Individual assignment

erikhma - Iteration 1

IN5480

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

How AI came about:

The American mathematician and logician John McCarthy is credited as the source of the first appearance of the term "artificial intelligence". While the term itself was only thought of and thereby "born" in preparation for a conference at Dartmouth College in the US in 1956 (LiveScience, 2014), the idea of intelligent machines had already been around for several years. Alan Turing wrote in the *London Times* in 1949 that "I do not see why [the computer] should not enter any one of the fields normally covered by the human intellect, and eventually compete on equal terms" (Grudin, 2009, p. 49). Isaac Asimov had also been working with similar ideas by introducing three laws of robotics through his collection of novels by the name "I, Robot" (ibid.).

Three definitions of AI & my own:

Definition of AI taken from Bratteteig & Verne, 2018 (p. 1-2) :

*"*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."

This is a relatively recent definition and shows the concern of the authors when it comes to "mimicry" of human intelligence and perhaps discretely pointing to AI's lack of true emotions and consciousness. The field of Participatory Design (PD) is generally focused on human-human interactions where sharing knowledge and mutual learning is key. While AI may be an extremely fast learner in some regards, it may lack other valuable human traits such as the ones mentioned previously.

AI defined by John McCarthy (according to Stanford University):

"The science and engineering of making intelligent machines"

This broad definition is starting to show its age a bit now as we have seen the rise of AI neural networks and the like which are able to "engineer" themselves through self-learning. The field of AI research is still very much focused on the science and engineering of making intelligent machines though, but likely in a different way than what John McCarthy would have experienced and imagined in the mid-1950s.

ICO definition of AI:

"AI is an umbrella term for a range of technologies and approaches that often attempt to mimic human thought to solve complex tasks. Things that humans have traditionally done by thinking and reasoning are increasingly being done by, or with the help of, AI."

ICO's definition is my preferred one, and it is quite similar to the definition by Bratteteig & Verne (2018). First off, it points out that AI is a broad term used to describe different technologies. This is very important to make clear as many different technological gadgets, programs, artefacts, etc. can be described by saying that they are artificially intelligent. This definition also includes a partial description of intelligence by saying that AI attempts to mimic human thought (e.g. logic, learning, problem-solving) to solve complex tasks. The latter being what we usually think of as requiring intelligence by humans, which Bratteteig & Verne (2018) also included in their definition of AI.

Based on these definitions I would like to incorporate parts of them into my own. I think the man-made aspect of John McCarthy's definition is interesting and deserves to be included. For the intelligence to be artificial it has to be created by humans - at least at some level. The ability to solve "complex" problems is also important to the definition. Such problems require varying levels of intelligence from humans and hence machines which are able to solve these should also be considered "intelligent" to some degree, even though they merely mimic human thought and are not able to be conscious or have emotions - yet.

Brief review of The 'Problem' with Automation: Inappropriate Feedback and Interaction, not 'Over-Automation' (Norman 1990):

The article addresses the notion of "over-automation" and presents a few dramatic examples where accidents happened due to a lack of human intervention when unexpected events occured. Automated systems were involved in most of these, and the author digs deeper into what *really* caused these accidents and how they could have been avoided. Taking a stand against the stance of "automation is too powerful", the author instead claims that it is not powerful enough due to how the presented systems acted in relation to their human operators under the unexpected circumstances of the examples given.

Subtle remarks such as "huh, this is weird" would be what a human pilot could say in an unexpected situation while flying, whereas the autopilot may just automatically adjust what it is doing in order to compensate for mechanical failures in the aircraft, without giving any indication to the human crew that something might be wrong. A lack of appropriate feedback from the automated systems, in

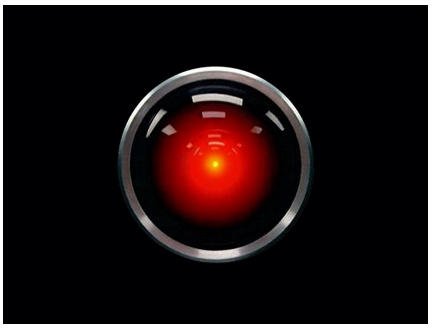
addition to what the author refers to as “mental isolation” on the part of the humans involved, are the culprits of these accidents according to the author.

Description of AI used by Boston Dynamics’ Spot robot:

“Out-of-the-box, Spot has an inherent sense of balance and perception that enables it to walk steadily on a wide variety of terrains. This form of AI that we call ‘athletic intelligence’ allows Spot to walk, climb stairs, avoid obstacles, traverse difficult terrain, and autonomously follow preset routes with little or no input from users.”

The way Boston Dynamics presents their use of AI is that it “allows” their robot to do certain things through what I would call mimicry of balance and perception. It appears as though they see AI as something that enables functional features through a kind of framework.

Film about human interaction with AI:



The movie *2001: A space odyssey* depicts the fictional AI character Hal 9000, who describes himself as “the most reliable computer ever made” who is “incapable of making mistakes or distorting information”. While he may appear helpful to the human characters in the film, the viewers are more likely unsettled by the way he is presented through close-up shots of his red “eye”. He communicates with the human characters with a natural male voice and slightly formal language. ((Spoiler alert!)) As many viewers could perhaps predict, he turns against the human characters when they start planning to disconnect Hal, which would effectively be killing him. Hal appears to have a state of consciousness, which slowly fades as his circuits are being disconnected by the humans.

1.2 Robots and AI systems

How the word “Robot” came about:

The word Robot has its roots in the Slavic language as robota, which means forced labour (“Robot”, 2021). It first appeared as a term to describe “artificial human bodies without souls” in a 1920s play by Czech writer Karel Capek (ibid.).

erikhma

Two different definitions of “robot”:

The Robot Institute of America defined a robot as:

“A reprogrammable, multifunctional manipulator designed to move materials, parts, tools, or specialized devices through various programmed motions for the performance of a variety of tasks” in 1979 (Thrun, 2004).

This is a very specific definition going into detail about the operations a robot should perform, but perhaps not broad enough to cover what we would consider a robot today. Thrun (2004, p. 9) also touches on this in the same article where he states that “Robotics is a field in change; the meaning of the term robot today differs substantially from the term just 1 decade ago.”. This seems to ring very true, looking back at this definition from over 4 decades ago.

Definition of robot by Britannica (2021):

“any automatically operated machine that replaces human effort, though it may not resemble human beings in appearance or perform functions in a humanlike manner.”

Britannica provides a much broader definition compared to the Robot Institute of America’s definition. Robots in 2021 are likely very different from the robots of old from several decades ago, and this shows in the definitions given.

My own definition of robot:

A machine which operates with a high degree of autonomy and is able to sense its environment and also makes decisions and acts based on what it processes about it.

The relation between AI and Robots:

I think AI and Robots are quite connected. Especially modern robots incorporate AI to a high degree in order to mimic human actions and perform tasks which would require intelligence if done by humans. A high degree of autonomy as part of my definition of robot would likely involve some form of artificial intelligence. Processing signals from sensing an environment and deciding on actions based on them requires some form of intelligence, at least I would imagine so as it involves solving relatively complex tasks.

About a contemporary robot:

“Dyret” (Dynamic Robot for Embodied Testing) is a four-legged robot designed and made at the Institute of Informatics at UiO. Its purpose is to use AI to “teach” itself how to walk on different surfaces, such as grass and varying types of carpets. This robot has the ability to change the length of

its legs in order to facilitate this. Human interaction is limited, but its “owner” and maker Tønnes Nygaard seems to have developed somewhat of an emotional bond to his creation (Torgersen, 2020).

1.3 Universal Design and AI systems

Description of Universal Design from DO.IT (2021) with an explanation:

“Universal design is the process of creating products that are accessible to people with a wide range of abilities, disabilities, and other characteristics.”

This definition shows the broadness and inclusive nature of Universal Design. A core element is making products more accessible and better for everyone, which also includes people “without disabilities” (although we can all have disabilities under certain circumstances).

The potential of AI with respect to human perception, human movement and human cognition/emotions:

AI is extremely able to mimic human cognition in particular. AI vision and speech recognition/reproduction has the potential to be vastly superior to their human counterparts. Measurements based on sensors such as LIDAR could be orders of magnitude more precise than what humans are able to produce. Autonomous/self-driving cars are only showing us a glimpse of what this type of technology can do in combination with AI.

Speech and language processing has already come a long way. We are now able to automatically annotate videos with audio in order to make them more accessible, although the fidelity of this technology is not yet in a great place. Reproduction of natural language is starting to take off in the tech sphere, as AI algorithms are able to be trained on audio examples of a specific person speaking and then eventually being able to produce words and sentences with their voice which that person has never said themselves (see for example https://www.resemble.ai/). This has incredible potential for converting books, learning material and other written works into something akin to audio books, read out loud by the AI-generated voice of the author.

The potential for AI to include and exclude people:

As already mentioned, AI has great potential to include people regardless of their abilities or disabilities. The potential for AI to exclude people can be seen in face recognition where the AI has been trained using a limited data set. If a facial recognition AI is trained with pictures of white men, it will become an expert at recognizing images of exactly that. However, if presented with images of a person with a different skin color or gender, this AI will likely seem less-than intelligent as it fails to

recognize the person being shown (Vox, 2020). The principle of “shit in, shit out” is very true when it comes to diversity. AI can be great at recognizing patterns in images, but if the collection of patterns they are being trained on is not diverse enough this could lead to exclusion of one or more groups of people based on their gender, skin color, etc.

Explanation of “understand” and “understanding” with regards to AI. Do machines “understand”?:

I think some AI are perhaps close to being able to “understand” certain things, as in they are able to make models of the physical world around them, but I think it would be near-impossible to make an AI fully understand something more abstract such as a complex concept. An AI capable of truly understanding would likely pass the Turing Test and be close to the intelligence level of the “replicants” seen in the Blade Runner movies. I believe that the ability to understand concepts the way a human does requires a consciousness in order to think critically about something. As far as I am aware, no AI is able to realize or “understand” that it does not understand something either. AI is great at faking understanding, though.

1.4 Guideline for Human-AI interaction

Description of 1 human-AI interaction guideline from Microsoft with a different example:

Guideline 11: Make clear why the system did what it did.

The autopilot discussed by Norman (1990, p. 586-587) in “the case of the loss of engine power” is a great example of why this is an important guideline. The crew of the plane was not sufficiently informed of the critical failure in the aircraft by the automated system and this nearly resulted in a tragic accident, a situation which could have been entirely avoided if the system made it clear why it was severely compensating to keep the plane stable.

Comparison of the Microsoft human-AI interaction guidelines and Norman’s 7 Fundamental Design Principles:

Norman’s first principle of discoverability is directly related to Microsoft’s first 2 guidelines of making it clear what the system can do and how well it can do it. Discoverability is all about is possible, given the current state of something. It also goes without saying that Norman’s second principle of feedback is closely connected to guideline 11, which I already covered. I would say that both Norman’s principles and Microsoft’s guidelines are quite focused on usability and user experience, although Norman is perhaps more on the side of helping with the latter.

References

LiveScience. (2014, December 4). A Brief History of Artificial Intelligence. https://www.livescience.com/49007-history-of-artificial-intelligence.html accessed 07/09/21 16:18.

Bratteteig, T., & Verne, G. (2018). Does AI make PD obsolete?: Exploring challenges from artificial intelligence to participatory design. *Proceedings of the 15th Participatory Design Conference: Short Papers, Situated Actions, Workshops and Tutorial - Volume 2*, 1–5. https://doi.org/10.1145/3210604.3210646

Manning, C. (2020, September). Artificial Intelligence Definitions. https://hai.stanford.edu/sites/default/files/2020-09/AI-Definitions-HAI.pdf accessed 07/09/21 19:14

Boston Dynamics. (n.d). https://www.bostondynamics.com/about#Q7 accessed 07/09/21 21:45.

Norman, D. (1990). The problem of automation: Inappropirate feedback and interaction, not over-automation. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, Vol. 327, No. 1241, Human Factors in Hazardous Situations (Apr. 12, 1990), pp. 585-593 (9 pages)

Robot. (2021, September 8). In *Wikipedia.* https://en.wikipedia.org/wiki/Robot#History accessed 08/09/21 20:10.

Thrun, S (2004). Toward a Framework for Human-Robot Interaction. Human–Computer Interaction, 19:1-2, 9-24, DOI: 10.1080/07370024.2004.9667338

Moravec, H. Peter (2021, February 4). *Robot*. *Encyclopedia Britannica*. https://www.britannica.com/technology/robot-technology accessed 08/09/21 20:32

Torgersen, E. (2020, August 27). No title. https://titan.uio.no/teknologi-informatikk/2020/det-var-et-stort-oyeblikk-da-dyret-kunne-ga-rundt-pa-laben accessed 08/09/21 23:34

DO.IT. (2021, April 9). What is Universal Design? https://www.washington.edu/doit/what-universal-design-0 accessed 09/09/21

ResembleAI. (n.d).
https://www.resemble.ai/ accessed 09/09/21 13:52

Vox. (n.d). Why algorithms can be racist and sexist.
https://www.vox.com/recode/2020/2/18/21121286/algorithms-bias-discrimination-facial-recognition-transparency accessed 09/09/21 19:55

UX Collective. (2020, February 4). Don Norman's seven fundamental design principles.
https://uxdesign.cc/ux-psychology-principles-seven-fundamental-design-principles-39c420a05f84