Individual assignment IN5480

Table of contents

tidemant@ifi.uio.no

AI - Concepts, definition and history

The term "AI"

The term Artificial intelligence originated in mathematics and engineering and started appearing during World War II when governments were funding bigger and stronger computers (Grudin, 2009). The British mathematician Alan Turing was involved with cracking the encrypted "Enigma" code during the time and played a key role in understanding the possibilities of computational force (Grudin, 2009). Even though Turing published the work "Computing Machinery and Intelligence" in 1950, it was the American mathematician and logician John McCarthy who first coined the term "Artificial Intelligence" in his workshop in 1956 (Grudin, 2009).

Definitions of AI

I feel like it is natural to start with the definition from John McCarthy. He explained AI like this; "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, 1998)

A newer approach to explaining the term comes from Wikipedia. I chose to use this because of its availability; it is the first thing that gets presented when people google the term AI, and therefore many people may base their knowledge on it. This explanation is also based on the thoughts of Stuart J. Russel and Peter Norvig who divides historical definitions into four groups; "thinking humanly", "thinking rationally" "acting humanly" and "acting rationally" (Russel & Norvig, 2009).

"In computer science, artificial intelligence (AI), sometimes called machine intelligence, is intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans. Colloquially, the term "artificial intelligence" is often used to describe machines (or

computers) that mimic "cognitive" functions that humans associate with the human mind, such as "learning" and "problem solving".” (Wikipedia, 2019).

The third definition is one that is mentioned within the “acting humanly” definitions from Russel & Norvig (2009). Rich & Knight (2009) are mentioning that this is “by no means a universally accepted definition”; “Artificial Intelligence (AI) is the study of how to make computers do things which, at the moment, people do better” (Rich & Knight, 2009). This definition avoids philosophical issues regarding the meaning of *artificial* and *intelligence* but still tries to give an outline of what AI is (Rich & Knight, 2009). The bold statement originated from the first version of their book in 1991 and may seem outdated today where there are computational tasks where AI clearly outperforms humans.

The way I coin the term “artificial intelligence” in today's setting; **is machine learning that is using algorithms to interpret and see patterns in big data mimicking intelligence and outputting something trivial**. Increasing processing power and larger data sets has skyrocketed the use of the term Artificial intelligence. I feel like AI is often mixed up with the term “general intelligence”, or that the machines learn the way a human would, which is not the case. The AI may interpret a picture of a cat, but it does not really understand what a cat is, it only matches the input to previous data.

Company that works with AI: Facebook

Facebook is presenting the title “Bringing the world closer together by advancing artificial intelligence” with terms like natural language processing, computer vision and conversational AI (Facebook AI, 2019). Facebook is speaking of AI as the thing to bring people closer, using the technology to strengthen their functions as a social media company. They are mentioning multiple times that they are developing AI “that has a positive impact on people and society” (Facebook AI, 2019), maybe trying to avoid the potential stigma that AI is something bad.

Movie that uses AI: Ex Machina

The movie Ex Machina is about Caleb, a 26-year-old programmer who wins a competition to spend a week at a private mountain retreat/facility belonging to Nathan, his CEO which is comparable to real-life Steve Jobs and Larry Page. Caleb will have to participate in a fascinating experiment in which he must interact with the robot girl Ava; the world's first true artificial intelligence. (Garland, 2014)

In the movie, AI is portrayed as extremely detailed humanoid robots that have been developed beyond the uncanny valley. The movie is proving that the Turing test may not be enough to distinguish an AI from a human. The human-robot interaction is based around speech and natural language between the human and the AI. Even though we can clearly see that it is a robot, the AI is learning from Caleb and the result is that it is able to trick him into thinking it is being fully conscious and intelligent. The movie also touches on an alternative of the Turing test; Mark O. Riedl's The Lovelace 2.0, which rather focuses on creativity to measure intelligence (Riedl, 2014). This can be seen in the movie as the humanoid robot has multiple scenarios of human-like creativity.

Robots and AI systems

The word "robot"

The word robot has been developed over centuries, and what was considered to be robots 20 years ago is now highly outdated (Dautenhahn, 2018). Robots have their roots in automation, doing automated or engineered tasks including water, falling weights, and steam in the early centuries (Brittanica, 2013). It has developed to the modern word "robot" or in Czech "robota", originally meaning hard work or slavery (Britannica, 2019). Going from something humanoid that mainly appeared in science fiction like in Isaac Asimov's science-fiction story Runaround from 1942 (Brittanica, 2019), to a more broad meaning of the word including lawnmowers and voice assistants. It is still a word that is changing rapidly, and we do not

have a common reference point when talking about “robots”, for all we know it may have a completely different meaning in 50 years (Dautenhahn, 2018)

Definitions of “robot”

The Oxford University Press (Lexico.com) has some different definitions of “robot”.

1. “(especially in science fiction) a machine resembling a human being and able to replicate certain human movements and functions automatically.” (Lexico, n.d.)
2. “A machine capable of carrying out a complex series of actions automatically, especially one programmable by a computer.” (Lexico, n.d.)

This is a good example of how split the definitions can be, on the one side you have a humanoid machine replicating human movement and functions. And on the other side you have a complex and automatic machine, often in the form of a computer doing processes a human could never compete with.

Another definition is from Cambridge dictionary; “a machine controlled by a computer that is used to perform jobs automatically.” (Cambridge dictionary, n.d.) In this definition, a robot takes on more of a role as an effective and industrial machine, rather than something creative and human-like.

I define a robot as a programmable machine that can carry out actions automatically. They may have pre-programmed operations, or sensing its surroundings and learning through AI. Even though I do have a strong mental model of the word robot being a humanoid or cyborg-like, this may be images from a previous and outdated understanding of the word. As the term “machine” often is described as a mechanical structure that uses power to control parts, I would also say that a robot has to be tangible and control some form of movement.

AI vs Robots

I feel that the difference between the two is largely related to the mechanical and physical aspects of it. When I think of a robot, I picture something tangible that is moving or

responding with physical feedback. Its hard to visualize AI in the same way, and I feel that modern robots often are a combination of an autonomous AI "brain" and a mechanical robot "body". You can still have robots that don't use AI, and you can have AI which is not a part of a robot, both are common. Robots can still do actions *automatically* without having AI, but *autonomy* does require a degree of artificial intelligence. It is similar in that both AI and robots may have a goal of being as human-like as possible, which also implies being intelligent and creative.

"Spot" - A contemporary physical robot

Boston Dynamics is famous for its life-like moving robots and has recently launched its new robot "Spot" which is terrifyingly good at mimicking the movements of a dog. There are plenty of applications for these kinds of self-stabilizing and autonomous robots, ranging from lifting, monitoring, inspections in environments not suitable for humans and much more. Humans can interact with it by giving it autonomous commands or being the operator of the robot (Boston Dynamics, 2019). The animal-like movement is remarkably natural, placing itself somewhere in the uncanny valley.

Previous videos released from Boston Dynamics where humans are pushing the robots around to show off their stabilization and recovery, has shown that people feel bad for the robot (Boston Dynamics, 2016, 1:25). We know these are just life-like robots, but because its movement is uncannily similar to a human, we react to it.

Universal Design and AI systems

Universal design was first coined by architect Ronald L. Mace as "a concept of designing products and environments for the needs of people, regardless of their age, ability or status in life" (Persson H, Åhman H, Yngling A. A. & Gulliksen J., 2014). It has its roots in other barrier-free and accessible design approaches (Persson et.al, 2014). Mace argued that the first definition was not broad enough, giving the definition; "The design of products and

environments to be usable by all people, to the greatest extent possible, without the need for adaptation or specialized design" (Persson et.al, 2014).
By this, it is understood that universal design is about including everyone to the greatest extent possible with your design. It is about making your design as simple, flexible and perceptible as you can to not discriminate some people from interacting with the design.

Potential of AI with respect to human cognition and emotions

AI is already being used in healthcare, and we are now seeing versions of social robots that are connecting with humans through AI. The potential for an AI to create emotions with people may be a good solution for the ones that has no other sources for this, especially the elderly and people with mental health issues. An AI can always be there to listen and talk to you, and it could offer a form of relief to have a robot to speak to that doesn't get mad or irritated, but rather a personal assistant just for you.

Some technology in this field already exists, like the PARO robot seal that is giving the therapeutic effects of an animal without the logistical difficulties at hospitals and care facilities (PARO, 2014). Or Hasbro's AI cat which is helping the elderly to remember taking their medication as well as having something to take care of. (Brown, 2017)

Potential of AI for including and excluding people

As mentioned earlier, AI is mainly machine learning training on big datasets. This makes the AI mimic real human behavior, often resulting in unwanted results. If the AI is trained on specific data, this will reflect the output and result. As it is limited to the datasets it is learning on, it may also be excluding people. An example of this is that facial recognition software wrongly recognized faces of well known African American women stated by Joy Buolamwini in her video "AI, Ain't I a woman?" (2018).

As time goes on, speach and image recognition and all other sub-categories of AI will improve. This may help inclusion in society, either it is blind people being able to talk to a voice-controlled AI assistant, or someone lonely getting someone to talk to.

tidemant@ifi.uio.no

Characteristics of AI-infused systems

AI-infused systems are everywhere, even when you don't think about it like in recommendations in Youtube or image processing in Snapchat. AI-systems are **learning**, **improving** and fuelled by **large data sets** (Følstad, 2019), which has only been available in the last years. These systems usually have a form of black box AI where you don't really see what is happening behind the scene, but rather put data in and get intuitive feedback back, the process between stays opaque (Følstad, 2019). The systems may help people in sections like health and industry, and can be seen as very useful to remove repetitive and straining jobs. Jobs that people have done for hundreds of years can now be replaced by robots and humans are taking a more controlling position rather than executing the jobs. AI-systems are still in its early stages tho, and they may demonstrate unpredictable behaviors (Amershi et al., 2019). This unpredictability can potentially be disruptive, confusing, offensive and even dangerous (Amershi et al., 2019). Systems may respond differently in each circumstance and are inherently inconsistent due to humans not quite understanding how the AI thinks (Amershi et al., 2019). One of the reasons that people may be negative of these technologies may be their high expectations of the system at what it may achieve (Kocielnik et al., 2019). AI functionalities will always operate at less than perfect accuracy, and therefore not be set with the user's expectations, which may disappoint and lead to an abandonment of the AI-infused systems (Kocielnik et al., 2019). They describe three techniques for setting correct expectations for the users; accuracy indicator, examples based explanation and performance control (Kocielnik et al., 2019).

Youtube as an AI-infused system

Youtube can be seen as one enormous data set for videos, and many hours worth of videos are uploaded each second (Covington et al., 2016). In essence, the users of the platform are mainly watching videos, but behind the scenes is a lot of machine learning, natural language

processing and AI-infused systems. Ranging from image recognition to filtering and recommendations, are all based on machine learning and AI in some way. The recommendations on Youtube is mostly based on deep neural networks (Covington et al., 2016). With all the hundreds of hours of content coming in constantly, they have to process everything to make relevant suggestions for you, which is done through a lot of hidden layers that are opaque for the users. The system is looking at a mix of previously watched and searched videos together with the new incoming videos and popular older videos that might be relevant (Covington et al., 2016). They use this in combination with your candidate profile and puts each video through a ranking process where millions of videos get turned in to dozens that hopefully will be relevant for you (Covington et al., 2016). Youtube profits on the users watching videos, and the users might see it as a positive to get relevant videos handed to them. But, the users do end up using more time watching videos, which in turn makes us go even deeper within the social media bubble, making it "waste" more of our time.

Human-AI interaction design

Amershi et al. (2019) talk about that HCI has been highly discussed for over 20 years in regard to principles and guidelines for creating AI interfaces. Although it has been a lot of positive that has come out of AI, the variability in design and high failure rates shows that designers and developers still struggle with creating AI-infused systems that are intuitive and effective (Amershi et al., 2019). If you design the system so that the users don't have as high expectations, and makes it obvious what the system is cabable of, their acceptance may be higher. They talk about the 18 general guidelines for understandability and how these may help the user accept and use the AI-infused system.

Kocielnik et al. (2019) discuss a lot of the same as the previous article, mainly focusing on how you can shape the AI expectations and how you help the users accept AI in general. They mention three techniques that you can use to set the expectations for the users; Accuracy indicator, examples based explanation and performance control. They talk about that the acceptance of the user will be higher if the user understands the limitations of AI.

They also talk about the impact of false positives and false negatives in the UX, and that few false negatives increase the perceived accuracy.

Firstly, Youtube is being very opaque about what is going on behind the screen. So “G11 Make clear why the system did what it did.” is not really followed. They do not make it obvious for people to see why the things that are recommended to them is actually recommended. There is no button or text explaining how they came up with the result and what it is based on.

They do have a module where they ask you to rate a previous video to “Help us make YouTube better” in the recommended section, so that they can further match the score with other videos. This can be seen as going under “G13 - Learn from user behavior”, youtube is clearly learning from the users behavior. If you are a new YouTube user and the first video you watch is a cooking video, your recommendations will be full of cooking videos. If you press “like” on a video, this also may affect how the machine learning processes new suggestions for you.

YouTube could be more transparent to the user in how they use your data to make suggestions. A simple and truthful explanation would suffice for most people to feel that they have some knowledge about the system.

Chatbots / conversational user interfaces

Følstad & Brandtzaeg (2017) talks about that in the current state of chatbots, one of the problems is that it can not hold a long conversational thread. They mention that even the state of the art assistants ultimately breaks down and that the input from the chatbot becomes irrelevant. Luger & Sellen (2016) says that we often get left in the void between experience and potential when talking to a computer. The success of the language processing gets less impressive when errors and unintelligible responses and bad dialogue gets in the way (Luger

& Sellen, 2016). To cross this gulf, the AI has to become better at understanding the drives and motivations of people and link this to behavior and expressions (Luger & Sellen, 2016). A simple conversation becomes highly complex when you have to account for factors like context, mood, tone and so on, it is not just the mere vocalizing of words (Luger & Sellen, 2016). If a machine cannot answer with these kinds of nuances and process each set differently, the usefulness of something like a chatbot may decrease.

G1: Make clear what the system can do.

A lot of people, me including can find out simple information through a google search or looking at a webpage. Therefore we often think that a conversation with someone, either it is a human or a machine, should not result in something trivial. We may expect to get an answer to a specific problem that we have, even though the chatbot only replies in short pre-programmed phrases. It is then really useful to make the users know what the chatbot is actually capable of from the start, rather than the user starting to ask random questions that go unanswered. If the chatbot cannot give you the answer, it is better to know this right away so you can talk to a human rather than going through a whole conversation that isn't going anywhere.

G2: Make clear how well the system can do what it can do.

This is similar to the first one in that you have to notify the users in some way what the chatbot can do, and if it can do it at all. There is no point in portraying that the chatbot is this magical genie that can answer all your questions when in reality, it cant. Make sure the users get a grip on what they can ask the chatbot and how complex your sentences can be.

tidemant@ifi.uio.no

Collaboration and levels of automation

Endsley (2011) talks about 12 steps of automation, ranging from manual control; where humans perform all aspects of tasks to full automation; where computers carry out all the tasks without human intervention. Somewhere in between these is where you find a lot of the robots that humans interact with daily.

One example that Philips et a. (2016) mentions is Boston Dynamics Big Dog which goes under the taxonomy “replace physical capabilities”. As the name suggests, the robot is designed to resemble the movement and structure of a big dog or a small mule, to successfully transport heavy equipment in rough or uncertain terrain (Philips et. al, 2016). The BigDog consisted of many sensors making it able to operate with a pretty high degree of autonomy, but still being controlled in some way by an operator. The BigDog could fit into the “Shared control” level of automation (Endsley, 2011), since it uses its sensors to execute on tasks, but the goal of the task was set by the operator in the first place. An example of this could be to follow a soldier with the help of a tracker on his foot, the human tells the robot what to do, but the robot calculates the best way of doing that task. It could also fit into the “Supervisory control” level of automation, where the computer generates recommended options and carries them out (Endsley, 2011). This is dependent on what mode the humans have set it to operate at, but the “supervisory control” may also fit in the situation where it autonomously follows a human and corrects itself choosing the best path. This is dependent on what part of the process you look at.

Since this was developed for military purposes, it had to be extremely reliable to not cause dangerous situations. By increasing the level of automation, the robot may produce unexpected actions like running away, reveal positions, or other actions which in turn could end up damaging people. The robot should ultimately just follow someone without them having to interfere or think about operating it, not restraining the people working with it. In this case, this may be too high of a risk, since stealth and fine movement may be needed in

the military operations. This movement may be unpredictable to control if the system is fully autonomous.

If we decrease the level of automation on the BigDog, it would maybe gain som stealth and the ability to move more precisely. It would be more reliable and not take unexpected actions, making it safer in the field. This would, on the other hand, be too reliant on humans, and may not perform the goal it was sent out to do. Having someone fully controlling it would be both cost-ineffective and may not be able to function at all.

Another example that Philips et. al (2016) brings up is the Paro robot which I mentioned earlier. The Paro robot seal is in the taxonomy of "provide comfort", and is often used in hospitals and care facilities (Paro, 2014). The Paro robot has multiple sensors, motors and microphones (Wikipedia, 2019), which help it respond to touching and sounds. Since all the sensors are hidden, it may seem like the robot is really automated, but it really just responds to simple interactions with the sensors so it may be on the "batch processing" level of automation. This is where the computer carries out a set of tasks commanded by a human (Endsley, 2011), which in this case is through touch and speech.

By increasing the level of automation, Paro may improve its connection to the humans by mimicking a real animal even better. It may respond with a response which increases the emotional level to the robot. Even though this might be good, it may get too automated, creating unexpected responses that may not be good for the patients using the Paro, since they are already vulnerable. If the dataset on which the AI would be trained on would be biased, it could also cause bad responses. Since it is a therapeutic tool, it is a fine balance between the reaction of the robot to human input. It may give too much response, making the patient lose its interest when all they want is just something soft to comfort them.

If the level of automation of a Paro goes down, it may lose a lot of its therapeutical value, since it would practically be just a stuffed animal. It may then lose some of its comforting abilities and some of the human-robot interaction, which people may look for. Even though it will lose some abilities, it might still create that emotional bond which is the main goal of the

robot. But if the automation and goes down, there will be less feedback to the human-robot interaction, and the zoomorphism might go down, weakening the bond between the human and the Paro robot.

References

Grudin, J. 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

McCarthy, J. (1998). What is Artificial Intelligence? Retrieved from http://cogprints.org/412/index.html

Russell, S. J. & Norvig P. (2009) *Artificial Intelligence a modern approach* (3. edition). Pearson. Retrieved from https://faculty.psau.edu.sa/filedownload/doc-7-pdf-a154ffbcec538a4161a406abf62f5b76-original.pdf

Wikipedia (2001, last edited 2019). Artificial intelligence. Retrieved from https://en.wikipedia.org/wiki/Artificial_intelligence

Rich E. & Knight. K (2009) Artificial intelligence (3. edition) Mc Graw Hill India. Retrieved from https://i4iam.files.wordpress.com/2013/08/artificial-intelligence-by-rich-and-knight.pdf

Facebook AI (2019) Homepage. Retrieved from https://ai.facebook.com/

Garland, A. (Director). (2014). Ex Machina [Movie]. UK: Universal Pictures.

Riedl, M. O. (2014) The Lovelace 2.0 Test of Artificial Creativity and Intelligence. https://arxiv.org/pdf/1410.6142.pdf

Dautenhahn, K., 2018. Some Brief Thoughts on the Past and Future of Human-Robot Interaction. ACM Trans. Hum.-Robot Interact. 7, 4:1–4:3. https://dl.acm.org/citation.cfm?id=3209769

Encyclopædia Britannica (1999, last edited 2013). Automaton. Retrieved from https://www.britannica.com/technology/automaton

Encyclopædia Britannica (1998, last edited 2019). Robot technology. Retrieved from https://www.britannica.com/technology/robot-technology

Lexico (n.d.) Definition of robot in English. Retrieved from https://www.lexico.com/en/definition/robot

Cambridge Dictionary (n.d.) Meaning of robot in English. Retrieved from https://dictionary.cambridge.org/dictionary/english/robot

Boston Dynamics (2019, 24. September). Spot Launch [Video]. Retrieved from https://youtu.be/wlkCQXHEgjA

Boston Dynamics (2016, 24. February). Atlas, The Next Generation [Video]. Retrieved from https://youtu.be/rVlhMGQgDkY?t=85

Persson H, Åhman H, Yngling A. A. & Gulliksen J. (2014) Universal design, inclusive design, accessible design, design for all: *different concepts—one goal? On the concept of accessibility—historical, methodological and philosophical aspects.* Retrieved from https://www.researchgate.net/publication/271657803_Universal_design_inclusive_design_accessible_design_design_for_all_different_concepts-one_goal_On_the_concept_of_accessibility-historical_methodological_and_philosophical_aspects

Paro robots (2014) Main page. Retrieved from http://www.parorobots.com/

Brown University (2017). Brown / Hasbro team to design smart robotic companions to assist seniors. Retrieved from https://www.brown.edu/news/2017-11-01/aries

Joy Buolamwini (2018, 28. June) AI, Ain't I A Woman? - Joy Buolamwini [Video]. Retrieved from https://www.youtube.com/watch?v=QxuyfWoVV98

Følstad, A. (2019, august). Interaction with AI - Module 2. Presented in IN5480 at Institute for Informatics. Retrieved from https://www.uio.no/studier/emner/matnat/ifi/IN5480/h19/undervisningsmateriale/interacting-with-ai-2019---module-2---session-1---handout.pdf

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

Kocielnik, R., Amershi, S., & Bennett, P. N. (2019). Will You Accept an Imperfect AI?: Exploring Designs for Adjusting End-user Expectations of AI Systems. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (paper no. 411). ACM. https://www.microsoft.com/en-us/research/uploads/prod/2019/01/chi19_kocielnik_et_al.pdf

Covington, P., Adams, J. & Sargin, E. (2016). Deep Neural Networks for YouTube Recommendations.
https://ai.google/research/pubs/pub45530

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

Luger, E., & Sellen, A. (2016). Like having a really bad PA: the gulf between user expectation and experience of conversational agents. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (pp. 5286-5297). ACM.
https://www.microsoft.com/en-us/research/wp-content/uploads/2016/08/p5286-luger.pdf

Endsley, Mica R.. Designing for Situation Awareness: An Approach to User-Centered Design, Second Edition CRC Press. (2011) (chapters 2 and 10)

Phillips, E., Schaefer, K. E, Billings, D. R., Jentsch, F., & Hancock, P. A., 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

Wikipedia (2006, last edited 2019). Paro (robot). Retrieved from

https://en.wikipedia.org/wiki/Paro_(robot)