Note: I have not been able to attend lectures due to quarantine. For this assignment, I have only used information that was possible to get online.

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

First, write a section about how AI came about, the history of AI. When, and by whom, was the term first used?

Conferences in the late 1940s and early 1950s discussed topics such as cybernetics and neural network models (Grudin, 2009). Among the participants in these conferences was Norbert Wiener who defined these topics as the study of control and communication in the animal and the machine (Grudin, 2009). The term 'artificial intelligence' was later coined by John McCarthy in 1956 (Grudin, 2009). This he claims, was to escape the association with cybernetics since according to John Mcarthy "It's [cybernetics] concentration on analog feedback seemed misguided, and I wished to avoid having either to accept Norbert Wiener as a guru or having to argue with him" (McCarthy, 1996). In the mid-1960s to the mid-1970s AI became a major research field, gaining monetary support from both the US and UK governments (Grudin, 2009). Due to a negative report on the prospect of AI in 1973, the UK's government funding almost went away completely (Grudin, 2009). This was also the beginning of what was called the 'AI winter' when AI ventures were scaled back sharply (Grudin, 2009). Not until the foreign threat of Japan's AI research that ushered in the 5th generation of AI, did the US funding of AI research increase again (Grudin, 2009). During this time new terminology was being used such as "expert systems," "knowledge engineering," "machine learning" and the term "AI" was used sparingly (Grudin, 2009). In the 1990s another AI winter was coming, with 2 billion spent on research, and with no tangible results, funding was cut once again (Grudin, 2009). However today this has changed. Much due to the reduced cost of storage and increasing capability of semiconductors, machine learning, and other AI technologies can be used on standard PCs (Grudin, 2009).

Then, find three different definitions of AI. Describe and explain these three definitions, for example by when it was defined, by whom and in what community. Based on these three definitions, make one definition yourself - and describe and explain your definition.

Artificial general intelligence
Artificial general intelligence (AGI) is a hypothetical computer program that can perform tasks just as well as a human or better ("DeepMind and Google," 2019). AGI also called "strong AI" differentiates itself from weak AI (Siri, Cortana, IBM watson) by being able to use reason, represent knowledge, plan, learn, communicate in natural languages (Adams et al., 2012). AGI was first coined by Mark Gubrud in 1997. In his paper, Gubrud brings up his point that AGI can be used in essentially any phase of industrial or military operations where human intelligence would otherwise be needed (Gubrud, 1997).

Weak artificial intelligence
Weak AI, also called Narrow AI are systems that are good at performing a single task or a limited range of tasks (Dickson, 2020). These systems often outperform humans in their specified domains but fail as soon as they are presented with a situation outside of their field (Dickson, 2020). Weak AI is not conscious or driven by emotions like humans are (Jajal, 2020). Examples of weak AI are Siri, Cortana, or IBM Watson.

Machine learning
Machine learning or ML for short is computer algorithms that improve automatically through experience from data (Goodfellow et al., 2016). Machine learning was defined by Arthur Samuel in 1959 (Samuel, 1959). Samuel wanted to prove that a computer can be programmed so that it will learn to play a game of checkers better than a person. ML can be divided into two categories, supervised learning and unsupervised learning (Goodfellow et al., 2016). Supervised ML takes data that already has the correct answer like an image labeled cat or not a cat and tries to predict the next image based on that data (*Supervised Machine Learning*, n.d.). It's supervised because we tell the algorithms what it got wrong. Unsupervised ML is used when there is no cat label and we do not know the label beforehand (Garvie & Frankle, n.d.). The goal is to find similarities and create labels.

Find one contemporary company that work with AI and describe how this company present AI on their web pages. In what way does this company talk about AI, as a product, as a service, framework or "idea"?

IBMs AI has had two successful publicity stunts, first, it beat the world's top chess player, Gary Kasparov in 1997, and in recent years it won a game of "Jeopardy!". On the IBM website, they present their AI called "Watson" as an AI service for enterprises. Watson will help you "unlock the value of your data in entirely new ways." IBM has two featured products, the IBM Watson Assistant and IBM Watson Discovery. Their Assistant is a conversational AI platform that provides fast and accurate answers to questions. It can automate the customer experience and derive insights from data. IBM Watson Discovery however is an AI-powered search technology, it uses text analytics and natural language processing to retrieve specific answers to questions. It understands business documents and identifies hidden trends within data.

I do find it interesting that IBM has little to no information about their ethics involving their AI, this was usually very prominent when visiting other companies' websites where they were promoting AI.

Select one documentary or a fictional film, book or game that is about the use and interaction with AI. Describe with your own word how human interaction with AI is portrayed in this work.

In his "Talks at Google," John Searle discusses consciousness in artificial intelligence. In the video, he brings up his famous Chinese room thought experiment. In this thought experiment, a computer takes Chinese characters as input and by following instructions of a computer program, it produces other Chinese characters as output. Suppose this convinces a human Chinese speaker that the computer is a human since all the questions that are put into the computer get a correct answer and thus it's able to pass the Turing test. But Searle makes the argument that the computer does not understand Chinese, it merely simulates understanding.

“Implemented computer programs by itself is never going to be sufficient for human understanding because human understanding has more than syntax. It has semantics.” Humans understand words and characters but also what they stand for in a real context.

Searle also brings up two senses in cognitive science that is observer-relative and observer-independent. Examples of observer-independent are mountains and molecules because they exist regardless of what anyone thinks. Observer-relative is stuff that only exists relative to the observers and users. Here Searle brings up the example of money as a piece of paper, but since we collectively have a special attitude toward it, it becomes money. Observer relative is created by human consciousness however the human that created this is itself not observer relative. We all know what consciousness is because we are all conscious beings but to define exactly what it is or reproduce it in a computer program is something that we do not know how to do. At the end of the talk, Searle explains his view that we first need to figure out how the brain produces consciousness before we can try to simulate consciousness.

1.2 Robots and AI systems

First, write a section about how the word Robot came about.

The word robot was introduced by the Czech playwright, novelist, and journalist Karel Capek in his drama show “Rossum’s Universal Robots” (RUR) in 1920 (*Science Diction*, n.d.). In Slavic, the word rabota means servitude of forced labor. In RUR the robots perform all the work that humans don’t want to do. This resulted in the mass production of robots since the people got comfortable with the robots doing all the hard work. At the end of the play, the robots rise up against the humans, killing everyone except one. They then realized that they had killed everyone who knows how to create more robots, resulting in their own demise.

Then, find two different definitions of “robot”. Describe and explain these definitions. Based on these definitions, make one definition yourself, and describe and explain this definition.

Merriam-Webster defines a robot as 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).

The acclaimed source of bestsciencefictionbooks.com defines the word robot as an artificial device or being, that is mechanical (*BestScienceFictionBooks.Com*, n.d.). In science fiction, a robot expresses one of three general feelings: pro-robot, anti-robot, and ambivalence (*BestScienceFictionBooks.Com*, n.d.). The plot of an anti-robot story features robots who follow hard wired code and tend to be benevolent servants, while an anti-robot story features a conflict or confrontation (*BestScienceFictionBooks.Com*, n.d.). Ambivalent stories explore the usefulness of robots, but also a sense of anxiety about human-like mechanical intelligence (*BestScienceFictionBooks.Com*, n.d.).

Discuss the relation between AI and Robots. Is "a robot" different from "an AI"? In what ways are they different and similar? Bring in the definitions that you described earlier about robots and AI for this discussion.

An AI 'learns' from its environment, this is done with computer algorithms that improve through the use of data. Over a time period (with more data), an AI will get better and more effective at completing whatever the task is. A robot does not operate like this,

Find one contemporary physical robot, either described in a research article - or a commercial robot, and describe how this robot moves and how a human user is interacting and using the robot in a specific situation.

In 2014 the company Savioke introduced their butler robot ALO (*Your Robot Butler Has Arrived*, n.d.). When released it was operating in a hospitality setting, the guests in a hotel could request a toothbrush or a charger from their room. Staff at the hotel would drop the item in a compartment on ALO and the robot makes its way to the guest, traveling independently between floors with the elevator and remotely rings the bell on a guest's door. ALO is using chimes and a display to interact and communicate with guests. Savioke claims that ALO provides two sorts of value, the first is that with the help of a robot, staff can focus on more important things than delivering toothbrushes (*Your Robot Butler Has Arrived*, n.d.). The second is that they believe ALO will delight guests in a way that travelers will make a point of visiting hotels that are using the robot, improving the guest's experience.

1.3 Universal Design and AI systems

Please find and describe a definition of Universal Design. Explain this definition, how you understand what Universal Design is about with respect to inclusion.

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 (*Centre for Excellence in Universal Design*, 2020). For me, universal design is a way to verify no one is left behind and making design accessible. There is a risk designers forget the user and base their design on their own perception of the world. Universal design exists to bring light to the marginalized communities and people. There is also a bottom-line incitement for companies to adhere to universal design since you are able to sell the product to a broader group of people.

Describe the potential of AI with respect to human perception, human movement and human cognition/emotions. You are encouraged to use examples.

Describe the potential of AI for including and excluding people. You are encouraged to use examples.

One large controversy for AI and specifically facial-recognition AI is racial bias (Garvie & Frankle, n.d.). If the person in a photo is a white man the facial recognition is right in 99 percent of the cases, however the darker the skin, the more errors arise (Lohr, n.d.). This fault in the AI is mostly due to a limited database with 75 percent male and 80 percent people with white skin (Lohr, n.d.). But what will happen when the AI tries to recognize gender? The LGBTQ+ community has previously been a marginalized group and still is today in many facets of society. Can we ever trust the AI to correctly

recognize gender? And in what situation is it actually relevant? I believe with more and more investments going toward commercial facial recognition technology, these questions will become essential to answer.

In WCAG 2.1 principles and in the Human AI-Interaction guidelines the concept "understand" and "understanding" is used. Explain briefly in what way you make sense of the concept "understand" and "understanding".

The user should be able to tell what the effect of using a function or element is and the language should clearly represent the intended effect.

Then address the question: Do machines understand?

I will agree with John Searle on this question, no machines do not understand, at least not in the way humans understand. Humans understand the semantics and the context behind words and characters. This is not something machines are able to do.

1.4 Guideline for Human-AI interaction

Please select one of the 18 guidelines from Microsoft, and describe this guideline with a different example than what is given by Microsoft.

G13 - Learn from user behavior. Personalize the user's experience by learning from their actions of time (Amershi et al., 2019). When I use search on my iPhone it first suggests four apps that I use often, however this recommendation is also based on location and time. In the morning it knows I usually check the weather at 9:00 so it suggests my weather application but later in the day my iPhone knows I'm not interested in weather so recommend me other applications that are more relevant to me at that time.

Search, and find one set of HCI design guidelines. Discuss briefly similarities and differences between the HCI design guidelines and the Human-AI interaction guidelines.

In Nielsen's 10 heuristics for user interface design 1# is visibility of system status (Nielsen, 1994). It says that the system should always keep users informed about what is going on, through appropriate feedback within reasonable time (Nielsen, 1994). G1 for Human- AI interaction says, make clear what the system can do. Help the user understand what the AI system is capable of doing (Amershi et al., 2019). G2 says make clear how well the system can do what it can do, help the user understand how often the AI system may make mistakes (Amershi et al., 2019). There is a strong overlap with these two guidelines, where the commonalities are about making the user understand what is going on.

References

Adams, S., Arel, I., Bach, J., Coop, R., Furlan, R., Goertzel, B., Hall, J. S., Samsonovich, A., Scheutz, M., Schlesinger, M., Shapiro, S. C., & Sowa, J. (2012). Mapping the Landscape of Human-Level Artificial General Intelligence. *AI Magazine*, *33*(1), 25–42. https://doi.org/10.1609/aimag.v33i1.2322

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for Human-AI Interaction. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3290605.3300233

DeepMind and Google: The battle to control artificial intelligence. (2019, March 1). *The Economist*. https://www.economist.com/1843/2019/03/01/deepmind-and-google-the-battle-to-control-artificial-intelligence

Dickson, B. (2020, April 9). *What is artificial narrow intelligence (Narrow AI)?* TechTalks. https://bdtechtalks.com/2020/04/09/what-is-narrow-artificial-intelligence-ani/

Garvie, C., & Frankle, J. (n.d.). *Facial-Recognition Software Might Have a Racial Bias Problem*. 6.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

Grudin, J. (2009). AI and HCI: Two Fields Divided by a Common Focus. *AI Magazine*, *30*(4), 48–48. https://doi.org/10.1609/aimag.v30i4.2271

Gubrud, M. A. (1997). *Nanotechnology and international security*. *1*.

Jajal, T. D. (2020, February 13). *Distinguishing between Narrow AI, General AI and Super AI*. Medium. https://medium.com/mapping-out-2050/distinguishing-between-narrow-ai-general-ai-and-super-ai-a4bc44172e22

Lohr, S. (n.d.). *Facial Recognition Is Accurate, if You're a White Guy*. 5.

McCarthy, J. (1996). *Defending AI Research: A Collection of Essays and Reviews*. CSLI Publications.

Nielsen, J. (1994). Enhancing the Explanatory Power of Usability Heuristics. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 152–158. https://doi.org/10.1145/191666.191729

Robot Science Fiction | BestScienceFictionBooks.com. (n.d.). Retrieved September 13, 2020, from https://bestsciencefictionbooks.com/robot-science-fiction.php

Samuel, A. L. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, *3*(3), 210–229. https://doi.org/10.1147/rd.33.0210

Science Diction: The Origin Of The Word "Robot." (n.d.). NPR.Org. Retrieved September 7, 2020, from

https://www.npr.org/2011/04/22/135634400/science-diction-the-origin-of-the-word-robot

Supervised Machine Learning: Crash Course Statistics #36. (n.d.). Retrieved September 7, 2020, from https://www.youtube.com/watch?v=jmLid2x9eKg&ab_channel=CrashCourse

What is Universal Design | Centre for Excellence in Universal Design. (n.d.). Retrieved September 8, 2020, from http://universaldesign.ie/What-is-Universal-Design/

Your Robot Butler Has Arrived. (n.d.). Savioke. Retrieved September 8, 2020, from https://www.savioke.com/blog/2014/8/11/your-robot-butler-has-arrived