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

Through the late 1940's there were founded work on topics related to what became AI, and in 1950 several publications were published on the subject. (Grudin, 2009, p. 49) However, the term artificial intelligence was first published in a call for participation in a workshop by mathematician and logician John McCarthy in 1956.

The first works on AI were done by engineers and mathematics, and their goal was to create technology that could rival human intelligence in the same way that computers rivalled humans in the field of arithmetics. (Grudin, 2009, p. 48)

Three definitions of AI

AI has been defined and described with different perspectives.

The writer Pamela (McCorduck, 2004) wrote a book on the history of AI where she define artificial intelligence as (…) that audacious effort to duplicate in an artifact what we humans consider to be our most important, our identifying, property—our intelligence.

Russel and Norvig have categorized definitions of AI into four categories where definitions gathered along dimensions of thinking versus acting and human-like behaviour versus rational behaviour. The designers Bratteteig and Verne has chosen a definition of AI for their paper that reflects that AI is intended to act humanly:

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 & Verne, 2018; Russell et al., 2010).

My definition

Reviewing these definitions, my description of AI could be that artificial intelligence is a term that describe a range of technologies which have developed from the idea to model computational processes on human capacities for reasoning, learning, problem solving, decision making and acting. Several directions of developments of AI technologies have been developed.

Similarities and differences between human autonomy and machine autonomy

For humans, autonomy can be understood as a widely agreed upon value that entails intentionality (Herstad & Soma, 2018) and the freedom to act independently (Finkelstein & Marcus, 2018). Even if humans do not always have the ability to act according to their intent and will, for instance if they are inhibited to do so by disabilities, they have the right to autonomy and to receive support to do so. Autonomy can be understood as the right to make decisions on concerns that affect us directly.

When machines are designed to have, or rather be, autonomous, this refers to the ability to act independently without interference or control from their surroundings. In machines, autonomy is maybe more of an ability than a right or a value.

Human and machine autonomy is similar in that decisions can be made without the intervention or control of others.

Similarities between the concept automatic and the concept autonomous

The concepts of automatic and autonomous are subject to ongoing discussions about their definitions an applications in technological systems, e.g. (Kaber, 2018). They are similar in that they suggest technology that executes tasks without humans directly interfering or controlling the automatically executed tasks.

Systems that are automatic or autonomous may also appear the same to users, in that neither require user's intentional input or control in the execution of the task.

Select one of the articles from the curriculum. Write a brief review from this article. If there is something that is unclear, or you disagree with, or that you question with the article – write about this.

A contemporary project: Self-driving vehicles in Norway

The Norwegian Public Roads Administration are currently running test drives for self-driving vehicles in several places in Norway. In Oslo they have deployed two named buses, Oda and Mads, in collaboration with the public transport company Ruter and Oslo municipality.

On their websites, the Norwegian Public Roads Administration present the buses as self-driving vehicles that are intended to become a service to the citizens of Norwegian cities in the future. They have purchased the technology, which they at one point describe as expensive, and they are testing it to detect problems and devise guidelines for future juridical regulation. They do not refer to the term artificial intelligent. Oslo municipality, on the other hand, use the term Smart transport in the Oslo region in their project title (STOR). In theory project description, the smart concept seems to do more than describe the technology and rather describe a vision to organize traffic and utilize recourses better. Ruter's website on the project seems to have been unpublished. One of the buses, Oda, has been exhibited in The Norwegian Museum of Science and Technology after two years of service in Oslo from 2019 to 2021. (see references for links to the mentioned stakeholders).

AI in fiction: Data

In the TV show Start Trek, the android Data possess artificial intelligence. He is an android so advanced that he is considered a synthetic life form and serves as an officer of Starfleet. He is portrayed in the star Trek shows The Next Generation and Picard, as well as in several of the Star Trek films. Data has multiple roles and functions as a member of the crew on the Federation starship Enterprise.

Human-android interaction

Humans interact with data trough natural language as well as other human communication forms, such as gesture and mimic. Humans in the Star Trek universe is well versed in interaction with another life forms of all kinds, and I think the portrayal of the interaction can also be understood in the context of communicating across different languages and cultures.

Skills and weaknesses

Data has formidable calculation skills and helps the crew out of dangerous situations by calculating surprising paths and manoeuvres for the Enterprise. He continuously learns new things. He is also

sentient, and shows friendship, loyalty and concern for his crewmates. He is portrayed as a gentle and kind person. Data's sentience lead him to understand that he is humanoid, but not human. His longing to discover his own humanity is a recurring theme in the TV show.

Hidden agendas

Even if Data is portrayed as an honourable and gentle being, his programming allows for hidden agendas that Data is not even aware of himself. Before becoming a member of a scientific expedition with the Enterprise, Data has been programmed to find a specific planet. He unknowingly directs the crew and the starship towards his goal and endanger them all. He discovers the truth too late but manage to use his extraordinary calculation powers to save them from being pulled into the planets heavy gravitational field.

1.2 Robots and AI systems

The word Robot

The word robot entered the English vocabulary when it appeared in the 1920 play R.U.R. or Rossu's Universal Robots by Karel Čapek (1880-1938) where it's the name of artificial workers ('The Origin Of The Word "Robot"', n.d.). According to (Szabolcsi, 2014), the archaic slavonic word robot means forced, unpaid labour while robota denotes paid uncompulsory work.

Two definitions of "robot"

"(…), the concept of "robot" is a moving target" (Dautenhahn, 2018, p. 2)

Sebastian Thrun contrasts two definition of robots to show the span of understandings of robot technology.

"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" (Russell & Norvig, 1995).

"An automatic device that performs functions normally ascribed to humans or a machine in the form of a human" (Thrun, 2004, p. 11): Merriam Webster's collegiate dictionary (1993).

The first one was written by the Robot Institute of America in 1979, and the latter is a dictionary definition from 1993.

My definition of robot

Based on the former definitions of robots, I would define robot as an automated technology that moves something. It performs functions and tasks for humans, but it's task can in some instances extend beyond what a human would do as robots can accomplish tasks that humans are not adept at

such as precision work in very small scales or in hostile environments. It's probably also important that it is reprogrammable.

The relation between AI and Robots.

Robots and Ais are concepts that describe a wide range of technologies. Therefore, some robots and AI systems will have overlaps, while other robots and Ais may not have much in common. The context in which both needs to be defined will further affect the choice of defining boundaries.

Possible differences, looking at the definitions mentioned so far, is that a robot moves something, an AI does not need to move something to be an AI.

Artificially intelligent systems will require that the system has abilities to handle contextual variances and make decisions autonomically. Robots, on the other hand, does not require intelligent technology to be defined as a robot. Robots may not always be autonomous in the sense that they can make decisions based on variations, but can probably pass as a robot with an entirely automatic setup.

Shopbot

The cnc router Shopbot is a commercial robot for routing. It is intended to conduct routing in workshops (shops), thereof the name Shopbot.

The Shopbot carries out computer numeric controlled routing where it spins a drill bit along three axis on a material surface. The drawing and programming of the paths that the Shopbot follow as it's working instructions is made by the operator. The operator install the material manually, and interact with the Shopbot via a software interface to set tool paths like depths, speed and force to the drawing, and starts, pauses and stops it's routine from the software interface.

1.3 Universal Design and AI systems

A definition of Universal Design.

> *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.*

I understand Universal Design to be the aim to design technology to be available to all potential users.

The potential of AI with respect to human perception, human movement and human cognition/emotions. You are encouraged to use examples. Please provide two examples of AI systems for including "more" users.

With AI capabilities of natural language processing, AI has the potential for translation between text and speech. This can potentially include users who have different challenges with text-based interfaces, like those with dyslexia or vision impairments. It may also help those with hearing loss by

capturing sound as text. Such technology could also summary and shorten texts to aid people with trouble focusing on reading for longer spans of time.

Examples of such systems are automatic captions of videos on YouTube and voice control for mobiles, PC and smart homes.

AI for image recognition can potentially help include users with vision impairments through being able to automatically add textual descriptions of images that can be read or read aloud by the technology. Face recognition technology can assist users who struggle with giving passwords or

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

AI technology can be applied for more inclusive technology by providing several modes of interaction for a given system that enable people with various abilities to use it.

However, algorithmic bias reflects human bias in for instance training data sets where some representations occur more often than others. Historical examples of this have been seen with facial recognition systems that is trained mainly on images of white faces and can't recognize black people's faces correctly.

Machine learning systems can learn that past and current situations are preferred situations and not account for the possibility that the data set does not reflect a wanted situation, like for example if recruiting for a position where there's historically been gender dominance.
(https://levity.ai/blog/ai-bias-how-to-avoid)

By consciously programming Ais to account for ideal states of inclusion, rather than historical situations, AI technology can hopefully be used for more inclusive technology.

The concepts of understand and understanding

Understanding is a context-dependent concept (*Understanding Understanding*, n.d.). In the WCAG 2.1 guideline 3 is named Understandable. In this context it seems to me that the guidelines points to the importance of considering what users know and try to avoid using specialized terms or erroneous use of pronunciation and terms that can easily lead to misunderstandings.
Understanding can also be understood on a continuum rather than as a binary. The ore one knows about a given element the deeper ones understanding of a sentence or a problem.

and Human-AI interaction design guidelines relates to

When it comes to accessibility of technology then I understand understanding as a state of mind necessary to make use of or operate the technology in the way that the technology is intended to be used or/and in the way that is productive for the user. The ability to understand require the ability to perceive the relevant item and draw on existing knowledge.

When designers make sure that their design is understandable to users we need to consider how deeply something needs to be understood to be used and what most users are likely to understand based on their ability to perceive our message through their sensory apparatus (seeing hearing feeling) and their probable knowledge consisting of educational linguistic and conventional knowledge. This implies that specific knowledge items like abbreviations need to be explained and ambiguous terms needs to be contextually specified.

In the case of guideline 3 in WCAG 2.1 I make sense of the guidelines as pointing to different aspects of technology that needs to be made clear and draw on what is likely to be the abilities and knowledge of the vast majority of users. In the wcag guideline mentioning *making things clear*, like spelling out abbreviations and in the Human-AI guideline 1 and 2 however, understanding seems to be pointing to the need of providing knowledge to users.

On understanding in the WCAG 2.1 and Human-AI guidelines understanding seem on the other hand to be pointing to the need of providing knowledge and information.

Machines don't understand. We say that they can learn, and they have the equivalence of knowledge through available data. They can also be context aware. They have the ability to use previous events as the basis of future decisions. And yet, they do not understand in the way that humans understand. They cannot be in a state of mind. There's no understanding continuum for a computer. They do not understand even a little.

1.4 Guideline for Human-AI interaction

One of the 18 guidelines for human-AI interaction from Microsoft, and describe this guideline with a different example than what is given by Microsoft. (what example?)

11 Make clear why the system did what it did.

HCI design guidelines

http://www.mhhe.com/engcs/compsci/pressman/information/olc/InterfaceDesign.html

Similarities between HCI guidelines and Microsoft's Human-AI guidelines

Provide help facilities that are context sensitive. HCI

Guidelines for recovery and handling of errors is present in both sets of guidelines:

> ***Permit easy reversal of most actions.*** *UNDO or REVERSE functions have saved tens of thousands of end users from millions of hours of frustration. Reversal should be available in every interactive application. HCI*
>
> ***09 Support efficient correction.*** *Make it easy to edit, refine, or recover when the AI system is wrong.*

Guidelines for making help functions available are also available in both sets of guidelines:

> *Provide help facilities that are context sensitive. (HCI)*
>
> ***07 Support efficient invocation. AI***
>
> *Make it easy to invoke or request the AI system's services when needed.*
>
> ***04 Show contextually relevant information.***

Display information relevant to the user's current task and environment.

Differences between HCI guidelines and Microsoft's Human-AI guidelines

Specific to HCI guidelines:

> *Be consistent. Use a consistent format for menu selection, command input, data display and the myriad other functions that occur in a HCI.*

Difficult to achieve in AI over time, maybe?

Capabilities and challenges specific to AI are reflected in the Human-AI design guidelines, such as

> *When wrong:*
>
> ***10 Scope services when in doubt.***
>
> *Engage in disambiguation or gracefully degrade the AI system's services when uncertain about a user's goals.*
>
> ***11 Make clear why the system did what it did.***
>
> *Enable the user to access an explanation of why the AI system behaved as it did.*
>
> *Over time:*
>
> ***13 Learn from user behavior.***
>
> *Personalize the user's experience by learning from their actions over time.*
>
> ***16 Convey the consequences of user actions.***
>
> *Immediately update or convey how user actions will impact future behaviors of the AI system.*
>
> ***17 Provide global controls.***
>
> *Allow the user to globally customize what the AI system monitors and how it behaves.*

Characteristics of AI-infused systems

> *AI-infused systems are ' systems that have features harnessing AI capabilities that are directly exposed to the end user' (Amershi et al., 2019).*

5 key characteristics of AI-infused systems

(Amershi et al., 2019) and (Kocielnik et al., 2019) identify a few key characteristic when exploring how to improve design for Human-AI interactions; uncertainty, inconsistency and opaqueness.

Uncertainty: probalistic

AI-infused systems are probalistic (Kocielnik et al., 2019). That means that they operate at less than perfect accuracy. (Amershi et al., 2019) identify a similar characteristic of uncertainty, which result in the systems often make errors, both false negatives and positives.

Inconsistency: constantly changing

AI-infused systems are constantly learning. They gather data through interaction and train and develop on large data sets. They are therefore impacted by user actions, such as user-generated content (Kocielnik et al., 2019). (Amershi et al., 2019) points to inconsistency as a key characteristic due to AI-infused systems sensitivity to context and small changes in input. They can also keep learning after deployment, in which case development can go both in favourable and unwanted directions (Yang et al., 2020)

Opaque: transparency issues

They are opaque, or black-boxed. Transparency issues (Kocielnik et al., 2019), how to mitigate, eg by showing decision rules. (Amershi et al., 2019) warns that these transparency issues, like for instance behind the scenes personalisation like filtering, can lead to unwanted information hiding.

Capability uncertainty

(Yang et al., 2020) arrive at two key characteristics from their analysis of existing research on design challenges for AI-infused systems; that they think are the roots of design challenges; capability uncertainty and output complexity.

Capability uncertainty has to do with understanding what an AI system can do. Initially that sounds, to me, in some ways to be a characteristic of the Human-AI designers. However, it is also a characteristic of the AI-infused systems themselves because their abilities are uncertain. They can, for example, be able to do something that they can't presently do at the arrival of a new data set. Their available capabilities include both those which are implemented, and all that are technically possible.

Output complexity

Outputs from AI-infused systems ranges from very simple, binary outcomes to very complex outputs with nearly infinite possible results (Yang et al., 2020)*. The relationships between inputs and outputs are also complex and not easy to describe. Amershi et al argues that very complex outputs that evolves over time challenges established design methods.*

Meetings scheduling assistant in Outlook

Outlook has a meeting scheduling assistant that makes itself known by highlighting and underlining phrases and dates in emails with blue text. If you hover the highlighted text, a window pops up, entitled suggested meeting, offering to put the meeting in your outlook calendar.

This system exemplifies that AI-infused systems makes errors and produce false positives when identifying all dates and phrases about events as suggestions for meetings. For example, if someone writes to you and tells you that they are busy at a certain date, or that they are really looking forward to a special occasion that you are not invited to, the scheduling assistant suggest you put it as a meeting in your calendar. This may be confusing to users.

The service seems to have some transparency issues as well. I have never been properly introduced to the service when using my Outlook email, for example. It's rather opaque in the sense that it does not signal how to ensure that it can better identify suggested meetings in emails written by me, for example.

Human-AI interaction design

Amershi et al. (2019) and Kocielnik et al. (2019) discuss interaction design for AI-infused systems.

(Kocielnik et al., 2019) set out to explore how to manage user expectations of AI-infused systems in order to increase user acceptance. They anticipated that having more realistic expectations of what AI systems can do will lead to higher user acceptance and satisfaction. They were able to adjust user expectations using three light-weight techniques not relying on deceptive techniques from marketing or in-depth explanatory models.

They explored how AI systems avoidance of certain types of errors in terms of false positives or false negatives impacted user's perceptions of the system and discovered that a focus on high accuracy, or low false positives, actually made users perceive the system as less accurate than high recall, low false negative, systems.

(Amershi et al., 2019) describe their method and process towards creating the 18 AI usability guidelines for Microsoft through collecting and synthesizing guidelines and research from the past 20 years.

We hope the 18 guidelines presented here and their validation process stimulate and inform future research into the development of domain-specifc guidance.

In creating and selecting the final guidelines, they focused on generalizebility and concrete design guidelines that could be evaluated and observed in a designed interface, like designguidelines that would work towards building trust rather than choosing "building trust" as a guideline.

They emphasize that they made trade-offs to make the guidelines generalizeable and call for further research into how designers make further trade-offs between guidelines in practice. They also remind designers that following guidelines cannot guard against unintended and unwanted ethical consequences in use. (Amershi et al., 2019)

How two of their guidelines apply to the meeting scheduling assistant

> *G8 Support efficient dismissal. Make it easy to dismiss or ignore undesired AI system services.*
>
> *G15 Encourage granular feedback. Enable the user to provide feedback indicating their preferences during regular interaction with the AI system.*

Guidelines 8 is met by making the suggestions for meetings unobtrusive, it's a subtle highlighting in emails that can be completely ignored without any consequences.

However, it's impossible to dismiss it actively. If you hover the highlighted text, I'm offered to act on the suggestion and confirm to put the meeting in my calendar. Guideline 15 suggest that it should be possible to convey my preferences for future interaction by for instance dismissing the meeting, giving feedback as to whether the system have identified a meeting or made an error.

Bender et al. (2021): On the Dangers of Stochastic Parrots

(Bender et al., 2021) examines potential risks associated with very big language models (LMs) and possible solutions.

The size of large LM make them consume considerable resources, especially in terms of power, and also in terms of hardware precious materials. They argue that the cost to the environment should be

the first consideration in building LM systems and that people that are the least likely to reap benefits of the systems are the people who pay the environmental cost first.

When the LMs grow larger, it becomes more difficult and eventually impossible to keep track of what the training data set consist of. This increases risks of biases and tendency of discriminating against minority populations. They propose not curating larger datasets than can be documented at the outset of a project.

They argue that it is important to understand why LM-based systems are successful, and that the size of the data set is not necesseraly the key factor. They argue that LMs are not actually performing natural language understanding, but rather function by manipulating linguistic form. Documenting the relation between data sets and succes could lead to more research on alternative ways of developing LM-based systems, hopefully decreasing the trend to create larger data sets. They call on reserachers to carefully weigh risks against benefits of applying data-intensive systems.

Chatbots / conversational user interfaces

Key challenges with chatbots

User expectations does not match the abilities of conversational user interfaces (Luger & Sellen, 2016). Chatbots failure to keep relevant conversation and provide correct and contextually relevant answers affect user engagement.

As other AI-infused systems chatbots operate with uncertainty inconstistency and opaqueness.

(Amershi et al., 2019) guidelines 1 and 2 can be applied to mitigate these problems. To make clear what the chatbot is designed for, and it's limitation can help users formulate appropriate questions and know whether the chatbot can help the user with their problem.

> *G1 Make clear what the system can do. Help the user understand what the AI system is capable of doing.*
>
> *G2 Make clear how well the system can do what it can do. Help the user understand how often the AI system may make mistakes.*

References

References

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

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

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

Dautenhahn, K. (2018). Some Brief Thoughts on the Past and Future of Human-Robot Interaction. *ACM Transactions on Human-Robot Interaction*, *7*(1), 1–3. https://doi.org/10.1145/3209769

Finkelstein, A., & Marcus, E.-L. (2018). Realizing autonomy: The phenomenology of independence and interdependence while living with Duchenne muscular dystrophy. *Disability & Society*, *33*(3), 433–453. https://doi.org/10.1080/09687599.2018.1427049

Grudin, J. (2009). AI and HCI: Two fields divided by a common focus. *AI Magazine*, *30*(4), 48–58.

Kaber, D. B. (2018). A conceptual framework of autonomous and automated agents. *Theoretical Issues in Ergonomics Science*, *19*(4), 406–430. https://doi.org/10.1080/1463922X.2017.1363314

Kocielnik, R., Amershi, S., & Bennett, P. N. (2019). Will You Accept an Imperfect AI? Exploring Designs for Adjusting End-user Expectations of AI Systems. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3290605.3300641

McCorduck, P. (2004). *Machines who think: A personal inquiry into the history and prospects of artificial intelligence* (2nd ed.). AK Peters.

Russell, S. J., Norvig, P., & Davis, E. (2010). *Artificial intelligence: A modern approach* (3rd ed). Prentice Hall.

Szabolcsi, R. (2014). The Birth of the Term Robot. *Advances in Military Technology*, *9*(1), Article 1.

The Origin Of The Word 'Robot'. (n.d.). *Science Friday*. Retrieved 8 November 2022, from https://www.sciencefriday.com/segments/the-origin-of-the-word-robot/

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

Understanding Understanding. (n.d.). https://doi.org/10.1086/443791

Luger, E., & Sellen, A. (2016). 'Like Having a Really Bad PA': The Gulf between User Expectation and Experience of Conversational Agents. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5286–5297. https://doi.org/10.1145/2858036.2858288

Yang, Q., Steinfeld, A., Rosé, C., & Zimmerman, J. (2020). Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3313831.3376301

Links for stakeholders in Self-driving Vehicles in Norway project
https://www.vegvesen.no/fag/fokusomrader/miljovennlig-transport/smart-mobilitet/forerlose-minibusser/

https://ruter.no/om-ruter/prosjekter/selvkjorende-kjoretoy/ (broken link)

https://www.vegvesen.no/fag/fokusomrader/forskning-innovasjon-og-utvikling/innovasjonspartnerskap/stor/

https://www.tekniskmuseum.no/21-nyheter/893-oda-den-selvkjorende-bussen-pa-teknisk-museum