Contents

1.1 Concepts, definitions and history of interaction with AI

A brief history

Artificial intelligence emerged as an important topic during WWII, when code breaking using computers became an essential tool in the war (Grundin, 2009, p. 49). Alan Turing, a mathematician and code breaker at the time, suggested that there were many areas in which a computer could compete or even surpass human intellect. The actual term *artificial intelligence* was first used in relation to a workshop by John McCarthy in 1956.

After a small lull in attention towards AI, interest picked back up in the 1960s and 70s (Grundin, 2009, p. 50). In 1960, Herb Simon suggested that within twenty years, machines would be capable of doing any work that a man could do. Research and funding into AI increased considerably and multiple influential figures in the field showed their belief and support. However, there was also an element of fear surrounding the topic of AI. There was a worry that when computers reached a certain point of intelligence that they would become autonomous and possibly even dangerous.

Interest in AI came and went in waves, but Turing's idea that computers would soon rival human intelligence remained consistently relevant.

Definitions

> *"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, 2007, p. 2)

This is the definition given by John McCarthy in his article titled *"What is artificial intelligence?"*. This article is described to be for the everyday person and it is stated that the statements made are not a consensus among researchers of AI. This definition proposes AI as the process of developing intelligent machines and the process of using computers to understand human intelligence. Through this definition it seems McCarthy is trying to suggest that AI is not simply the task of computers trying to imitate humans, and that computers are not confined to the biological restraints that humans are.

> *"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".* (Bratteig & Verne, 2018, p. 1)

This definition is provided in the article *"Does AI make PD obsolete? Exploring challenges from Artificial Intelligence to Participatory Design"* by Bratteig and Verne. This is a more recent article that discusses AI in relation to participatory design (PD), where PD refers to the involvement of future users in a design process. This definition presents AI as the task of developing computer systems that mimic and imitate humans. Furthermore, there is a focus on the elements of human intelligence that can be particularly tricky to emulate, such as rational thinking or decision making within a certain context. This definition is therefore quite different from the one McCarthy presents.

> *"Artificial intelligence is that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment."* (Stone et. al., 2016)

This definition is presented in the article *"Artificial Intelligence and Life in 2030: One Hundred Year Study on Artificial Intelligence.".* Therefore, this definition is most likely influenced by a number of differing opinions and research, considering the article consists of thorough review of the field of AI. Here humans are not directly mentioned, instead "an entity" is used. Furthermore, intelligence is described as what makes a certain entity function, with foresight, in its environment, not demanding that this intelligence or functionality must be human.

> *"AI is the process of providing a non-living organism with intelligence, where intelligence refers to the ability to act and react efficiently in line with varying contexts, similar to the way living organisms operate"* (My definition)

It seems as though AI is generally thought of as the process of developing something, therefore my definitions begins with defining AI as a process. Furthermore, I appreciate the suggestion that AI doesn't necessarily have to be directly compared with human intelligence, and that is instead possible to identify the aspect of human intelligence that we are actually trying to emulate. This being the ability to act and react in different contexts and not in line with a set of rules, which is typically the way a computer operates.

Brief article review

I have chosen to read *Does AI make PD obsolete?; exploring challenges from Artificial Intelligence to Participatory design* by Bratteig & Verne (2018). This paper discusses the idea of AI affecting the use of PD in a negative way. AI is often considered to be able to understand a person’s needs possibly even before they have expressed their needs. PD is in many used to achieve precisely that, understand a user’s needs, therefore this paper discusses the idea that AI may make PD slightly irrelevant. Even though AI is often thought of as threat, I believe that it can never fully rival real human interactions. As is mentioned in the paper, AI is great if you a person that fits a statistically “average” person, but if you differ from this average then AI may not actually understand you that well.

Contemporary company that works with AI

A particularly relevant contemporary company that works with AI is Amazon. The way in which Amazon uses AI is not generally visible to the user, however the company does give a thorough description of their take on AI. AI is defined as *“the field of computer science dedicated to solving cognitive problems commonly associated with human intelligence, such as learning, problem solving, and pattern recognition.”* (Amazon Web Services, 2021). The company also connects AI to the term machine learning, and proposes that AI uses data as fuel, and becomes “smarter” with more data. A few use cases are presented as examples, including Amazon Alexa use of speech recognition and online shopping with personalized content recommendations. Furthermore, AI is offered as a service by Amazon, through their Amazon Web Service platform, where they aim to put machine learning in the hands of every developer (Amazon Web Services, 2021).

AI in fiction

The Netflix series *“Black Mirror”* introduces some sort of AI in almost every episode. The episode *“Be Right Back”* is particularly interesting. In this episode a young woman’s boyfriend dies in a car accident and she is presented with and accepts the opportunity to receive an identical copy of him with the use of AI technology. This AI version of her boyfriend is created through his previous messages, phone calls and videos of him. After living with this “fake” boyfriend for a while, the young women becomes increasingly uncomfortable with the technology. The AI version of her boyfriend does not have her real boyfriends negative attributes and is in many ways “too perfect”. This shows one of the ways in which AI may struggle to portray human intelligence, seeing as computers generally strive for perfection and humans are naturally imperfect. In addition, this example shows how when humans interact with AI, we very often find it extremely uncomfortable when something almost resembles a real human being, but at the same time is lacking important human characteristics.

1.2 Robots and AI systems

The history of the term “robot”

The word “robot” comes from the Czech term for “forced labour” (Simon, 2020). It was first used by Karel Capek in 1921 in his play named Rossums’s Universal Robots. From this the original idea of a robot was a piece of lifelike technology that would carry out tasks for you that you didn’t particularly want to do yourself.

Definitions

> *"A machine controlled by a computer that is used to perform jobs automatically"* (Cambridge English Dictionary, 2021)

This definition is somewhat short and non-descriptive. However, it does state that a robot should be some type of machinery that has been programmed, using a computer, to perform a task automatically. This implies that a robot should be able to carry out jobs without constant human intervention.

> *"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, 2018)

This definition builds somewhat on the former definition where it states that a robot is autonomous, meaning it is capable of carrying out tasks with some sort of freedom from human intervention. Furthermore, this definition claims that a robot should be able to sense it's own environment and act out tasks and computations in the real world. This implies that a robot can not purely exist in an somewhat abstract form, but that it must interact with it's surroundings in some sort of way.

> *"A robot is a machine that can to some degree act autonomously and that is aware of, and can act or react to, its environment"* (My definition)

My definition includes the aspect of a robot being somewhat autonomous, seeing as I find this to be an integral aspect of what we consider a robot. It should be able to make decisions without continuous input from a user or human. Furthermore, these decisions should be a result of an understanding of the robots physical environment.

The relation between AI and Robots

One of the main differences between robots and AI seems to be that fact that a robot needs to take the form of a machine that exists physically, AI does not need to take a tangible form. Furthermore, AI generally focuses on the ability of technology to emulate a humanlike intelligence, that is rational and can make complex decisions. The definitions of a robot we have looked at to not refer to this sort of high degree of intelligence. One thing that is similar for AI and robots it the idea of being able to act and react in a certain environment and therefore act somewhat autonomously.

A contemporary physical robot

One of the most advanced and humanlike robots that has been developed is "Sophia the robot" (Hanson Robotics, 2021). When it comes to Sophia's movements this robot is most humanlike when it comes to facial expressions. The physical movement of Sophia's limbs is more limited, being quite jittery and slow. This robot is very advanced when interacting with humans, being able to understand language and read facial expressions to carry out a substantial conversation. Sometimes Sophia is operating in a fully AI autonomous mode, and other times she more consistently controlled by a human, having more of a script and set objectives.

1.3 Universal Design and AI systems

Definition

> *"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."* (National Disability Authority, 2021)

This definition describes how a situation should be designed in a way that allows as many people as possible to thrive in that situation, regardless of their individual characteristics. Very often, new products or services will be developed in a way that causes the end product to be mainly designed for the average person. In this way, a large chunk of the population is often forgotten and the design is assumed to be "good enough". Universal design aims to ensure that this kind of discrimination does not happen and that no matter your situation, you will be designed for and not be forgotten.

The potential of AI

One example of AI that provides more inclusion of outlying groups of people is a technology that aids people who have lost the ability to move or speak to communicate with others (Willett et. al., 2020). This technology gains information from the motor cortex when the person is attempting to write a message, in this way the attempt is "decoded" and is translated to real text. This example shows how technology targeting human perception, cognition and movement can be used to help various individuals.

Another example of AI that contributes to inclusion is a technology that can aid the visually impaired in perceiving their surroundings (Grayson, 2020). This technology in particular consists of a device that is worn on an individuals head that senses which other individuals are in the vicinity and can give information about their location, identity and gaze-direction. The user can then be given this information in an auditory form. This technology can help visually impaired people be more confident and natural in their interactions.

However, AI doesn't exclusively provide inclusion for all groups of people, it can also lead to exclusion. An important example of this comes when considering the discrimination that has been uncovered to occur with the use of face recognition systems. Buolamwini & Gebru (2018) explain that face recognition algorithms are often claimed to be approximately 90% accurate, but that this accuracy is exclusive to a certain "kind" of person. In general, for an accuracy rate this high you would have to be a light skinned male, whereas being a dark skinned female would in some cases reduce this accuracy rate by 34%.

"Understand" and "understanding"

In my opinion “to understand” refers to the ability to receive information, process it and then act on that information. Machines are in many ways capable of this sentiment, where they can receive input, calculate the input and then give output according to their given algorithms and rules. However, understanding does not always follow specific rules and limitations, instead the way humans understand can be very complex and not always intuitive. Nuances in body language or context are things that machines are not very good at perceiving and this can limit their “understanding”.

1.4 Guidelines for Human-AI interaction

Microsoft Guideline 4: *“*Show contextually relevant information – *Display information relevant to the user’s current task and environment”*

This guideline refers to the importance of providing the user with information that suits the context they are in right at that moment. An example of this can be found on Netflix when you are watching the beginning of an episode of a TV show and Netflix will ask you if you would like to “Skip the intro”, or when Netflix asks you if you are still watching. Here the system is reacting to the specific context a user is in right at that moment.

I have chosen to compare Microsoft’s 18 guidelines for human-AI interaction with Norman’s (1988) Seven Principles of Usability. Being able to receive visual cues about what is happening in a situation and therefore also receiving feedback seems to be an important similarity between these two sets of guidelines. Furthermore, making it clear what a system is capable of doing is also a similarity. However, there are also some differences. For example, Microsoft’s guidelines mention trying to hinder biases and stereotypes being reinforced, something that Norman’s principles do not comment on.

2.1 Characteristics of AI-infused systems

AI-infused systems can be described as *systems that have features harnessing AI capabilities that are directly exposed to the user* (Amershi et al., 2019). Considering that AI-infused systems are constantly learning and change in line with the context they are in, they are often quite inconsistent and unpredictable. The dynamic nature of AI-infused systems leads to a large amount of mistakes and errors. Another key characteristic of AI-infused systems is that the expectations of the end-user are often inaccurate (Kocielnik, 2019). These inaccurate expectations are often a result of the fact that it is often not conveyed what these systems are actually capable of doing. Yang et al. (2020) also introduce inaccurate expectations as a key characteristic of AI-infused systems. Furthermore, this article also comments on that it is not only the end-users expectations that may be lacking, but also that of the designer. Designers may also struggle with considering all the different outcomes of an interaction with an AI-infused system, making the design process of these systems complicated.

An example of an AI-infused system is Tesla cars and their "full self-driving capabilities" (Tesla, 2021). This AI-infused system reflects many of the above mentioned characteristics. First and foremost, the self-driving capabilities that Tesla promises can be somewhat unpredictable. Seeing as this technology is constantly learning it can stumble upon some unfortunate errors, such as mistaking the moon for a traffic light (McFarland, 2021) or speeding up and crashing into the car in front (Levin, 2018). There are a considerable amount of accidents, some fatal, that have taken place while a Tesla vehicle was in autopilot mode (Othman, 2021). It seems that a main cause in some of these cases was the expectations the driver had of what a self-driving car could actually do. Many drivers assume that a self-driving car does not need a human to be present at the steering wheel, otherwise what would be the point of a self-driving car? This shows a clear lacking in the end-users expectations of what this AI-infused system can do.

2.2 Human-AI interaction design

In the article by Amershi et al. (2019) problematic characteristics of AI-infused systems are discussed. As mentioned above, this characteristics are for example the inconsistent and unpredictable nature of these systems. Throughout the article 18 guidelines for human-AI interaction are suggested, their intention being to avoid the negative outcomes that may occur while using AI-infused systems, for example end-user errors. These guidelines are categorized according to where in the user-journey they may be relevant; *initially*, *during interaction*, *when wrong* or *over time.* These guidelines cover aspects such as providing the user with sufficient information, being considerate of social norms and learning from the users behaviour.

Kocielnik et al. (2019) focus on the expectations that users have when it comes to AI. Here it is suggested that these expectations make it difficult for the user to accept an imperfect AI-system. As shown in the article by Amershi et al. (2019), providing the user with important information, for example about what the system can do, can be very beneficial in reducing errors and enhancing the user experience. Kocielnik et al. also suggest that giving the user enough relevant information about the AI will help in managing their expectations and making them more accepting of an AI system that is not perfect.

The first guideline I will consider in regards to Tesla's self-driving cars is G2: *Make clear how well the system can do what it can do* (Amershi et al., 2019). As mentioned previously, many people may overestimate the abilities of a self-driving vehicle. Automated cars use extremely advanced technology and are capable of incredible things, however they are still a relatively new innovation. Complex abilities, like how humans can understand many nuanced facets of a situation, are not something self-driving cars necessarily are so good at. An effort to communicate to the population just how well self-driving cars can actually drive by themselves would contribute to their ability to handle and interact with this kind of technology.

The second guideline I will discuss in the light of Tesla's self-driving cars is G6: *Mitigate social biases* (Amershi et al., 2019). This guideline consists of ensuring that the language and behaviours of the AI system do not reinforce unfair biases. It is difficult to gain a complete understanding of how well Tesla fulfils this guideline when it comes to their self-driving cars, however this guideline introduces a very important ethical conversation. In the event of an impending crash, who should the car choose to save? Most likely a self-driving car will be designed in a way that it will protect the people in the car, because who would buy a car that doesn't protect them. However, this means that individuals that perhaps cannot afford this degree of advanced technology will be in a more comprised situation, where their safety is intentionally not prioritized.

Bender et al. (2021) present various problematic aspects concerning large language models. The environmental and financial implications are discussed, where it is presented that the cost of language model training and development can be very substantial, both in dollars and in CO2 emissions. These risks are considered in comparison to the benefits of such large language models, where it is discussed how these models usually benefit a certain group of people, usually English speaking, and negatively affect other groups of people, for example environmental changes that impact more marginalized communities. Furthermore, the article reflects on how we may assume that if we were to take large amounts of data from the internet, we would have a dataset that is representative of all the worlds population. However, there is a certain type of person who typically has access to the internet, and also that voices their opinions on the internet. Therefore, assuming that the internet provides us with a representative dataset may enforce dominant viewpoints in the first world, thereby increasing the power imbalance and inequality. The article ultimately suggests a way forward to avoid the risks that are mentioned. In summary, it is recommended to spend a large amount of time and resources on the planning of data collection, and considering in detail what the risks and goals of the data collection are. Instead of just consuming huge amounts of data from convenient sources, more time should be spent on assembling datasets that actually suit the task at hand.

2.3 Chatbots / conversational user interfaces

Whereas in other design projects the visual layout and graphical aspects may be of great importance, it is the art of conversation that becomes crucial when dealing with chatbots. Chatbots are unique in this aspect and therefore present complex challenges. First and foremost, a chatbot must be able to understand and respond to input in an appropriate manner, an aspect that Følstad & Brandtzæg (2017) present as a key challenge. This input can be presented in an infinite amount of different ways, where the user doesn't just have the option of clicking a certain button, but can express themselves in a unique way using language. Another key challenge is the fact that we must shift our focus from designing an object to designing a service.

Adherence to the guideline G1: *Make clear what the system can do* could help with the aspect of what input the user might provide. If the user is aware of what a certain chatbot is there to help with, then the input they give may be more suited to the responses the chatbot is capable of giving. This also includes making it clear what the system can't do and giving feedback when the chatbot doesn't understand, and what to give as correct input. A way for chatbots to adhere to this guideline may be to present to the user what kind of input is preferred, for example the use of a few key words instead of complete sentences.

When it comes to the guideline G2: *Make clear how well the system can do what it can do* could be very influential when it comes to the interaction between human users and chatbots. Often, chatbots are designed to resemble humans and human conversation as much as possible, this may give users the expectation that this technology with understand and respond in the same way a human will. If the limitations of a particular chatbot were made more clear, the expectations of users may also be managed. A way for chatbots to adhere to this guideline may be to avoid resembling very human-like system. For many users the human aspect of chatbots is often what makes them feel comfortable and confident while using them. Therefore, there is an important balance to be considered here.

3.1 Feedback

The "wish" I received from the first iteration was directed towards the "AI in fiction" section of the assignment. Here it was suggested that I explain a little more in thoroughly how the example I introduce portrays humans interaction with AI, seeing as I mainly just explained the specific details about the example. Therefore, I have now gone a little more in depth about how I feel this example of AI in fiction shows how humans and AI may interact.

The "wish" I received from the second iteration was directed towards the "Chatbots/conversational user faces" section of the assignment. Here the feedback proposed that I give some suggestions on how chatbots could fulfil the two guidelines that are presented, not just explain why it is important that they follow them. Therefore I have provided some suggestions on how chatbots could follow these guidelines.

Sources:

Amazon Web Services. (2021). *What is Artificial Intelligence? https://aws.amazon.com/machine-learning/what-is-ai/*

Amazon Web Services. (2021). *Machine Learning on AWS.* https://aws.amazon.com/machine-learning/

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

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

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

Grayson, M., Thieme, A., Marques, R., Massiceti, D., Cutrell, E., & Morrison, C. (2020). *A dynamic AI system for extending the capabilities of blind people*. Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems.

Grudin, J., (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.

Hanson Robotics. (2021). *Sophia's Artificial Intelligence.* https://www.hansonrobotics.com/sophia/

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)

Levin, S., (2018) *Tesla fatal crash: 'autopilot' mode sped up car before driver killed, report finds.* The Guardian. https://www.theguardian.com/technology/2018/jun/07/tesla-fatal-crash-silicon-valley-autopilot-mode-report

McCarthy, J., (2007) *What is Artificial Intelligence? Computer Science Department*. Stanford University. http://www-formal.stanford.edu/jmc/whatisai.pdf.

McFarland, M., (2021) *How Tesla can sell "full self-driving" software that doesn't really drive itself.* CNN. https://edition.cnn.com/2021/10/09/cars/tesla-fsd-legal/index.html

National Disability Authority, (2021), *What is Universal Design.* http://universaldesign.ie/what-is-universal-design/

Norman, D., (1988), *The Design of Everyday Things,* New York: Doubleday, 1990.

Othman, K., (2021) *Public acceptance and perception of autonomous vehicles: a comprehensive review.* AI Ethics 1, 355-387. https://doi.org/10.1007/s43681-021-00041-8

Robots.ieee., (2021) *What is a Robot?* IEEE.org. ieee.org/learn/what-is-a-robot/

Simon, M., (2020) *The WIRED Guide to Robots*. Wired. https://www.wired.com/story/wired-guide-to-robots/

Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., Hirschberg, J., Kalyanakrishnan, S., Kamar, E., Kraus, S., Leyton-Brown, K., Parkes, D., Press, W., Saxenian, A., Shah, J., Tambe, M., & Teller, A. (2016). *Artificial Intelligence and Life in 2030: One Hundred Year Study on Artificial Intelligence.* Stanford University. http://ai100.stanford.edu/2016-report

Tesla, (2021), *Autopilot.* https://www.tesla.com/autopilot

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

Willett, F.R., Avansino, D.T., Hochberg, L.R. *et al.,* (2021), *High-performance brain-to-text communication via handwriting*. https://doi.org/10.1038/s41586-021-03506-2

Yang, Q., Steinfeld, A., Rosé, C., & Zimmerman, J. (2020). *Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design.* In Proceedings of the 2020 CHI conference on human factors in computing systems (Paper no. 164). (https://dl.acm.org/doi/abs/10.1145/3313831.3376301)