IN5480 individual assignment

Table of Contents

Introduction

This report is divided into three parts. Module one discusses concepts, definition and history of Artificial intelligence, robots and universal design. It also contains some examples of real-life and media to explore these technologies. Module two discusses AI-infused systems with focus on chatbots and how we need to manage the exaptation's to the user. Finally, module 3 explores the level of autonomy in robots and side-effects what may occur if we increase or decrease autonomy in our robots.

Module 1

Early history of AI

The concept of AI is nearly as old as the modern computer. Already in the late 40s, computer science pioneers like Alan Turing was public talking about computers eventually have intellect matching humans.

And in the early 50s the fields of cybernetics, defined by Norbert Weiner as "the study of control and communication in the animal and machine" (Jonathan Grudin, 2009). Was discussed in New York on a series of conferences. This; together with dissuasions about neural networks models became the basis for mathematician and logician John McCarty to Coin the term artificial intelligence in 1956. Today artificial intelligence, machine learning, deep learning and neural networks are often used interchangeably even though they often contain very similar properties. They simulate a kind of intelligence or a way to learn new information.

The development of AI has had many ups and downs during the 20th century. A mix of founding, interests of researchers and technical limitations has made booms and droughts. One of the first big booms for AI came in the 60s. With the cold war and the space race kicking off, both the US and UK founded a lot of scientific research, and AI was one of them.

Researchers and scholars were optimistic and in 1965 I.J Good proclaimed: “More probable than not, within the twentieth century, an ultraintelligent machine will be built and it will be the last invention that man need to make.” (Jonathan Grudin, 2009). Marvin Minsky and several other researchers agreed that a machine with human level intelligence would arrive by 1985. This did not happen, but we can easily see the influence their statements have had on media and popular culture.

Some definitions of AI

> “*[…] a machine with the general intelligence of an average human being. I mean a machine that will be able to read Shakespeare, grease a car, play office politics, tell a joke, have a fight. At that point the machine will begin to educate itself with fantastic speed. […]”* - Marvin Minsky (Jonathan Grudin, 2009 / Time Magazine, November 20, 1970)

Minsky’s quote fit well in the optimism around technology and science present in the 70s. The seed of innovation since the world wars astonishing, and many believed that no technology was to far fetched to be invented in just a few decades. As i interpret Minsky, he talks about an AI that behave and act like humans. Has needs for emotional stimulation such as being capable to be angry, have a fight, the joys of a joke and play office politics.

> *“AI is a subfield of computer science aimed at specifying and making computer systems that mimic human intelligence or express rational behaviour, in the sense that the task would require intelligence if executed by a human.”* (Bratteteig and Verne, 2018)

As a contrast to Minsky, Bratteteig and Verne have a view of AI from a more modern viewpoint. They don't talk as much about the AI having real human capabilities. But talk more about AI as

an imitation of human behaviour. If we want to put an unsiencefitic label on it: Minsky talks about AI like it has a soul. Bratteteig and Verne does not.

> "*The capacity of a computer to perform operations analogous to learning and decision making in humans, as by an expert system, a program for CAD or CAM, or a program for the perception and recognition of shapes in computer vision systems*". (dictionary.com, 2019)

This last definition of AI is very generic, as it's from a dictionary. But this last one is more about how a system can take decisions like a human, without being one.

I think AI is about how a computer can acquire new skills, or make decisions without the need for human input. And the possibility to adapt its current knowledge to new uses and use these skills in a human like behaviour.

Siri by Apple, an example of AI today

Siri, by Apple, is maybe one of the most famous AI "assistants" in the world. And as many other products made by Apple, they try to present it as some technological magic. They present Siri as a convenience you cannot live without. Siri's slogan at the moment is: "Siri does more than ever. Even before you ask. "In many ways, they don't talk about the tech behind Siri, they don't even mention AI on its product page. Apple only focus on the many features, how easy the product is to use, and the safety of the product.

How AI portrayed in Westworld

In the TV series Westworld is about a theme park made to look like an old western movie. The park is filled with robots, who all have AI personalities who play a role in the park's narrative. In the early episodes of the series, we see how the humans in the park behave around the robots.

Many, if not most, use their time in the park to live out a western fantasy, without any consequences. They shoot, kills and rape as many robots as they can in the park. Just because they can. It's a sinister viewpoint; to infer that most humans turn into power hungry psychopaths just because they can't get punished for their actions.

Later in the series the show explores the concept of sentience, and if AIs can develop free will. If you look at the interactions between robots and humans, you can clearly see it changes over time. The humans treat them more as a fellow man after the inevitable uprising.

How the word robots came about

The Czech word for "forced labour" and was first used in the play: R.U.R by Karel Capek. Later, the word have evolved and some definitions today are:

"A machine resembling a human being and able to replicate certain human movements and functions automatically." - lexico.com, 2019

"A robot is a physically embodied artificially intelligent agent that can take actions that have effects on the physical world"- Anca Dragan, roboticist of UC Berkeley

The relation between AI and Robots

A robot is a kind of machine who allow a computer to move in a 3-dimensional space. This may allow the computer to do not only calculation tasks, but also physical tasks. A robot is is the embodiment of a computer. If an AI is the "brain" on a machine a robot is the "body".

Maybe the most contemporary robots in society today, is the robot cleaners. Many have robots vacuums, and they have sensors that make them capable to navigate rooms and open spaces. Different models have different strategies on how they move, but most modern models uses AI or other algorithms to åathfind an optimal path for cleaning. Usually humans don't interact a

lot with vacuum cleansers, mostly because they are designed to run when the owner isn't home.

Universal design

Universal design can be defined as the design of products and environments to be usable to the greatest extent possible by people of all ages and abilities. (Story et al., 1998) This definition fits my view of universal design well. Universal design is about including as many as possible, and be conscious of the restrictions and exclusions of groups inherit in the design of a product or environment.

It's possible to use AI to analyse human emotions and reactions. In the medical field, i may be used to help patients with expression impairment, like cerebral parse to express themselves. We can use AI to reduce the miscommunication between patient and caretaker.

In many ways, AI today are exclusive. Many face recognition software's have an hard time to recognise non-white people. And voice-controlled systems are often only usable if you have a good English accent. On the other side; if AI are design the right way, we may manage to remove a lot of biases and executions made by humans today. Like the possibility to make websites more accessible to all, or AIs who make decisions without individual biases, like humans.

Module 2

Characteristics of AI-infused systems

Most AI infused systems are "*Systems that have features harnessing AI capabilities that are directly exposed to the end user*".(Amershi et al., 2019). Typical examples of this can be music

recommendations, email filtering and voice assistants. We can usually differentiate “normal” algorithms and AI infused system on a few key characteristics that I will describe in this section.

Today all AI infused systems really on large datasets data and the emerging capability to possess. The data used by the AI-infused system can differ hugely, but often they try to make an make a profile out of age, sex, location and previous behaviour, where the latter is the most important.

Out of the large datasets collected of the users the systems learn and try to predict a certain outcome. Learning is the second characteristic. In a mail application, this may be to characterize emails based on what you and other users have characterized similar emails as before. A sub characteristic of this is that the system should always try to improve itself and correct errors in its behaviour. And errors is inevitable.

Because the system is learning, change over time, and are using huge and complicated datasets, the behaviour may become uncreditable. An example of this is that you may experience a different google response if try the same queries on different machines. This may be because the system doesn’t know what kind or answers to be tailored to you, or you may get answers tailored to someone else.

Example of a AI-infused system

Google’s Gmail application is a AI-infused system on multiple levels. One application Gmail uses AI, is to automatically sort mail into 3 main categories. Primary, Social and Promotions. Without knowing the details of the black box that do this sorting for me, I can assume it analyses the text and meta data.

Just after this feature came out to the public, I had some problems with it. Emails from Facebook always got placed in the social category, this was a problem because not all mails from Facebook are equal, some are more important, and I want them in my primary inbox. But after some time, emails containing important information like security warnings or emails

regarding my commercial Facebook pages got placed in the primary category. This is a good example of the AI learning what I usually read, and changes accordantly.

Human-AI interaction design

Amershi et al. (2019) article is mainly about general guidelines on how to design for understandability and control in AI-infused systems. We can't always really on old design precipices and understanding when working with AI, because of the complex underlying mechanisms of AI. The 18 guidelines presented in the article helps designers to convey the underlying functionality and inform the users about and why actions are made.

The main take-away in Koscielny's article is the importance of adjusting user expectations in a AI-infused system. If the user understands the limitations of an AI system, their acceptance of the system will likely be higher. Koscielny also writes about the cost of error recovery, and that a system with few false negatives have an increased perceived accuracy for the end user then systems with few false positives. (Kocielnik et al., 2019)

One way Gmail follows Amershi's guideline G17 (Provide global controls) is how the user can turn on and off certain AI-infused functionality. In Gmail, you can off the categorisation from the general settings, or if you want, turn on more categories then are enabled by default. You can also manually move email between categories. In some ways Gmail do not comply completely, because you can't tweak how the categorisation works on a general level, but it still gives a lot of control to end-users.

I will say that Gmail do not follow G1 (Make clear what the system can do) particularly well. I find no explanation on what and how the system are sorting my emails. I can deduct from category titles and email content what the broad functionality are, but this is never explained or confirmed in the system.

If Gmail docs had given the user a little more control over the categorisation and had been clearer on how it categorise its content, I may have had a better experience in in how I perceive the functionality. And this may especially be true for new users.

Chatbots / conversational user interfaces

Luger and Sellen describes a conversation agent (CA) as "dialog system often endowed with humanlike behaviour" (Luger and Sellen, 2016), and mention one challenge with CA is that a lot of research today mostly focus on the technical perspective or use of CA in controlled environment or for a specific context. The cap between expectations and capability of the systems in an everyday context is one of the key challenges we face today in the design of conversational agents. This cap may make users hesitant to use CA for critical or advanced tasks, even though they are capable of these kinds of tasks.

One of the challenges Følstad og Brandzæg highlights are the biases included in the design of conversational agents and other AI systems. CAs are primally made by a western, mail dominant tech industry. (Følstad and Brandtzæg, 2017) For a AI infused system, the dataset is everything, and if the dataset includes biases, the resulting system will as well.

If we design conversation agent agents with Amershi's guideline G1 (Make clear what the system can do) and G2 (Make clear how well the system can do what it can) in mind, we can hopefully lessen the cap between expectations and capability. The trust between machine and human are central do overcome this challenge. By have a deeper understanding on how the underlying system work, we may also see, find and remove some of the biases mentioned by Følstad and Brandzæg. Users will know it's a system fault, and not a user fault.

Modul 3

In this module I will use some of the examples of human-robot-collaboration from *Human-animal Teams As an Analog for Future Human-robot Teams* (Phillips et al., 2016) and describe and assess which level of autonomy they have according to the scale provided in *Designing for Situation Awareness* (Endsley, 2011). I will also reflect on some advantages and disadvantages that may occur if we increase or decrease the level of autonomy.

In Endsleys book, er can find 12 levels of autonomy described. The lowest one is *Manual control* which is when humans preform all aspect of the a given task. At the other end is *full automation*, which is when a robot or AI can perform all aspects of a task without any human intervention or input.

Industrial robots

On level five, *batch processing,* we find my first case: industrial robots. Most industrial robots today I would say fit on this level. A computer carries out a set of task or actions predefined by a human. For example, robots in a car manufacturing plants mount, assemble, weld and lift standardised parts to make a standardised car. The robots itself do not make any meaningful decisions or even decide the order to out the parts. All of these kinds of decisions are pre-programmed by a human. Often, the only way the robot can influence its actions, is by making small adjustment based on its sensors. Often, just to make sure the parts are aligned. To summarise, the robot only real responsibility is to implement the tasks set out by humans. The benefits of industrial robots today, is that they can remove a lot of physical stress on factoryworkers and mechanics by reducing the amount of lifting required and to mount components on hard to reach locations. This is what Phillips calls *Multiply physical capabillites,* and in her arcicle industial robots are compared to elephants lifting trees as a exsample of this.

If we lower the level of autonomy to industrial robots, many of the benefits by using them will disappear. If the robot only can to one task, like on autonomy level 4 *Action support*, and not a series of task without human instructions, production time will surly increase. And if lower it even further, to level 3 *Situational awareness support*, it would just be a lift with measuring instruments controlled by a human. In a car factory setting this would not be very benefitable,

but in a non-standardised setting, it could. Like in a car repair shop. The repairs, and cars, may not be totally standardised, but could still benefit from a robot to lift an engine and to assist the mechanic to place engine in the right place or orientation in relation to the rest of the car.

I do not believe an increased level of autonomy would be the best solution in a car factory setting I have descried. In many ways. Before the car manufacturing robots, we have today, we used human labour for the same tasks. The humans back in the earlier days did not have a lot of input in the car manufacturing. They performed a set of tasks defined by their possession on the assembling line. Usually a standardised car has a standardised order to be manufactured.

Search and rescue

Today, we have drones that are used in many different applications. One of these applications is to extend or augment our cognitive capabilities. In high risk environments, like a collapsed building. There are many dangers working in or near a collapsed building, like asbestos risk, the risk of further collapse and the general danger of walking on debry. Today we can use drones to find wounded people before sending in a rescue team. We can also use the drones to map out, or spot dangerous parts of a site that need to be avoided. The drones can fly by themselves and use infrared cameras to find and highlight wounded for the requeue workers.

The scenarios described above, is all under the what I would classify as level two autonomy: *information queuing.* The drones are controlled by humans and apply a viewpoint not possible without them. The drones can scout and report back findings without human interference, except from initial setup and defending area to search.

The only level we can lower the autonomy to is level one: *Manual control*. In this scenario this would mostly bring us to the level as hobby drones or rescue drones. The rescue workers may still get the benefits of an overview over the search site, but humans will be behind the controls. One disadvantage may be that the searching requires more specialised drones operators, and this may drain resources form other part of the operation, or limit the number of drones used because of lacking manpower.

One advantage of using manual control is that humans may notice ques or patterns that may be out of the capabilities of the drones. If the humans rely to much on the highlights form the level 2 system, important information can be missed and crate bad situations.

If we increase the level of autonomy of rescue drones, to for example level 4: *Task support*. We may tell the drones to autonomous fetch people in need. A human operator would oversee the rescue site and tell the drone rescue the humans found. If we can achieve this level of autonomy, we can minimize the risk of humans, and can rescue humans in situations that would be to risky for humans.

Appendix

According to the feedback received from iteration one and two I have added introduction to my report, and done some spellchecking. I also added a table of content.

References

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P., Inkpen, K., Teevan, J., Kikin-Gil, R., Horvitz, E., 2019. Guidelines for Human-AI Interaction, in: CHI 2019. ACM.

Bratteteig, T., Verne, G., 2018. Does AI Make PD Obsolete?: Exploring Challenges from Artificial Intelligence to Participatory Design, in: Proceedings of the 15th Participatory Design Conference: Short Papers, Situated Actions, Workshops and Tutorial - Volume 2, PDC '18. ACM, New York, NY, USA, pp. 8:1–8:5. https://doi.org/10.1145/3210604.3210646

Dictionary: https://www.dictionary.com/browse/artificial-intelligence Accessed 27 Sept. 2019.

Endsley, M.R., 2011. Designing for Situation Awareness: An Approach to User-Centered Design, Second Edition, 2nd ed. CRC Press, Inc., Boca Raton, FL, USA.(Only chapter 2 and 10)

Følstad, A., Brandtzæg, P.B., 2017. Chatbots and the new world of HCI. interactions 24, 38–42. https://doi.org/10.1145/3085558

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 - CHI '19. Presented at the the 2019 CHI Conference, ACM Press, Glasgow, Scotland Uk, pp. 1–14. https://doi.org/10.1145/3290605.3300641

Luger, E., Sellen, A., 2016. "Like Having a Really bad PA"□: The Gulf between User Expectation and Experience of Conversational Agents, in: Proceedings of CHI 2016.

Phillips, E., Schaefer, K.E., Billings, D.R., Jentsch, F., Hancock, P.A., 2016. Human-animal Teams As an Analog for Future Human-robot Teams: Influencing Design and Fostering Trust. J Hum-Robot Interact 5, 100–125. https://doi.org/10.5898/JHRI.5.1.Phillips

"The Origin Of The Word 'Robot.'" Science Friday, 2011,www.sciencefriday.com/segments/the-origin-of-the-word-robot/. Accessed 27 Sept. 2019.

Jonathan Grudin, 2009. AI and HCI: Two Fields Divided by a Common Focus. AI Mag. 30. https://doi.org/10.1609/aimag.v30i4.2271

Story, M.F., Mueller, J.L., Mace, R.L., 1998. The Universal Design File: Designing for People of All Ages and Abilities. Revised Edition. https://eric.ed.gov/?id=ED460554.