IN5480 - Specialization in Research in Design of IT: *Individual assignment*

By Thea A. Nordgulen (Theaano)

Module 1: Individual assignment

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

The story of artificial intelligence is according to the article “AI & HCI: Two fields divided by a common Focus” (2009: 48)) by Grudin “ ... a social history, focusing on the forces affecting interdisciplinary work that spans the two fields”. AI has its origins back to the 1950s, when World War II created an engagement in code-breaking (Grudin (2009: 48)). Alan Turing was one of the first to comment on artificial intelligence. He believed that computers could cover fields of human intelligence and that they could compete against each other (Grudin, 2009, p.49). The term ‘artificial intelligence’ originates from a workshop in 1956, written by John McCarthy (Grudin, 2009, p.49).

In 1960 came the first argumentation and translation programs (Grudin, 2009, p.49). In addition, the publication of the article “Man-computer symbiosis” by J.C. R Licklider drew much attention to artificial intelligence and the use of computers to achieve this (Grudin, 2009, p.50). In the 1960s, the view of AI was very optimistic and researchers had big plans for what AI could do (Grudin, 2009, p.50). I. J Good also commented on how human survival depended on an extremely intelligent machine that would be built within the 20th century, which would then be the last thing humans themselves needed to build (Grudin, 2009, p.49).

In the 1970s to the early 1980s, there was a great decline in interest in artificial intelligence (Grudin, 2009, p.52). In 1977 an empirical study of “natural language understanding” was published by an AI group that also began to establish itself within HCI, where among others Winograd, Norman and BoBrow were very involved in the study (Grudin, 2009, p.52). In the early 80's, however, AI became a popular theme again. Turing's attitude remained relevant, where “machines would soon rival human intelligence, then educate themselves 27/7 and leave homo sapiens in the dust” (Grudin, 2009, p.53).

In the 1990s, there was again a decline in interest in artificial intelligence. A turning point Grudin mentions is when Deep Blue's in 1997 beat chess master Kasparov in a battle with chess. Today, AI and general machine learning can be utilized by ordinary computers through browser access to servers (Grudin, 2009, p.55). In addition, another wave of great interest in AI came with the internet. It was now estimated that it would take 30-40 years before ultra-intelligence could exist (Grudin, 2009, p.55).

McCarthy (1998)

“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's (1998) definition is based on the fact that artificial intelligence is a scientific field in itself and a design of the creation of intelligent systems. Here one can see artificial intelligence in light of the fact that one uses computers to understand humans and that its intelligence is not necessarily visible to us. McCarthy uses a perspective that is very technical. He draws a line between man and machines, where he points out that intelligence has a fundamental difference, where one can be used to understand the other.

Bratteteig & Verne (2018)

"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 and Verne's (2018) definition focuses on how one wants artificial intelligence to be able to imitate human intelligence. Bratteteig and Verne, who both work in the field of Participatory design, carry a discourse of the methodology when it comes to the definition of artificial intelligence. Here it can be seen in the context of how artificial intelligence is directly aimed at humans and not a machine perspective.

High-level Expert Group on Artificial Intelligence (2019)

"*Artificial intelligence (AI) refers to systems that display intelligent behavior by analyzing their environment and taking actions - with some degree of autonomy - to achieve specific goals. AI-based systems can be purely software-based, acting in the virtual world (eg voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (eg advanced robots, autonomous cars, drones or Internet of Things applications)* ".

The High-level Expert Group on Artificial Intelligence, on the other hand, defines artificial intelligence as part of a theory and framework for analysis (2019). They link it to a more technical perspective, where AI can take the form of both software and hardware. They also mention that artificial intelligence can be seen as a way to achieve a specific goal, through using AI's ability to analyze.

My definition:

"Artificial intelligence is defined by implemented algorithms that address different aspects of human intelligence and can be used as tools to effectively analyze and calculate how to mimic human intelligence to achieve a goal. Artificial intelligence can be in the form of both hardware and software, but is based on autonomous systems that are pre-implemented".

Artificial intelligence can be all possible areas within algorithms that address the imitation of human intelligence. We can use AI to simplify solutions and analyze content, where AI continuously builds on each other and itself in order to optimize further development. Nevertheless, AI is implemented by humans and will always have to behave within a framework that humans have established.

The company ‘Analytics Software & Solution’ (SAS) is a company that works with artificial intelligence. On their website, they point out how they want their software to contribute to increased efficiency, creativity and new opportunities. They refer to AI as an automated solution that is intelligent. They mention that "... our AI technologies support various environments and scales to meet changing business needs", where they want others to use their solutions which include machine learning, 'natural language processing', forecasting and optimization. They thus offer AI software that can be used for analysis and the creation of new solutions, which are used by others, as a product.

The episode ‘Hang the DJ’ is an episode in the fourth season of the series ‘Black Mirror’. Here is a digital solution, where the people who live in the world where the episode takes place, are put into a simulation of their own love life. Here they have an artifact called 'Coach', which consists of a small, round screen that can interact with the "players". Human interaction with AI is here projected by humans living in a simulated world, where calculations are made based on the conditions they go through during the simulation. They thus interact unconsciously with the system through action, but through ‘Coach’ they can interact directly with the system.

‘Coach’ tells the “player” when they have a new relationship and they can, among other things, check how long they will be with the new girlfriend. In the episode, the main characters Frank and Amy find each other and agree not to check the time they have together. Frank gets stressed because he thinks Amy is the perfect one for him and therefore breaks the promise between them. Thus, the time they have with each other goes down at Coach. They therefore want to escape from the simulation, but are caught in the simulator forever.

1.2 Robots and AI systems

The term 'Robot' was first used in a play called 'Rossum's Universal Robots' (RUR) by Karel Capek in 1921. According to Wired, in the article 'The Wired Guide to Robots', the word 'robot' comes from the Czech Republic and can be directly translated to 'forced labor' (Simon, 2020). Nevertheless, to this day it is very controversial what a robot actually is. In the article “Some brief Thoughts on the Past and Future of Human-Robot Interaction” by Dautenhahn it is mentioned that what we define as robots is highly dependent on context and in what technological time perspective we are in (2018, 4.2). Thus, according to Dautenhahn, we do not necessarily have a clear reference to what a robot actually is and do not even have “... a sample of widely agreed-upon reference points” (2018, 4.2). Robots are a constantly evolving topic and it is therefore difficult to define the robot's history to date.

Robots.IEEE

“A robot is an autonomous machine capable of sensing its environment, carrying out computations to make decisions, and performing actions in the real world” (Robots.IEEE (2020)).

The Oxford dictionary: Robot

"A machine capable of carrying out a complex series of actions automatically, especially one programmable by a computer" (Lexico (2019)).

Both of these definitions are based on the fact that a robot is a machine. IEEE refers to how machines can mimic human senses and take them as input to interpret and calculate the environment. Thus, they can also perform actions in the real world in a similar way that humans can. Here, the IEEE refers to how robots can interpret the outside world and how this can be done on a general basis from a technical perspective on input from the environment. Oxford dictionary, on the other hand, defines robot as something purely technical and a programmed machine that can perform complex actions. It is emphasized that they can most easily perform actions that can be programmed and do not connect it directly to human characteristics.

My definition:

"Robots are machines that have functionalities that can take input from the outside world. Robots also have an ability to interpret inputs and apply them to the benefit of interacting with the environment on an autonomous level. Machines are particularly suitable for performing simpler physical tasks, but by using metrics for artificial intelligence, this room for maneuver will increase considerably".

In my definition, I want to focus on the fact that robots interact with the outside world and that its ability to interpret any input is essential for machines to be able to perform tasks that give robots exactly the uniqueness it has in being able to mimic human behavior and movement. In Schulz's article, they look at how movement in the home and "classified the movement in relation to humans and their movements" uses animation to make the robots more familiar and easier to understand, which will also provide a good starting point for the future (2018, p.242). Research related to such problem areas is important when it comes to expanding the field of robotics.

There are some fundamental differences between artificial intelligence itself and robots. Robots can have artificial intelligence, where algorithms for that embrace it to mimic human intelligence, something we see in the definition of McCarty (1998). Yet AI does not have to be "biologically observable" (McCarthy, 1998), where robots try to perform actions in the real world (Robots.IEEE, 2020) and one can possibly see this as a design of the biologically observable, to which McCarthy refers. In addition, a robot is defined by both the IEEE and the Oxford Dictionary as something that will perform actions. AI must be able to analyze and carry out actions in order to achieve specific goals and be in the form of both software and hardware (European Commission, 2019), where robots, on the other hand, must take input from the outside world in order to carry out actions in the physical world.

We see that Norman, in the article «'The problem with Automation: Inappropriate Feedback and Interaction, not 'Over-Automation' (1990)" links intelligence levels to robots. He believes that the intelligence level of robots is not high enough and that one must either improve the intelligence or lower it, so that it should be appropriate for humans to use it. Both AI and robots use both similar and different perspectives from human intelligence, something we see in virtually all definitions except Oxford Dictionaries definition of robots. Artificial intelligence is about performing tasks that human intelligence can perform (McCarthy, 1998) and preferably in a more efficient way. In addition, it involves algorithms such as voice recognition, translation, visual perception and decision making (Lexico, 2019), where robots can be programmed to use artificial intelligence to collect input from the outside world in which it will act.

The robot dog ‘Sony Aibo’ has built-in visual perception, so that it can recognize rooms and it can also remember actions and words. Sony (2017) mentions in the article ‘Entertainment Robot‘ aibo ’Announced’ that “aibo can form an emotional bond with members of the household while providing them with love, affection, and the joy of nurturing and raising a companion”. The owner of the robot dog can interact with Aibo by having built-in sensors such as "... detect and analyze sounds and images". Thus, Aibo can also interpret facial expressions, words and recognize faces (Sony (2017)). Aibo uses artificial intelligence to interpret these impressions in the form of inputs, and will build up a more complex impression of the surroundings and increase the possibility of interpreting them. Sony instead tried to mimic the movements of a dog. The owner can interact with the robot dog by talking, patting it and through facial expressions.

1.3 Universal Design and AI systems

Universal Design (2020)

“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. An environment (or any building, product, or service in that environment) should be designed to meet the needs of all people who wish to use it. This is not a special requirement, for the benefit of only a minority of the population. It is a fundamental condition of good design. If an environment is accessible, usable, convenient and a pleasure to use, everyone benefits. By considering the diverse needs and abilities of all throughout the design process, universal design creates products, services and environments that meet peoples' needs. Simply put, universal design is good design. ”

Universal Design's (2020) definition of universal design is based on creating solutions that can be used by most people, including that most people should be able to understand how to use a solution and that it is accessible to most people. The definition emphasizes that one should base the development of the solution based on requirements from needs to the general public. What is meant by this is that one should design in the best possible way in every aspect for as many people as possible. Universal design is generally about including as many people as possible in its solution, which Universal design thus includes in its definition, where the design should make it usable, accessible, have a satisfactory design and convenient to use (Universal design, 2020).

One can use algorithms that, for example, interpret visual perception for the visually impaired, and thus can interpret the room so that they can orient themselves around just as other people could have done. We can also use AI to recognize movement patterns in people with a form of physical disability, among other things, and thus use the analysis of this to create solutions that can be used for as many people as possible. One can also use AI to, among other things, recognize faces (European Commission, 2019), and thus analyze this to detect how people react to a certain action and then adapt the actions performed based on the interpretations the artificial intelligence calculates.

Artificial intelligence can also exclude humans. Wired documents a case in the article "The best Algorithms Struggle to Recognize Black Faces Equally" (2019). Here, artificial intelligence contributes to exclusion, where several of the very best face recognition algorithms fail to identify people who are dark-skinned, as well as people with lighter skin (Wired, 2019). If AI is not built from scratch with a focus on including as many as possible, such exclusion will occur frequently. AI can also contribute to inclusion, among other things by a solution based on voice recognition in the home can contribute to the inclusion of those with physical disabilities who do not have the opportunity to move to the same degree.

The term 'understanding' means giving meaning to the environment, situation and / or context one is in. Thus, they can interpret this information and decide how to act on it. This can also be transferred to computers, where the computers can interpret different sources of information based on the algorithms they can use. Thus, they will be able to conclude and draw conclusions based on this information. Nevertheless, I believe that humans have an ability to understand aspects that machines cannot yet understand, which are not only based on logic and rationality, but also emotions and empathy. In other words, humans have a unique ability to detach themselves from the iron cage of rationality and act on human terms.

1.4 Guideline for Human-AI interaction

Microsoft Guideline 4: "Show contextually relevant information - Display information relevant to the user's current task and environment".

This guideline is based on using AI to show relevant information to the task the users complete there and then. This can be done, among other things, by creating a list on Spotify. When there are no more songs in the list, Spotify's algorithm will resume playback based on which songs you have added to the list and the title of the list. You will also see a relevant list during playback of the list.

I chose 'Norman's Seven Principles' as my HCI design guidelines (Batterbee, 2020). Similarities between Norman's seven principles for HCI and Microsoft's Guidelines for Human-AI Interaction are mainly the focus on easy navigation through the solution and feedback to the user. This is consistent in both lists and is repeated in many of the points, especially at Microsoft, but also at Norman's seven principles.

Module 2: Individual assignment

1 Characteristics of AI-infused systems

In the mandatory curriculum of this assignment, the scope is targeted towards how AI-infused systems are hard to design. Amershi (2019), Kobielnik (2019), Yang (2020) and Laio (2020) all have a common aim of identifying and understanding why this is such a difficult task and how to deal with this issue by providing guidance and framework through research on the field. I will explain and identify the key characteristics which are presented through the articles mentioned above and future discuss them through an example of an AI-infused system known today and how these characteristics can be drawn to the area of this system.

In the article 'Guidelines for Human-AI Interaction' (2019), Amershi et al addresses the issue of AI-infused systems probability of acting in unpredictable ways and how it is a significant challenge to identifying these issues and the uncertainties in how to handle such challenges (2019, s.1). AI-infused systems are defined by Amershi as "...systems that have features harnessing AI capabilities that are directly exposed to the end user (Amershi et al, 2019, p.1). The article also argues how the components of AI often are considered insufficient due to the lack of understanding of the capabilities of artificial intelligence (Amershi, 2019, p.1). Related to the lack of understanding, is the issue of management error prevention (Amershi, 2019, p.2). To understand the sources of errors, you need to understand the behaviour of the system.

Amershi et al also provides 18 guidelines in designing human-AI interactions, and argues how the guidelines are a result of balance between specialization and generalization (2019, p.11). The guidelines should provide insight on four top-level categories 'Initially', 'During interaction', 'When wrong' and 'Over time' (Amershi, 2019, p.4). The goal of the paper is to offer guidance to the general designer when designing human-AI interactions. The multiple categories of guidelines might relate to the key characteristic of AI-infused systems being quite complex and thus hard to design and conduct.

Kocielnik et al argues how expectations of the end-user to AI-infused systems might not be accurate and addresses the issue of not conveying what the system actually is able to do (2019, s.1). The article provides guidance for shaping the expectations of the end-user of such systems and establishes three different techniques for doing so. These techniques should help by setting the expectations of the user to a more accurate level and make the use of AI-infused systems more efficient. Characteristics of AI-infused systems which are covered in Kocielnik et al, is perception of accuracy and high recall (2019, s.2).

Perception of accuracy relates directly to the user's perception of how accurate the system is, while high recall is connected to the occurrence of false positives in the system (Kocielnik, 2019, s.7). They suggest three mechanisms for adjusting the expectations of the user, which is 'Accuracy indicator', 'Example-based explanation' and 'Control slider' (Kocielnik, 2019, p.5). All these mechanisms are established through a general goal of being able to influence the expectations in different ways, through differing design elements of visualisation and description. This shows how a key characteristic of AI-infused systems is the challenge of balancing the preventions of both false positives and false negatives by adjusting high recall versus high precision of accuracy (Kocielnik, 2019, p.11).

Yang et al also identifies how the unpredictable nature of AI-infused systems do in fact make such systems hard to design in a proper manner (2020, p.1). In the article, they mention two sources of challenges related to designing the capabilities of artificial intelligence, which are related to AI's capabilities uncertainties and the complexity of outputs of AI-infused systems (Yang, 2020, p1). Yang et al mentions how this is highly connected to designers' understanding of AI capabilities and how traditional HCI is not properly suited to design AI-infused systems (2020, p.3). Yang et al further contributes to a framework which identifies these attributes and provides a method to analyse AI-infused systems through four levels combining the appearance of capability uncertainty and output complexity (2020, p6).

The article "Questioning the AI: Informing Design Practices for Explainable AI User Experience" by Liao et al (2020) addresses the issue of making artificial intelligence understandable and covering the gap between the algorithms, the practices of handling AI-infused systems (Liao, 2020, p.1) and the user needs (Liao, 2020, p.7). This, again, emphasizes how AI capabilities are difficult to design, due to the lack of understanding and knowledge in the field. In the article, Liao et al provides an explanation framework through a 'Explainable Artificial Intelligence' (XAI) question bank (2020, p.6). They identify seven categories in a framework for guidance of making the capabilities of the AI-infused system explainable and how this accordingly is related to the user needs (Liao, 2020, p.9).

An AI-infused system which relates to the framework provided by Yang et al (2020) is the movie-recommendation function of Netflix. The AI capabilities are responsible for calculating which movies you might like, but how this is done could be quite unclear for the designer with insufficient knowledge about the algorithm. At the same time, the system is quite complex comparing thousands of movies and series based on previous watchings. This might be related to a 'level 2' system in which Yang et al (2020) presents. There is also a challenge in the balance between high precision and high recall, where the recommendations might or might not be appropriate for the user.

In the lecture «Interaction with AI – Module 2» Følstad (2020) identifies tre charaterictics inspired by the guidelines presented in Amershi (2019). 'Leaning' is the first identified characteritic, where the main focus is on providing a dynamic learning curve (Følstad, 2020, p.13), making clear what the system can / can not do and how well this can be done (Amershi, 2019, p.3). The next charactertic presented in the lecture, is 'Improving', which refers to the guideline grouping of «when wrong» (Følstad, 2020, p.24). The AI-infused system should be able to improve when mistakes are done and potential feedback is given. 'Black box' and 'Fuelled by large data sets' are the two remaining charactericsts presented in the lecture, which referes to the uncertaincy of how the system works and how it calculates the output (Følstad, 2020, p.30) and the concern of using data appropriately according to the assigned issue and doing this in a right manner of safety (Følstad, 2020, p.38).

Netflix's recommendation capabilities are directly exposed to the user, as Amershi et al (2019) argues is a characteristic of AI-infused systems. According to the guidelines, the recommendations are quite clear and offer recommendations in a 'movies you might like' -manner. As an end user, there are few or none uncertainties about what this function is capable of, but as a designer it might be hard to adjust the algorithm in a more clear and efficient way. This can also be viewed in the light of Liao et al (2020), where there are few explanations on how the recommendations are made both to the end user and designers.

2 Human-AI interaction design

In the articles of Amershi et al (2019) and Kocielnik et al (2019), the issue of how to properly design human-AI interaction is addressed. Kocilnik et al (2019, p5) focus on guidance on how to shape the expectations on AI-infused systems by taking account for high precision systems (false positive) and high recall systems (false negative). The main takeaway of the article is that as a designer, you have to make clear the capabilities of the system by providing visualisation and description on what the end user can expect. This will contribute to adjusting the expectations of the user more equal to the actual capabilities of the system (Kocielnik, 2019, p.12).

While Kocielnik et al mainly focus on how to shape the expectations, Amershi argues in a slightly more general manner how the interaction of the system should be managed. Amershi et al discuss the four categories which frame the 18 guidelines in a way of what the system should provide when in a specific state of when the guidelines should be applied and in the given situation (Amershi, 2019, p3). Both Amershi et al (2019) and Kocielnik et al (2019) carried out complex studies trying to carry out these results and they argue how it is a difficult challenge to even predict and manage the behaviour AI capabilities. The article provides two approaches of more efficiently designing Human-AI interactions and what might become a challenge through their suggestions of guidelines and mechanisms.

The first guideline of Amershi et al (2019), which I chose, is G9 "Support efficient correction". As discussed, the recommendation function of netflix does not take account for false positives and might show movies you would not like. Netflix do not let the user edit or suggest filters for this recommendation, which means Guideline 9 is not followed. They might improve the system by letting the end user adjust this recommendation by providing filters and settings for it. The second guideline I chose is G13 'Learn from user behavior'. I think the algorithm of Netflix recommendation function is interesting, because it will interpret for example the genre, theme, length, etc of the movie/series which you are watching and make up recommendations of the data. The more movies and series you watch, the more the algorithm will learn about the user's preferences and further make more accurate the recommendations based on them.

3 Chatbots / conversational user interfaces:

The occurrence of chatbots are relatively new to the HCI field and the task of designing them properly requires new approaches to the design of human-AI interaction (Følstad, 2017, p.40). In the article "Chatbot and the

New World of HCI", Følstad and Brandtzæg mention how a key challenge of designing chatbots is making sure the chatbot will answer appropriately according to input (2017, p.42). Another challenge is doing the accurate calculations of the algorithms, so the chatbot will be able to interpret input from the user (Følstad, 2017, p.41). In designing chatbot we must also shift the focus from designing an object, to designing a service (Følstad, 2017, p.41). Due to this change, we have to refocus to make the user reach their goal and provide data-management and accurate interaction to solve the problem of the user.

A more technical challenge could be understanding how the algorithms of different chatbot operates and how to control them in a given context. In the lecture "Interacting with Artificial Intelligence" (2020), Goodwin argues how the interplay between input-, output and hidden layers make up the deep neural networks of which more technical chatbots operate. In Appendix 2, we got to experience this complexity and how getting to know the capabilities of the AI is essential for further design of Chatbots. Designing more complex chatbot requires more technical knowledge, but also aspects of psychology related to perception and interpretation in the future.

Guideline 1 'Make clear what the system can do' and guideline 2 'Make clear how well the system can do what it can do' presents guidance which should be applied initially in interaction between humans and AI (Amershi, 2019, p.3). Contributing to explanations of what the system can do is essential to the user for understanding what capabilities the system provides. When missing out on what functionality a chatbot has, it might seem useless or less useful to the end user. It can also result in the rejection of the chatbot. In providing a clear description on what the chatbot is able to do, the use of it will be way more efficient and the area of problem solving will be clear and might be considered a useful tool.

Guideline 2 might be a bigger challenge to accommodate when implementing a chatbot. It obviously depends on the complexity of the chatbot, where the output complexity will vary accordingly to how many layers and databases of which it is "built". It might be useful to draw lines from Kocielnik et al (2019), where the mechanisms for adjusting expectations. These mechanisms could be used in manner of visualising and/or describing how well the Chatbot performs its capabilities, which also could result in higher user satisfaction. It is useful to the user to know whether he/she could interact with the chatbot in a 'one-way' kind of communication or if the chatbot performs complex machine learning and problem solving.

Module 3: Individual assignment

In the article 'Human-Animal Teams as an Analog for Future Human-Robot Teams: Influencing Design and Fostering Trust' by Phillips et al, they are presenting how to use the interactive partnerships between humans and animals to establish mental models (2016, p.103). The mental model should provide a starting point for the functional requirements, implementation and use of Robots as a teammate and not just a tool (Phillips, 2016, p.102). This approach suggests a contribution to align the general mental model of assumed expectations of what a robot is capable of performing and what it actually can do.

Phillips et al address this by looking at how the structure of the relationship of Human-animal teams unfolds and identifies the elements which create both benefits and limitations for the proposed robot solutions (2016, p.103). The aim of this view is to adopt the process of which "Service animals can augment and extend natural human functions" (Phillips, 2016, p.105) . Further they provide a suggestion of developing the potential of robots as team mates similar to the partnership between animals and humans in task performance in the physical, emotional and cognitive fields (Phillips, 2016, p.118). In this assignment, the chosen point of convergence is the emotional and cognitive aspects of human-robot teams.

The emotional aspect of Human-Animal teams address the benefits of indefintication of 'human social' skills and the interplay between humans and animals in developing social perception capabilities and competence (Phillips, 2016, p.105). Phillips et al mentions how animals could benefit as teammates by " ...providing comfort and companionship" (2016, p.105), but they also stress how it is a significant challenge to imitate such human perception and consideration of emotional interpretation. Characteristics of animals as supportive and loving creatures increase a humans willingness to trust the animal (Phillips, 2016, p.106), which has a potential of adoption in human-robot partnerships.

The issue of this area is to enhance and identify the social and emotional elements of which animals are emotionally beneficial to **humans**. The main issue is how these elements can be adopted to allow robots to be comforting, trustful and provide a feeling of closeness to the emotions of humans (Phillips, 2016, p.106). Phillips et al also present the ability of animals to detect "... deficiency in emotional capabilities" (2016, p.106) in humans, which could be of interest in the development of human-robot partnership. This serves as an insight to using robots in behavior skill training or emotional therapy for both children and humans with emotional difficulties (Phillips, 2016, p.107).

An example of which is implemented and designed to provide comfort and closeness is 'Ollie the Baby Otter' (Ackerman, 2015). This robot was developed in the course 2.009 - 'Product Engineering Processes' by a student at MIT. 'Ollie' is developed in the cause of therapy, related to depression and anxiety due to dementia (Ackerman, 2015). The partnership with Ollie will keep the patients company and make use of sensors to interpret the touch of the user. These inputs will result in different feedback outputs such as purring and movement (Ackerman, 2015). This human-robot interaction is inspired by the partnership between humans and animals in the shape of an otter and the elements of which "animal-assisted therapy can help reduce stress and agitation, minimize feelings of isolation, and give people something to touch, and be touched by" (Ackerman, 2015).

The other element of focus is cognitive aspects of human-animal teams, concerning the areas of perception and cognitive thinking. Simultaneously there exists a great challenge in capturing the cognitive architecture of which animals and humans are able to perform (Phillips, 2016, p.107). The scope of human-animal partnership is on developing robots with a diversity of cognitive capabilities based on sensory information and interpretation in animals (Phillips, 2016, p.107). This sensory information provides additional information to humans, which might not be able to perceive this information themselves through the capabilities of human senses or replacement for cognitive constraints.

An example of cognitive benefits provided by robots developed through a lens of human-animal partnerships is a robot who is using ultrasound for navigation in " ... unfamiliar environment" (Lavars, 2018). The robot is named 'Robat' and navigates through a sonar system by exuding ultrasonic waves of sound, just like a bat does in the dark to both navigate and for hunting (Lavars, 2018). Robat was the first robot which fully is able to navigate through this kind of technology and promoted a great potential for further development (Lavars, 2018). Echolocation is also used by dolphins, which has served a great potential for the fishing instruy of locate areas of both fish and other sea creatures (Lawrence, 2018).

Levels of automation and reflections

In the article 'Human-Centered Artificial Intelligence: Trusted, Reliable & Safe' (2020) Schneiderman presentes a two-dimensional 'HCAI' framework. The aim of the article is contributing to the discussion of developing " ... high levels of human control AND high levels of automation" (Schneiderman, 2020, p1). The framework provides a table which presents the first dimension 'Control', where the different ends of the scale represents human- vs computer control. The second dimension presents 'Automation', where the scale goes from low to high. Systems which have high human levels of human control and high level of automation (in the top-right area of the table) are considered TLS (trusted, reliable, safe) systems (Schneiderman, 2020, p.8).

Ollie the baby otter provides a high level of automation, where the robot interprets and provides output without the help of humans except turning on the power. Thus, Ollie does not allow for any human control except for the interaction of which the human does with the robot and the 'off' -button. Ollie will then be located in the right-bottom of the two HCAI dimensional framework, which is a characterised system of high levels of computer control and little human mastery, except for input for sensory interpretation. This influences the level of human control and locates Ollie closer to the middle in the bottom-left corner.

Robat is able to move autonomously in an unfamiliar environment. This serves as a tool for humans in environments where humans are not able to navigate due to sight disabilities in given context. In the article, Lavars (2018) mentions how "the work can open the door to an advanced breed of robots that use sound in this way to make their way through unfamiliar environments". This shows the potential of using the concept of Robat as a teammate in the future. Since Robat is supposed to navigate itself, I assume this makes a reference to low human mastery and high level of automation and locates Robat a bit further from the middle in the HCAI framework than the location of Ollie.

An advantage of making Robat more in control of humans, is the ability of earning cognitive benefits of additional sensory of ultrasound. In such a case, humans can navigate in through the interface of Robat though environments the humans themselves are not able to navigate safely. Thus, this might result in higher error rate due to human errors and Robat might serve more as a tool than a teammate. Ollie, for instance, could serve both advantages and disadvantages in adjusting the levels of control and automation. Including more human control, might make the comforting animal-like behavior less "authentic". At the same time, allowing for high levels of automation might promote unsatisfactory and unpredictable behaviour in Ollie.

Since both Robat and Ollie serve little human control and moderate to high levels of automation, the explainability is more relevant to the develop- and maintenance of the robot. At the same time, both of the robots do not, at the current time, need interpretation of high-dimensional inputs (Hagras, 2020, p.29). Thus, Ollie presents a need for an XAI model which presents the capabilities, strength and weaknesses of the robot

(Hagras, 2020, p.30). In the further development of Robat, there is a significant need for XAI models explaining how to interact with a potential teammate and get hold of the additional information through deep explanation (Hagras, 2020, p.31). It is also essential to know how Robat interprets the environment, what it makes decisions and what Robat are / are not able to provide in the partnership with humans.

References

Ackerman, E. Ollie the Baby Otter Is a Therapy Robot That's Actually Affordable. (March 25, 2015). IEEE Spectrum. Available through URL: https://spectrum.ieee.org/automaton/robotics/medical-robots/mit-ollie-the-baby-otter-therapy-robot.

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). Microsoft. Redmond, WA, USA. 2019. ACM. (https://www.microsoft.com/en-us/research/uploads/prod/2019/01/Guidelines-for-Human-AI-Interaction-camera-ready.pdf)

Batterbee, Ian. Don Norman's seven fundamental design principles. 2020. UX Collective. Available from URL: https://uxdesign.cc/ux-psychology-principles-seven-fundamental-design-principles-39c420a05f84.

Bratteteig, Tone. Verne, G. Does AI make PD obsolete? exploring challenges from Artificial Intelligence to Participatory design. Available from URL: https://dl.acm.org/doi/10.1145/3210604.3210646.

Chernofsky, J. Illustration of a robot and a human. Is artificial intelligence the future of customer service? (2016). The NextWeb. Available from URL: https://thenextweb.com/future-of-finance/2016/12/08/is-artificial-intelligence-the-future-of-customer-service/.

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).

Følstad, A. (2020). Interaction with AI – Module 2. Session 2. SINTEF. Avaliable through URL: https://www.uio.no/studier/emner/matnat/ifi/IN5480/h20/Undervisningsmateriale/interacting-with-ai-2020---module-2---session-2---handouts.pdf

Goodwin, Dr M. (2020). Interacting with Artificial Intelligence. University of Oslo. Available from URL: https://www.uio.no/studier/emner/matnat/ifi/IN5480/h20/Undervisningsmateriale/in5480-lecture-notes-13-oct-2020.pdf.

Grudin, Jonathan. 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.

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

High-level Expert Group on Artificial Intelligence. “A definition of AI: Main Capabilities and Disciplines” (April 8, 2019). European Commission. Available from URL: https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines.

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)

Lavars, N. Bat-inspired uses ultrasound to navigate unfamiliar environments. (September, 2018). New Atlas. Available through URL: https://newatlas.com/robat-bat-robot-echolocation/56242/.

Lawrence, C. How Sonar Tech and Echolocation Bring Smart Tech to Fishing. (October 10, 2018). IoT Zone. DZone. Available through URL: https://dzone.com/articles/how-sonar-tech-and-echolation-bring-smart-tech-to.

Liao, Q. V., Gruen, D., & Miller, S. (2020, April). Questioning the AI: Informing Design Practices for Explainable AI User Experiences. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (paper no. 463). ACM. (https://dl.acm.org/doi/abs/10.1145/3313831.3376590)

McCarthy, John. What is Artificial Intelligence? Computer Science Department. (November 12, 2007). Stanford University. Available from URL: http://www-formal.stanford.edu/jmc/whatisai.pdf.

Norman, D (1990). The problem of automation: Inappropriate feedback and interaction, not over-automation. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, Vol. 327, No. 1241, Human Factors in Hazardous Situations (Apr. 12, 1990). Available from URL: https://www.jstor.org/stable/55330?seq=9#metadata_info_tab_contents.

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

Sas.com. Analytics Software & Solutions. 2020. Available through URL https://www.sas.com/en_us/home.html.

Robots.ieee. What is a Robot? IEEE.org. Hentet 09.09.2020. Tilgejnglig via URL: *https://robots.ieee.org/learn/what-is-a-robot/*.

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. Available from URL:

/studier/emner/matnat/ifi/IN5480/h19/undervisningsmateriale/schulz-et-al_2018_classifying-human-and-robot-movement-at-home-and-implementing-robot-movement.pdf.

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

Simon, Matt. The WIRED Guide to Robots. Wired. 2020. Available from URL: https://www.wired.com/story/wired-guide-to-robots/.

Simonite, Tom. The Best Algorithms Struggle to Recognize Black Faces Equally. 22.07.2019. Wired. Available from URL: https://www.wired.com/story/best-algorithms-struggle-recognize-black-faces-equally.

Smith-Renner, A., Fan, R., Birchfield, M., Wu, T., Boyd-Graber, J., Weld, D.S., and Findlater. L. 2020. No Explainability without Accountability: An Empirical Study of Explanations and Feedback in Interactive ML. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–13. DOI: https://doi.org/10.1145/3313831.3376624

Sony. Entertainment Robot “abio” Announced. (November 1. 2017). SonyInfo. Available from URL: https://www.sony.net/SonyInfo/News/Press/201711/17-105E/.

Universal Design. What is Universal Design. Universal Design.ie. Available from URL: http://universaldesign.ie/What-is-Universal-Design/.

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)

Appendix: Feedbacks

Iteration 1

In the feedback of ‘Iteration 1’ I got a longer text which did explain what my fellow student did think about the text. She said that she thought it was a good text and that there were not any significant flaws in the text. Though, she did comment on one particular paragraph, the discussion of the different guidelines. She felt like I should discuss this further in the next delivery, which I really agree on. The last, but not least, comment she made was on the reference list and the style.

Iteration 2

The second feedback was very helpful as well. My fellow student commented that she did find the text interesting and easy to read. She did not have any comments on the actual text, but the format of it. She informed me that I should edit the article to fulfill the formal requirements, which I will do in Iteration 3. This was again related to the reference style and that the document was delivered through Google Docs and not PDF. Also, the references should be on a seperate page as for the rest of the text.