users expectations on the capability of different AI-systems. In the article, they try to form different expectation adjustment techniques that prepare the users for AI imperfections. At first, they end up with five different suggestions, where one was rejected, one was partially supported and three were accepted. They end up moving forward with these three. The first one is an accuracy indicator, where they specify the accuracy of the system to the user before he or she uses it. The second one is example-based explanations, while the last is performance control. A way for the user to adjust the performance of the system themselves.

Design guidelines - Spotify

G13 - Learn from user behavior

"Personalize the user's experience by learning from their actions over time"

Spotify is constantly collecting data from their users. As mentioned in the previous chapter, they use this information to create tailor made playlists and give song recommendations based on what the user has listened to previously. The recommended lists and songs constantly update as the user picks a new song, and it will always be up to date.

G15 - Encourage granular feedback.

"Enable the user to provide feedback indicating their preferences during regular interaction

with the AI system."

You do have some control of the recommendations given to you. For example, if you like a song, you can give it a heart. Spotify will then add it to your favorites and recommend more songs and albums by this specific artist or band. One setback is that sometimes you listen to songs that are outside of your regular taste. These songs will also be given equal emphasis in the algorithm, and there is currently no way of telling the AI system that you in fact weren't too happy with their recommendations, which would probably solve a lot.

2.3 Chatbots/conversational user interfaces

Key challenges in the design of chatbots

Følstad mentions a few different key challenges when it comes to the design of chatbots. Two of these are "Conversation as design object" and "Necessary to move from UI design to service design",

Conversation as design object refers to the difference of designing a chatbot, relative to other user interfaces. Usually when you design, you have a graphical interface made up from different design elements. When it comes to chatbots, everything is more hidden. You have to design by interpretation instead of visualization.

Necessary to move from UI design to service design refers to the change of designing for a specific goal, to focusing on the interactive systems. Følstad and Brandtzæg says that "we need to move from seeing design as an explanatory task to an interpretational task". I would say this means to stop focusing on what the user can or cannot understand but shift the focus on understanding the users and their needs.

Key challenges with guidelines G1 and G2 in Amershi et al. (2019)

Guideline G1 and G2 in Amershi et al. (2019) states: "Make clear what the system can do" and "Make clear how well the system can do what it can do"

One problem that can be minimalized by following these guidelines is the users expectations of the AI systems. Unrealistic expectations will lead to frustration and dissatisfaction. It is

therefor important to clarify what the system can or cannot do before to actual interaction. This will also save the user a lot of time by not having to get to know/understand the system beforehand. When the user expectations do not live up to reality, people will easily give up and abandon the system. However, if they are aware of the limitations beforehand, they are more acceptable of it.

3.1 Human AI collaboration

With the increased technology of robots, we are now starting to view them more as interactive companions rather than a programmable machine.

In this assignment I will look at two examples of human-animal robots from Philips at al. (2016), describe their level of autonomy in light of Sheneidermans' research (Sheniderman, 2020) and reflect on their advantages/disadvantages if we were to change their autonomy-level. I will also reflect on the robots' current and needed explainability with the help of Hagras (Hargras, 2018) article.

What are levels of autonomy?

Jason Walker, co-funder describes an autonomous robot as a robot that can perceive its environment, make decisions based on what it perceives and/or has been programmed to recognize and then actuate a movement or manipulation within that environment (Waypoint Robotics, 2020). In Shcneidermans article, he presents 10 different levels of automation made by Sheridan and Verplank, to help create better, more trustworthy computer applications. These levels are ranked from 1 to 10, where level 1 is dependent on the human making all the decisions, while level 10 requires no human interaction or input at all. The goal of this system is to end up with solutions that have both high levels of human control **and** high levels of automation (Sheniderman, 2020).

Paro, the therapeutic robot

The first human-animal robot I want to look at is the robot Paro. Paro is a therapeutic companion created to resemble a baby seal and was created as a replacement for real-life animals in places such as hospitals and care facilities. Its purpose is to relieve depression and

anxiety, but it can also be used to prevent loneliness. Paro recognizes patterns in the users' behavior and actions. It learns and adapts from these and will eventually become quite personalized.

Considering Paros functionality, I would place it as a level 7 on the Sheridan and Verplanks autonomy scale. "The computer executes automatically, then if necessarily informs the human" (Shneiderman, 2020). The robot is able to make its own decisions but needs input by the human end-user to adapt and gradually learn/develop a unique personality suited for the owner.

Decreasing Paros level of autonomy could for example take away the robots ability to remember previous interactions, and it would respond in predefined patterns. With an action-reaction pattern like this, it will lose its distinctiveness, and would probably feel more like a toy rather than an interactive companion.

Increasing the level could perhaps result in the robot being unpredictable and therefore unfit for the environment it was created for. On a more positive note, giving it the ability to make even more intelligent decisions could also result in the user viewing it more as an independent companion rather than a robot.

Big Dog, the military robot

The second human-animal robot I want to look at is the military robot, Big Dog. Big Dog is, like the name states, a robot that resembles the physics of a dog. It was created by Boston Dynamic in 2004 (Boston Dynamics, 2020), with the main purpose of helping the military navigate through rough terrain using built in sensors. It also carries cargo in order to reduce soldier load.

Looking at BigDog in light of the Sheridan-Verplank levels, I would place it around 8. "The computer informs the human only if asked" (Shneiderman 2020) It has a high level of control of its actions, it is aware of its surroundings and adjusts after them in a very accurate way. Most of its actions are self-governing, but it does need input from humans from time to time.

Decreasing BigDogs level of automacy could result in the robot needing more supervision,

which would make it less of a resource. It would be an extra asset to keep track of during missions.

Increasing BigDogs level of automacy could result in being too independent. It would make its own decisions. One of BigDogs more important traits is that it has to be able to work together with the soldiers as a team. It would be unreliable if it were completely autonomous. On the other hand, with increased autonomy comes increased effectiveness. It could possibly be able to perform actions faster, when it doesn't need the approval of a human.

Explainability in human-animal robots

An explainable AI is an AI in which the actions can be easily understood and analyzed by humans (Hagras, 2020). For the AI to be explainable, it needs to be able to justify its action/behavior to the end-user. AI systems are often fueled with a lot of data, and without seeing the clear context, it can often be hard to understand why the AI does what it does. Hagras presents five different key focus areas for this: transparency, causality, bias, fairness and safety.

Looking at the first robot, Paro, in light of these key traits, it seems to be quite self-explanatory. As it both looks and sounds like a baby seal, users will expect that it acts like it as well. This fulfills the transparency part of explainability. Paro establishes trust by connecting with their users in long term sequences, and makes it clear what it can do from the very beginning. It does not show signs of any bias and treats all users equal. Paro also scores high on safety, as it can be somewhat predictable, and won't make any unforeseen actions. The consequences of Paro not being fully explainable is not that crucial, all though it might not be able to fulfill what it was initially made for.

When it comes to the second robot, BigDog, I have been trying to get some more information of its functions from extern sources, but none which helps me understand the abilities to a full extent. I feel however, that the tasks that this robot performs are pretty important, and the need for transparency is absolute when working as a team in military operations.

Appendix – Peer Review

Iteration 1:

The feedback I got from the Peer Review was very helpful. They used to "two stars and a wish" method where they said my writing was good, and they were happy with my choice of definitions for the different chapters.

For the wish, I was told to reflect more upon the relationship between AI and robots and explain my current reflections more in detail. I understand this wish, as I had completely misunderstood one of the terms, I used to explain the differences. I have made edits to this in the second iteration.

In addition to these changes, I've made a table of contents, sorted out some headlines and cleaned up a little bit.

Iteration 2:

The feedback I got from the Peer Review on iteration two, was that I needed to cite my sources better. It was not clear which articles I was talking about at all times. I completely agree with this. I have taken this into consideration when writing iteration 3.

Also updated the table of contents, as it was a bit off on the last one.

References

Grudin, J. *AI and HCI: Two Fields Divided by a Common Focus*. AI magazine 30, no 4 (September 18, 2009). https://aaai.org/ojs/index.php/aimagazine/article/view/2271

McCarthy, J. (1998). *What is Artificial Intelligence?* Available from http://cogprints.org/412/index.html (Retrieved: 09/09/2020).

Lexico (2020) *Artificial Intelligence.* Available from: https://www.lexico.com/definition/artificial_intelligence (Retrieved: 09/09/20).

Lexico (2020) *Robot.* Available from: https://www.lexico.com/definition/robot (Retrieved: 09/09/20).

B.J. Copeland (1998-2020) *Artificial Intelligence.* Available from: https://www.britannica.com/technology/artificial-intelligence (Retrieved: 09/09/2020).

Ghost (2020) *About ut.* Available from: https://driveghost.com/about/ (Retrieved 09/09/20)

Merriam Webster (2020 *Robot* Available from https://www.merriam-webster.com/dictionary/robot
(Retrieved 09/09/20)

Softbank Robotics (2020) *Pepper.* Available from: https://www.softbankrobotics.com/emea/en/pepper (Retrieved 09/09/09).

Difi (2017) Kva er universell utforming? Available from: https://uu.difi.no/kva-er-universell-utforming (Retrieved 09/09/020)

Forskrift om universell utforming av IKT-løsninger (2013) *Forskrift om universell utforming av informasjons- og kommunikasjonsteknologiske (IKT)-løsninger* Available at: https://lovdata.no/dokument/SF/forskrift/2013-06-21-732 (Retrieved: 09/09/20)

J. Nielsen (1994) *10 Usability Heuristics for User Interface Design.* Available at:https://www.nngroup.com/articles/ten-usability-heuristics/ (Retrived: 09/09/20)

Kocielnik, R., Amershi, S., & Bennett, P. N. (2019). *Will You Accept an Imperfect AI?: Exploring Designs for Adjusting End-user Expectations of AI Systems.* Available at: https://www.microsoft.com/en-us/research/uploads/prod/2019/01/chi19_kocielnik_et_al. pdf) (Retrieved 20/10/20)

Følstad, A., & Brandtzæg, P. B. (2017). *Chatbots and the new world of HCI. interactions.* Available at: https://dl.acm.org/citation.cfm?id=3085558(Retrieved 20/10/20)

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*. Available at (https://www.microsoft.com/en-us/research/uploads/prod/2019/01/Guidelines-for-Human-AI-Interaction-camera-ready.pdf (Retrieved 20/10/20)

Spotify (2020) *About us* Available at: https://www.spotify.com (Retrieved 20/10/20)

Phillips, E., Ososky, S., Swigert, B. and Jentsch, F (2016) *Human-animal teams as an analog for future human-robot teams, Proceedings of the Human Factors and Ergonomics Society* Avaialable at: https://doi.org/10.1177/1071181312561309 (Retrieved 12/11/20)

Shneiderman, B., *Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy* (2020). Available at https://arxiv.org/abs/2002.04087v1 (Retrieved 12/11/20)

Hagras, H., *Toward Human-Understandable, Explainable AI, Computer*. Available at: Ahttps://ieeexplore.ieee.org/document/8481251 (Retrieved 12/11/20)

Walker, Jason (2020) *What are autonomous robots?* Available at: https://waypointrobotics.com/blog/what-autonomous-robots/ (Retrieved 12/11/20)

Parobots (2020) *PARO Therapeutic Robot* Available at: http://www.parorobots.com (Retrieved 12/11/20)