Iteration 3 - AI

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

The history of how AI came about begins during World War II, where the potential importance of computers was established through its role in code breaking. A leading code breaker at the time, Alan Turing, wrote in the London Times that “I do not see why [the computer] should not enter any one of the fields normally covered by human intellect, and eventually compete on equal terms” (Grudin, 2009). The term “artificial intelligence” (AI) was first used in 1956, in a call for participation in a workshop, written by John McCarty. AI evolved over the decades, and eventually took a turn described by McCarthy who wrote that “[the goal] was to get away from studying human behavior and consider the computer as a tool for solving certain classes of problems. Thus AI was created as a branch of computer science and not as a branch of psychology.” The most recent turning point in the interest of AI came in the year of 1997, where a machine defeated the world chess champion. Further on, events such as launching remotely controlled robots on Mars, the availability of the internet and recommendation systems within it, as well as reduced costs for storage, processing and access, lead to the big interest we have of AI in today's society.

Definition 1: “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.” (Verne & Bratteteig, 2018).

Verne is a Senior lecturer while Bratteteig is a professor, both at the University of Oslo.

Definition 2: "Artificial Intelligence (AI) refers to the simulation of human intelligence in machines that are programmed to think like humans and mimic their actions. The term may also be applied to any machine that exhibits traits associated with a human mind such as learning and problem-solving." (Investopedia, 2020)

Investopedia is an American website based in New York City that provides investing and finance education.

Definition 3: "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." (Britannica)

Britannica is a general knowledge English-language online encyclopaedia

The definition by Bratteteig and Verne talks about AI as a field of computer science, while Investopedia defines AI as a simulation, and lastly Britannica writes both about the theory and the development of the systems. These three definitions do not contradict each other, but they focus on three different aspects of what AI is about. The definitions also vary in age, as Bratteteig and Verne's paper is written in 2018, the Britannica definition is from 1998 or earlier (it is difficult to pinpoint when it was actually written as is existed in book form previously , but it was uploaded online in that year), and the Investopedia definition unfortunately has no written date.

My definition of AI would be:
"AI is a machine performing tasks that requires human intelligence, as well as being able to mimic human cognitive processes". This definition is inspired by the other definitions above, but more focused on the aspects I find important with AI. To me, AI is not a theory or a field, but the actual system performing the tasks and how that system is perceived.

The company Computas is a Norwegian provider of IT solutions and consulting services in technological innovation. They talk about AI as a theory and a development process, writing on their website that “Artificial intelligence is the theory and development of computer systems capable of performing tasks that require human intelligence” (Computas), very similar to Britannica's definition. They also explain it as taking what is seen as human characteristics, and transferring them to a machine.

The movie Her (2013) is a fictional movie about a man who falls in love with an AI. The plot shows a recently divorced man who feels lonely and misses his ex-wife, but finds company in a new operating system called Samantha. She exists on his computer and his phone, and they have very human-like conversations, which creates the feeling of her being an actual person. In this movie, AI is portrayed as a system that can be perceived as a real human, and a system that can evolve and think on its own. This is shown in the end of the movie, where the system (Samantha) decides on her own to explore outside the operative system, and leaves.

1.2 Robots and AI

The term “Robot” was first used in 1920, by Karel Čapek in his play R.U.R (Rossum's Universal Robots), although he names his brother as the original inventor of the term. However, the word ‘robot’ was not new, as it had been in the Slavic language for a long time, having the word *robota* with the meaning “forced laborer”. In the play, Čapek told a story about artificial human bodies without souls doing the work that humans did not want to do, and the word *robot* then fit with the explanation of these artificial workers.

Definitions of AI

Definition 1: The Robot institute of America defines 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" (Thrun, 2004).

Definition 2: "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)" (Merriam Webster Dictionary)

The two definitions are different from each other, in the way that the first definition only talks about it being a manipulator and its functionality, while the second definition mentions the machine's resemblance to a living creature as well.

My definition of a robot would be:
"A machine made by humans with the purpose of doing complex tasks" Whether the machine resembles a living creature or not, is not essential to me when defining a robot. The important part is that the machine is able to do complex tasks, with a purpose set by the designer of the machine.

Relation between AI and robots

When talking about a robot, one generally only talks about the tasks that the machine is able to do, while the AI adds the element of human intelligence. A robot is only defined as "a manipulator designed to move materials", it is more of a physical tool in a way. A definition of an AI often includes "mimicking human intelligence" (Verne & Bratteteig) and cognitive skills like "visual perception and speech recognition" (Britannica), which can make an AI more of a collaborative partner than a simple tool, like a robot is.

Movements of the robot vacuum cleaner

A commercial robot that I have experience with, is the robot vacuum cleaner. This is a simple robot, programmed to analyze a room and then clean it by driving back and forth until it has covered the entire space. Usually, the human does not have to interact with it much for it to function. Either, the human user has to program it once with time intervals, and then it will do the job on its own periodically, or the human user has to go turn it on manually every time they want the floors cleaned. These robot vacuums have a "home station" that they return to when finished cleaning the floors, which also acts as a charging station.

1.3 Universal Design and AI systems

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." (Universaldesign.ie)

This definition talks about access, understanding and usability, by all people. These three words are essential when talking about universal design, because when something is universal, that means it is for everyone, regardless of their abilities or disabilities. If something is designed to leave any one user group out, then it is not universally designed.

The potential of AI

AI can be used to aid companies in customer service, by being used in chatbots, where they are to be perceived as humans in customer service, able to help customers with simple tasks and questions. AI also has the potential to help aid

people with movement, like having self driving cars for people who struggle with driving themselves.

AI has the potential to help include everyone, as they can help people with difficulties within perception, movement and cognition. For example voice controlled AI (like Siri) are used a lot as a tool for visually impaired people, to help navigate as well as perceive information. Voice controlled AI can also end up excluding people, for example people who have trouble speaking or have a speech impediment.

Understand and understanding

To me, the concept of "understanding" something means that you perceive the *meaning* of what is said or done. I see this as a human skill, and not something machines can do to the same extent. When a machine "understands" a command, it only reads it and its following actions are then programmed, it does not have its own personal understanding of the word or the action that has taken place. I think that a machine can be perceived to understand something, but not truly understand it as it only "understands" what it is told that something means.

1.4 Guideline for Human-AI interaction

Guideline from Microsoft

G13: Learn from user behavior.

AI should learn whether a user is experienced or inexperienced, and then respond thereafter so as to aid the user in the best possible way, whether that be explain more in detail or skip explanations all together.

Similarities and differences between the HCI design guidelines and the Human-AI interaction guidelines

I have chosen Donald Norman's seven principles of interaction between human and computers. There are a lot of similarities between the two sets of guidelines, as they both write about the importance of the user understanding. Donald Norman's guideline "*Make things visible*" and the Human-AI interaction guideline *"make clear what the system can do"* are both about showing the user what the system's abilities are. When it comes to differences, Donald Norman's principles are more specifically about understanding, whereas the Human-AI interaction guidelines also cover social elements and efficiency.

Characteristics of AI-infused systems

When talking about Artificial Intelligence (AI), we differentiate between three different types of AI ;

- Artificial super intelligence: When doing something that matches or goes beyond human capabilities.
- Artificial general intelligence: When the AI mimics general human intelligence.
- Artificial narrow intelligence: When the AI focuses mainly on simple and narrow tasks.

The AI that we are familiar with in today's society is the artificial narrow intelligence, which is what I will be referring to when writing about AI-infused systems and their characteristics in this text.

Characteristics of AI

Følstad mentioned, during the first lecture of module 2, four key characteristics of AI-infused systems; Learning, improving, blackbox and fuelled by large data sets.

Learning within the area of AI implies that the system is very dynamic, and that it is designed for change. An AI is never "complete" when it comes to learning about the users and how to cater to them. Amershi et al (2019) writes about how these systems often have an element of uncertainty, where errors are common, often producing false positives and negatives, which can be confusing and offensive to the users.

Improving refers to how the AI evolves through the learning process, and becomes better at understanding the user and making the right actions. A way to learn would be through making mistakes, and then getting feedback from the users on these. Amershi et al (2019) mentions how AI-infused systems often are personalized in the way that they have filters built in "behind the scenes", that act on behalf of their users, such as filtering content. However, they might not always be aligned with the users' preferences, which is something that often can be improved through the system learning about the user and its preferences.

The Blackbox refers to how AI is viewed by many, as black solid boxed, where you are unable to see what happens inside. The "phenomenon" of AI can be very confusing and difficult to understand, when all you see as a user usually is the result. This also makes AI-infused systems difficult to use, especially for inexperienced users. This is supported by Kocielnik, who then suggests that "providing explanations will lead to higher perceptions of understanding how the AI system works" (2019).

Being *fuelled by large data sets* means that the AI learns through receiving loads of information, collected through interaction with users. Through receiving this data, either actively or passively, the system improves over time to learn how to cater better to the users.

Netflix

I have chosen the service Netflix and its recommendation-system as my AI-infused system. This service helps you find new movies or series by suggesting some based on what I have watched before. As well as my view-history, they also base their recommendation on whether i liked a movie/series or not, which I tell them through using thumbs up ("I liked this") or thumbs down ("this was not for me"). Netflix allows you to have several profiles as well, so that I can get individual suggestions based on only my preferences, and not the rest of my family as well. I find the system to be very intuitive, as the symbols for "like" and" dislike" is very universal, and its placement next to the movie/series makes it easy to understand what you are reacting to. It also is clear that the system learns through user feedback, as the suggestions become better and better consistent with the amount of feedback I give. However, it is not clear *how* Netflix chooses the movies/series that they suggest to me, and so the blackbox-element of AI is very relevant here.

Human-AI interaction design

Amershi et al. (2019) propose 18 design guidelines for human AI interaction as a way to better the communication within a field that is currently advancing quickly. They present several issues with human AI interaction, such as producing false results, hidden personalization and unexpected changes that come with misunderstood learning over time. With these guidelines, they are hopeful that it will "result in better, more human-centric AI-infused systems", and that their synthesis can facilitate further research.

In the Kocielnik et al. (2019) paper, they "explore techniques for shaping end-user expectations of AI-powered technologies prior to use and study how that shaping impacts user acceptance of those technologies". They focus on different types of

methods for setting expectations before initial use of an AI-based system, more specifically they use an Scheduling Assistant to explore this subject. They explore two different versions of the Scheduling Assistant, designed with two different focuses on the types of errors to avoid.

Guidelines

I have chosen guideline 1 and 9, where I will discuss how Netflix meets or deviates from these guidelines.

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

I believe that Netflix could do better on this guideline concerning their services. It is not obvious that it is possible to use the “thumbs up” and “thumbs down” functions, as they are never presented when joining the service. They are also a bit hidden from the user, and many are not aware of this functionality. I believe that it would better the user experience if more users were made aware of the functionality, as they would most likely receive better recommendations when using it.

G9: Support efficient correction. ***Make it easy to edit, refine, or recover when the AI system is wrong.***

If Netflix suggests a movie/series that I do not like, I have the option to mark it as “not for me”. What happens then is that the movie/series is “greyed out” and made to stand out from the rest. This is a very easy and efficient way of marking that the AI made a bad suggestion, and then simultaneously make me aware to avoid it in the future. However, I believe that the best way would be for it to be removed from my suggestions all together, as I clearly am not interested in watching it. I tried to refresh the service, but it was still present in greyscale.

Chatbots / conversational user interfaces

There are several implications that can occur when designing a chatbot, and I will firstly discuss three implications presented by Følstad during our lectures, as well as findings from a paper written by Yang, Steinfeld, Rosé & Zimmerman.

Conversation as design object

When designing other user interfaces, the focus is usually on different graphical design features such as navigation and interaction with elements. With the user interface of a chatbot, it is much more blank where most services are hidden from the user. This creates a difficult task for the designers, which have to move from design being an explanatory task, to seeing design as an interpretational task of understanding what the user needs, and then how this can then be presented.

Necessary to move from UI design to service design

When designing chatbots, the attention has changed from previously being focused on the user's goals, towards a focus on the interactive system, much like the ways of service design. Chatbots will be differentiated by their convenience in accessing the context of conversational threads (Følstad & Brandtzæg 2017).

Necessary to design for networks of humans and bots

When designing for typical HCI systems, one usually designs with one user and one system in mind. However, the future of the use of chatbots within systems can lead to the need for multi-agent systems, systems where both human users and chatbots will work together. This already exists in several places, and they have encountered problems surrounding their communication. Følstad & Brandtzæg mentions a case where different bots on Wikipedia went around reworking each other's work on articles, which is very unnecessary and unproductive.

Rapid prototyping and testing

When working with HCI, one core practice is" rapid prototyping, assessing the human consequences of a design and iteratively improving on it" (Yang, Steinfeld, Rosé & Zimmerman, 2020). However when working with AI, this is very difficult as the process of developing an AI is very intricate and comptex. One can never fully anticipate all the different developments and consequences that can occur during the learning process of an AI. These consequences can be harmful, an example being the chatbot Tay from Microsoft who ended up being a racist and mean chatbot after learning from users wanting to harm it.

Adherence to guideline 1 and 2

I think the main problem with chatbots today, is that the users do not understand what the chatbot is created to do. Many users have unrealistic expectations about its functionality, and therefore can end up angry and annoyed after having tried to use a chatbot. If they have a certain issue, and are met with a chatbot that implies that it is there to solve the problem for them, but ends up not being able to, the user will feel upset and that they have wasted their time. This will reflect badly both on the company/system using the chatbot, as well as the concept of chatbots itself. If guideline 1 is followed, this might be avoided in a lot of situations. If the chatbot makes clear what it is capable of doing, the user might understand when the chatbot is not able to help them, before spending time trying.

This is also connected to the issue of "how well can the chatbot do what it can do". For example, if a chatbot says that it can book cinema tickets for you, but then doesn't disclose that it is not able to take into account if the user is handicapped, then it can be frustrating to use for those in that situation. It is therefore also not only important to make clear what the system does, but also guideline 2, make clear how well it can do it. If a chatbot has limits, for example when a chatbot is in its learning stages, it will with certainty make a lot of mistakes. If the chatbot discloses

this as an introduction to the user, the user will not be surprised when the chatbot meets its limits, but rather be prepared and more understanding of the situation.

Module 3 - Human AI collaboration

Mental models refer to structured, organized knowledge possessed by humans that describe, explain, and predict a system's purpose, form, function, or state (Phillips et al., 2016). These mental models impact how people interact with the world around them, and they are continuously being updated. This is especially important to take into consideration when talking about systems with which people are unfamiliar, such as AI and robots. Phillips et al. explore in their article why the analog of a human-animal team can be a useful mechanism for supporting human mental models of robots. For instance, humans generally feel like they have authority over animals, and therefore this can impact the interaction by creating a natural interaction hierarchy between humans and robots. Robots however differ in their levels of autonomy, and the new goal is to seek high levels of human control AND high levels of automation, which is more likely to produce computer applications that are Trusted, Reliable & Safe (Schneiderman, 2020).

I will further present two different robots, both which have taken the shape of animals, and explore their functionality, their levels of autonomy through using the 10 levels of autonomy from Sheridan and Verplank, and their explainability.

BigDog

BigDog is a legged robot designed to function essentially as a pack mule and traverse terrain not accessible by wheeled or tracked vehicles (Philips et al., 2020). It was created by Boston Dynamics in 2005, with the intention of using it in military situations to assist soldiers. The legs of the robot contain a variety of sensors, as well as a laser gyroscope and stereo vision.

Level of autonomy:

The goal of the program behind BigDog is to load the robot with several hundred pounds of supplies and have it follow a dismounted soldier over rough terrain (afcea.org). Another goal for the robot is for it to be able to follow a soldier at a preset distance down a path, and also for the soldier to be able to change this distance with a command. To define which level of autonomy the robot is at, I need to divide its functionality. The balance and movement of the robot is close to level 10 as it is described in Schneiderman (2020), as the robot keeps its balance when moving without any input from the human. However the navigation is on level 1, as the human makes all the decisions on where to go. If they were to increase the level of autonomy on the navigation, this could mean that the robot would be able to adapt to the situation around and maybe aid the soldiers even when they were not able to tell it to do so. If they decreased the level of autonomy, it might struggle with being able to assess the terrain and make its own decisions on how to be able to traverse it, needing help from a human with every decision.

Explainability:

Hagras (2018) defines XAI as an AI in which the actions can be easily understood and analyzed by humans. The functionality of the robot is explained well through its look and movement. It is shaped a little like a mule, which is an animal typically used by humans to carry things across a distance. Its four legs also communicate that it is supposed to walk. I therefore believe that BigDog can be defined as an XAI because its actions can be understood. However, I am not familiar with the technology behind it so I can not really make a comment about that, but I don't think that this is not that relevant to the user of the robot as they do not need to understand it to be able to use it well.

Paro

A small seal-like robot designed to alleviate depression by providing companionship to elderly individuals (Philips et al., 2020). Paro has been found to

reduce stress, improve the patients relaxation and motivation, it stimulates interaction between patients and caregivers, and improves the socialization between patients, as well as between patient and caregiver (Paro robots). This is an example of a robot where appearance is very important to the design, and the attributes are made to match the animals characteristics.

Level of autonomy:

Paro can move its limbs and body, and it responds to petting by moving its tail and opening and closing its eyes. It actively seeks out eye contact, responds to touch, cuddles with people, remembers faces, and learns actions that generate favorable reactions. It also responds to sounds and can learn names, including its own. Using the 10 levels of autonomy by Sheridan and Verplank (Schneiderman, 2020), I would place Paro on level 8. It makes its own decisions about how to behave, but can inform the humans about it if they interact with it. If the level of autonomy would increase, then it means that the robot would decide everything on its own, ignoring the human. The robot would then lose its primary functionality, as it would not be a two-way interaction with the human. If the level of autonomy decreased, it would lose a lot of its personality as the human would need to approve every action, and it would not be able to resemble an animal in the same way, but end up being more of a static robot instead of an interacting one.

Explainability:

I think Paro is an XAI because its behaviour is visible to the user, and you get feedback through actions when you interact with it. Considering the primary users of the robot is elderly, who can suffer from illnesses, memory loss and other disabilities, I believe this level of explainability is all that is needed for this kind of robot.

Feedback

Iteration 1:

I got feedback about the fact that I did not make my own titles, but rather used the questions in the assignment. I will consider changing this for the final delivery, but as of now I will keep them in to better be able to navigate around as well as make changes to the assignment before final report delivery.

I also was asked to provide more about the people giving the definitions (of AI i presume), so I have tried to add in some information about that.

Iteration 2:

The only feedback I got was that there was a paragraph that was split between two pages so that it made it difficult to read, so I have now made sure that the paragraph is all on one page.

References

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

Verne, G, Bratteteig, 2018, *Does AI make PD obsolete?; exploring challenges from Artificial Intelligence to Participatory design*, https://dl.acm.org/citation.cfm?id=3210646

Frankenfield, J, *Artificial Intelligence (AI),* retrieved (09.09.20) from: https://www.investopedia.com/terms/a/artificial-intelligence-ai.asp#:~:text=Artificial%20intelligence%20(AI)%20refers%20to,as%20learning%20and%20problem%2Dsolving

Copeland, B, *Artificial Intelligence*, retrieved (09.09.20) from: https://www.britannica.com/technology/artificial-intelligence

Computas, *Kunstig intelligens' rolle og funksjon*, retrieved (10.09.20) from: https://computas.com/tjenester/skyteknologi/kunstig-intelligens

Thrun, S., 2004. *Toward a Framework for Human-robot Interaction*. Hum.-Comput. Interact. 19, 9–24. https://www.tandfonline.com/doi/pdf/10.1080/07370024.2004.9667338

Merriam Webster Dictionary, *Robot*, retrieved (10.09.20) from: https://www.merriam-webster.com/dictionary/robot

Universaldesign.ie, *What is universal design*, retrieved (10.09.20) from: http://universaldesign.ie/What-is-Universal-Design/

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)

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)

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)

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 the 2020 chi conference on human factors in computing systems (Paper no. 164). (https://dl.acm.org/doi/abs/10.1145/3313831.3376301)

Henry S. Kenyon, Afcea.org. *BigDog Steps Into Action*, (retrieved 15.11), from: https://www.afcea.org/content/bigdog-steps-action

Hagras, H., Toward Human-Understandable, Explainable AI, Computer, 51, 9, 2018, 28- 36 https://ieeexplore.ieee.org/document/8481251

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 https://www.researchgate.net/publication/306003139_Human-Animal_Teams_as_an_Analog_for_Future_Human-Robot_Teams_Influencing_Design_and_Fostering_Trust

Shneiderman, B., Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy, arXiv.org (February 23, 2020). https://arxiv.org/abs/2002.04087v1

Parorobots.com (Retrieved 15,11), from: http://www.parorobots.com/