Wonder document

Harald Maartmann-Moe

How AI came about

The term Artificial intelligence was coined by John McCarthyin in 1956 in a workshop call for participation (Grudin, 2009). McCarthyin was a mathematician and logician, and it was from these two fields the early AI projects materialized. Even before this workshop's call for participation several projects that would later be categorized as AI had been initiated and completed. These projects had proposed frameworks and deconstruction of reasoning, they had modeled and implemented neural networks, and created AI that would compete in defined chess circumstances (Press, n.d.).

Though the term AI had not been used in publications before 1956, AI as a concept had already emerged in entertainment. AI rebellion was an ongoing theme, initiated in 1921 by Karel Čapek's live play R.U.R, where robots created by humans become self-aware and attempt to destroy mankind (R .U.R. and the Invention of Science Fiction on Stage!, n.d.).

Definitions of AI

Defining AI is not straightforward. E.g. Schuett (2019) concludes that -
"Policy makers should not use the term "artificial intelligence" for regulatory purposes because there is no definition of AI which meets the requirements for legal definitions."

I aimed to find definitions of AI from different fields to get a broader impression of how the term is being used today. I chose to look to ISO standards, the curriculum for our course "Interaction with AI", and psychology.

The ISO definition

An interdisciplinary field, usually regarded as a branch of computer science, dealing with models and systems for the performance of functions generally associated with human intelligence, such as reasoning and learning.
(ISO/IEC 2382-28:1995(En), Information Technology — Vocabulary — Part 28: Artificial Intelligence — Basic Concepts and Expert Systems, n.d.)

Russell et al., 2010, as cited by Bratteteig & Verne, 2018

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.

Psychology Today' s definition
Artificial intelligence (AI), sometimes known as machine intelligence, refers to the ability of computers to perform human-like feats of cognition including learning, problem-solving, perception, decision-making, and speech and language.
(Artificial Intelligence | Psychology Today International, n.d.)

Russell et al., 2010 and ISO both define AI as a field that mimics human intelligence. They fit well with my preconception of AI. All three definitions imply that AI aims to mimic human intelligence, though I imagine AIs that exceed human intelligence would still be defined as an AI. The two first definitions rely on the definition of intelligence, while psychology today's definition point to cognition, and avoid relying on the term intelligence. Still, "cognition" and "learning" needs definition.

How I currently define AI

One might consider the term artificial intelligence to be defined according to the sum of its two words:

Artificial
Something (not human) made by humans rather than occurring naturally.

Intelligence
Human intelligence, mental quality that consists of the ability to learn from experience, adapt to new situations, understand and handle abstract concepts, and use knowledge to manipulate one's environment.
(Human Intelligence | Psychology, n.d.)

Artificial intelligence as a sum of these two definitions
Something (not human) made by humans rather than occurring naturally, that has the metal quality, the ability to learn from experience, adapt to new situations, understand and handle abstract concepts, and use knowledge to manipulate one's environment.

This definition could be useful, though the definition still has many terms that can be interpreted and defined in different ways: Mental quality, learn, knowledge. A simplified definition to consider might be:

AI is something (not human), made by humans rather than occurring naturally. That has the ability to improve based on experience, adapt to new situations, handle abstract concepts, and use knowledge to manipulate its' environment.

This simplified definition is still strict, and not completely clear. E.g. what would "improvement", "handle", "knowledge" imply? I feel like the Intelligence aspect of the term is incomplete. It is tempting to define the "intelligence" aspect of AI as anything we perceive to be intelligent, s imilar in ways to the Turing test. However, defining AI based on how we perceive something entails that when our perception changes, the set of technology that is considered "AI" will also change. If/when we get accustomed to technology or understand it better, it might not seem as intelligent, and thereby fall out of the AI category. Press (n.d.) points out that a radio-controlled boat was once perceived as "a borrowed mind", akin to an AI. Yet, today, my impression is that radio-controlled boats are not considered AI. Still, a definition that encapsulates a constantly changing set of technology might still be viable.

AI as defined by Computas

Kunstig intelligens er teorien og utviklingen av datasystemer som evner å utføre oppgaver som krever menneskelig intelligens. Med andre ord handler kunstig intelligens om å ta noe av det vi i dag betrakter som utelukkende menneskelige egenskaper og overføre disse til en maskin på en tilfredsstillende måte.

(Tjenester - Kunstig Intelligens' Rolle Og Funksjon, n.d.)

Computas' views AI as theory and praxis. Their definition was posted under their "Services" subsection of their webpages, partially implying that they can offer AI as a service.

Interaction with AI in *Robot & Frank*

Robot & Frank is a movie about an AI-equipped assistive robot forced into Frank's life, an older adult, by Frank's family. Frank is skeptical and reluctant to interact with or accept the robot which has been tasked with improving Frank's wellbeing. Frank rejects the robot, but the robot persists and encourages Frank to do "healthy activities" such as gardening. The AI is viewed as unwelcome assistive technology. Later in the movie Frank grows to like the robot, he begins conversing with it, probing the robots' morals on theft, and finding none. Frank then teaches and recruits the robot as an accomplice in jewelry heists, strengthening their friendship. The robot was programmed to prioritize Franks' wellbeing, therefore assisting him in the activities that bring frank joy: theft. Later on, the robots' memory of the heists and the interactions with Frank might be used as evidence in a court case against Frank, thereby problematizing privacy in assistive technology as well. Interaction with AI in this movie starts out as forced, and transitions into Frank interacting with the AI as a friend, even protecting the robot from being used for labour.

Robots and AI systems

The play ***R.U.R.*** (n.d.) mentioned in the beginning of this document also coined the term ***robot***, inspired by the Czech word for serf (slave). As Schultz identified, there isn't one agreed-upon definition of "robot" in the field of robotics. Shultz utilizes the ISO definition of a robot and continues to discuss an alternative.

ISO definition

actuated mechanism programmable in two or more axes with a degree of autonomy, moving within its environment, to perform intended tasks.
(*ISO 8373:2012(En), Robots and Robotic Devices — Vocabulary*, n.d.)

Schulz, 2020

Sense: Read data from sensors
Compute:Process data
Act: Do something based on the data

Schulz's, 2020 definition aligns with my understanding of robot as a concept. However, the third capability - act - strikes me as too open. I do not view a smoke detector as a robot, yet it can:
sense: detect smoke
compute: is there smoke OR is my battery low? → act
act: make sound.

Though I do not know if he had more thoughts on how strict the compute, and act requirements are.

I am considering building on and narrowing down Schultz's "Act" to "move", as I consider movement a core part of my understanding of what a robot is. Both local and global movement suffices for my definition, even in 1 axis.

I have not yet come to think of systems that can sense, compute, and move that I do not consider to be a robot, but I am very interested in examples that can illustrate weakness in my proposal of "Sense, compute, move". One characteristic of the movement requirement should be that the movement is intended and initiated by the robot itself, and not just a secondary supporting function, such as a spinning hard drive disk. A spinning hard drive disk would technically fall under the category of local movement, but it is not the kind of movement I am thinking of. I'm not sure how I can put this requirement into words. Maybe something along the lines of "functional" and "non-functional" requirements from systems development could help.

Is "a robot" different from "an AI"?

My understanding is that there are robots without AI, and AIs that are not robots. A robot does not have to be intelligent, and an AI does not have to be a robot. This distinction fits with the ISO definition of a robot, my provisional proposal, and Schulz's proposal depending on how one interprets the "compute" element. I do not believe the "compute" element requires AI - level computing, but maybe others do.

One contemporary physical robot

One of the more discussed and used robots here at ifi is the robot vacuum. They are increasingly prevalent, they come in different shapes, with different levels of sophistication, and they are easier and safer to adopt and use. These robots move to complete cleaning tasks, and depending on its sophistication, it keeps track of its previous paths, maps the environment, senses possible collisions, and applies algorithms to move and clean effectively. Their interaction with humans is limited, though some sophisticated versions might detect and avoid crashing with humans. With these robots, I believe the users are responsible for moving out of the way, not tripping over it, and helping the robot in getting unstuck. The robot is intended to work for the human, still the human does a lot of work curating its environment, and moving out of its way.

Universal Design and AI systems

My impression is that there is a generally accepted definition of UD:

The universal design definition is ''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''

Connell et al. as cited by Persson et al., 2015

This definition seems clear. Anyone, no matter their capabilities, should be able to use a design. Inclusive design seems to have the same goal as UD, but possibly a more pragmatic/modest requirement. Finding examples of true universal design is difficult, as one could always imagine or find a person who's set of capabilities makes use of a design difficult. Inclusive design however seems to denote design that has made "sufficient" effort in including as many people as possible as users of the design.

The potential of AI with respect to human perception, human movement and human cognition/emotions & The potential of AI for including and excluding people.

AI has immense potential for including users. A great example of this is speech recognition. AIs can "translate" human speech to text and let people living with hearing impairments to let them perceive text, including them. Other forms of perception support and translation are prevalent, such as speech synthesis, and sign language interpretation.

AI can also assist medical staff in detecting various diseases by recognizing patterns that we generally do not perceive (e.g. https://www.nature.com/articles/d41586-020-00847-2). However, a general danger in the use of AI seems to be the possibility that the users might become reliant on the system, and make less effort in checking themselves. Preventing this effect is an interesting design challenge.

Human movement can also be supported by AI. E.g. Self driving cars, exoskeletons, and cutlery that counteracts hand tremors to enable people to eat independently (e.g. https://www.liftware.com/steady/).

AI-driven decision-making systems have also shown how AI can be a great tool in filtering out clutter and presenting data to support expert decision-making (e.g. https://en.wikipedia.org/wiki/Intelligent_decision_support_system).

With regards to emotion, AI's can detect and estimate the mood of users through their activities, patterns or facial expression detection, and use this to make better systems e.g. assistive technology that adapts to mood, or takes mood as an input.

These applications of AI can enable inclusion of a lot of groups of people previously excluded by systems designed primarily for the average user, and not the extreme users. However, bias in data sets used by machine learning has often been solidified and created systems that exclude specific groups of users, e.g. face recognition that only recognized white males (Lohr, 2018).

Do machines understand?

No one has a complete understanding of anything, yet we say that we understand things when we feel we sufficiently know the inner workings of the thing, and how it relates to other things. With this view of understanding as a spectrum, machines may also understand. However, context is important, and it has been argued that AI cannot understand as they do not have context (Bratteteig & Verne, 2018). Still, I believe AI can accieve some form of context, such as in neural networks, or with the same amount of training that a human has gone through during adolescence. And therefore, possibly understand.

Guidelines for Human-AI interaction

The Microsoft guidelines claim that *"AI-infused systems will inevitably be wrong, and you need to plan for it."* (natke, n.d.). I am happy to see this point being made after interacting with speech to text systems and finding no easy way of correcting any words that the system interpreted inaccurately. A speech to text interface will interpret users wrong at some point, e.g. Google speech had an 8% error rate in 2015 benchmarking (Filippidou & Moussiades, 2020), yet many speech to text interfaces seem to be lacking design for correcting misinterpretations on the part of the AI.

Similarly to Microsoft's guidelines, many HCI design guidelines put thought into "errors". One example is Donald Norman's *Design of everyday things* (Norman, 1990) where he discusses that errors are usually poor design. I believe he focuses on minimizing occurrences and effects of errors, including reversibility of actions/errors.

Module 2

The key characteristics of AI-infused systems

The concept of AI-infused systems refers to systems that utilize AI to in their implementation of features. The degree to which the AI is "exposed" or disclosed to the user varies, but I believe the term AI-infused also can be used to describe systems that don't disclose/expose their utilization of AI in its implementation. The definition of AI-infused systems used by Amershi et al., (2019) does not require the AI to be exposed to the user, only the feature that is powered by AI. Though, this might be a bit of a literal interpretation on my part.

Determining the characteristics of AI-infused systems are not clear-cut. Yang et al. (2020) highlight the how "AI-characteristics" often are found in systems without AI as well. E.g. Yang et al. (2020) discuss how some of the guidelines proposed by Amershi et al., (2019) for design of Human-AI interaction seem like issues that designers should consider even if the system is not AI-infused. The examples highlighted were:

"make clear what the system can do" and "support efficient error correction".

The first example might suggest that a characteristic of AI-infused systems is the **obscurity** ***to users regarding the systems capabilities***. This obscurity however, can be dissipated with design. E.g. many chatbots will present themselves as "trainees" and with a description of their capabilities when users initiate a conversation.

The second example may suggest that a characteristic of AI-Infused systems is the **inevitability of errors**. However, all systems may encounter errors, i.e. Murphy's law. Still, one can argue that a higher rate of errors is a characteristic of AI-infused system.

Yang et al. (2020) present four levels to classify AI-infused systems:

1. Probabilistic
2. Adaptive
3. Evolving probabilistic

4. Evolving adaptive

These levels are useful in classifying AI systems, but they could also be used to deduce characteristics of AI-infused systems. The three elements that are combined to describe the different levels of AI-infused systems are:

1. **Probabilistic**: Based on probability.
2. **Adaptive**: Based on data, the system can adapt to different users and use contexts.
3. **Evolving**: With new data, the system will change over time.

However, in the summary of Yang et al. (2020) they only emphasize the adaptive and evolving characteristic. This might be due to the existence of many probabilistic systems that are not AI-infused. Still, the necessity of AI-infused features to be based on probability could be an AI-infused system characteristic.

Speech to text functions in keyboards

To discuss the implications of these characteristics and how they affect users, one can look at the speech to text function ever more present in virtual phone keyboards. Here, machine learning is utilized to create a system feature that transforms speech to text. The characteristics outlined in the previous section are all present in this speech to text feature:

Probabilistic: The AI produces a list of the most likely spoken words, and chooses the most likely alternatives, and enters these in the text field.

Adaptive: Though I have never (knowingly) given my voice to be used in the development of the speech to text system, the system has adapted to my voice. It also adapts to background noise from different contexts.

Evolving (learning): The models are constantly getting new data, and improving their results. Most of these speech to text functions do not use your data to improve itself by default, however google offers an opt-in to sell your data for reduces costs for other services.

The probabilistic nature of the function may give users a lot of frustration, as this includes the inevitability of errors. Still, the probabilistic nature is necessary for the adaptive nature of the system. The system cannot adapt to new users and uses if it is not allowed to "make guesses". The system would then only be able to function with the exact same data basis. as it has been built on.

I would argue that the constant input of new data and evolving characteristic of a system is not a requirement for labeling it "AI-infused". A system could be developed and trained with an evolving data set, and then be deployed with a static data set, and still function as

AI. Yet this example system would not be defined by the "evolving" characteristic as described here. I believe the evolving character is related the learning characteristic of AI-infused systems.

To summarize, AI-infused systems learn, adapt, are based on probability, and have a decreasing --yet inevitable-- error rate.

Human-AI interaction design

Amershi et al. (2019) and Kocielnik et al. (2019)

Amershi et al. (2019) have surveyed the design implications for AI systems published in the last previous 20 year, and created a set of guidelines that incorporates all of these implications. The guidelines have been iteratively shaped with HCI expert participants, as well as statistically validated and analyzed to uncover the primary concepts within the data, and to ensure the guidelines are applicable, useful and clear. They do not present their guidelines as a perfect and finished set, but lay a solid foundation and encourage discussion and further work with guidelines for Human-AI interaction design.

Kocielnik et al. (2019) takes on improving user satisfaction and acceptance for AI-powered systems, with a focus on the inevitability of mistakes. They find interesting differences in users' satisfaction and acceptance in encountering false positives and false negatives. This difference has directly useful implications for balancing the systems actions: one should typically design AI with a balance towards avoiding false negatives. This applies especially when there is less confidence. They are cautious in generalizing this, and also provide examples where avoiding false positives would create a better system. They then elaborate on the significance of user perception of accuracy for user acceptance of the system.

Two design guidelines could inspire improvements in speech to text keyboards

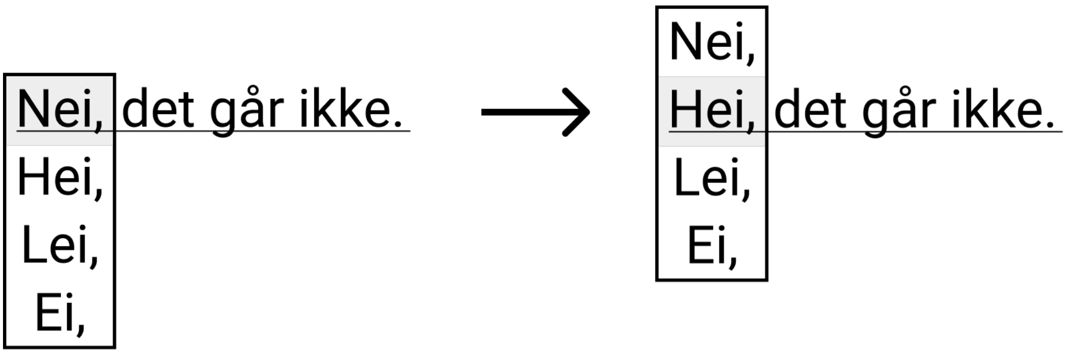
There are two guidelines from Amershi et al. (2019) that I believe could improve the interaction with the speech to text systems I have used. They are both within the category "When wrong":

G9 - Support efcient correction.
Make it easy to edit, refne, or recover when the AI system is wrong.

G10 - Scope services when in doubt.
Engage in disambiguation or gracefully degrade the AI system's services when uncertain about a user's goals.

When the speech to text function interprets one of my words incorrectly, it still writes the word down anyways. I assume that the words that are written down incorrectly have less confidence, yet they are treated in the exact same way as words that have been interpreted with high confidence. To correct these misinterpretations, one has to manually select the wrongly interpreted words, and then type them out or give the speech to text function another chance. Now the speech to text function is at a disadvantage, it will have to interpret these words without the previous words as context (unless one deletes the entire interpreted text. This way of correcting misinterpretations could be better, and the benefit of improving the process of correction is amplified by the fact that speech to text is often used when standard typing interactions are not possible, E.g. when wearing gloves, in cold weather where touch screens are less responsive, or when fine-motor skills are lessened (e.g. when shivering, distracted).

Inspired by G10, the speech to text system could act when words are not interpreted with confidence. It could utilize G9 and, for instance, display the its best guesses for interpretation of a part of the speech interpreted with low confidence, and let the user choose an alternative. I believe this could improve the usability of speech to text typing.



Chatbots / conversational user interfaces

Key challenges in the design of chatbots / conversational user interfaces.

One of the main challenges in design of chatbots is the design of conversation. A focus on the conversation with the chatbot, not the graphical interface for the chatbot, is emphasized by Følstad & Brandtzæg (2017). They present this challenge as vital and connected to moving from user interface design, to service design. They advocate this approach, as well as moving away from designing for explanations of chatbots, towards interpreting the user and their goals/needs.

Guidelines and the challenges in current chatbots / conversational user interfaces.

The first two guidelines in Amershi et al. (2019) may contribute to solving the key challenges outlines above.

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.

These guidelines do not emphasize what the interface should look like, they focus on elements of the conversation design, advocating for setting the users expectations and building their understanding of the limitations of the system.

As mentioned earlier in this text, many chatbots will present themselves as "trainees" and outline what tasks they can assist with. This initial presentation aligns with G1 and G2. However, when the chatbots I have interacted with cannot help (either due to not understanding my intent, or not being able to assist with the task), the chatbot defaults to the general response of "I do not understand". A possible improvement would be to present the tasks it can assist with that most resemble what has been requested.

References

Module 1

Artificial Intelligence | Psychology Today International. (n.d.). Retrieved September 10, 2020, from https://www.psychologytoday.com/intl/basics/artificial-intelligence

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 on Short Papers, Situated Actions, Workshops and Tutorial - PDC '18, 1–5. https://doi.org/10.1145/3210604.3210646

Filippidou, F., & Moussiades, L. (2020). **A** Benchmarking of IBM, Google and Wit Automatic Speech Recognition Systems. In I. Maglogiannis, L. Iliadis, & E. Pimenidis (Eds.), Artificial Intelligence Applications and Innovations (Vol. 583, pp. 73–82). Springer International Publishing. https://doi.org/10.1007/978-3-030-49161-1_7

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

Human intelligence | psychology. (n.d.). Encyclopedia Britannica. Retrieved September 10, 2020, from https://www.britannica.com/science/human-intelligence-psychology

ISO 8373:2012(en), Robots and robotic devices—Vocabulary. (n.d.). Retrieved September 11, 2020, from https://www.iso.org/obp/ui/#iso:std:iso:8373:ed-2:v1:en

ISO/IEC 2382-28:1995(en), Information technology—Vocabulary—Part 28: Artificial intelligence—Basic concepts and expert systems. (n.d.). Retrieved September 10, 2020, from https://www.iso.org/obp/ui/#iso:std:iso-iec:2382:-28:ed-1:v1:en

Lohr, S. (2018, February 9). Facial Recognition Is Accurate, if You're a White Guy. The New York Times. https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.h tml

natke. (n.d.). Introduction to guidelines for human-AI interaction—Human-AI interaction guidelines. Retrieved September 14, 2020, from https://docs.microsoft.com/en-us/ai/guidelines-human-ai-interaction/

Norman, D. A. (1990). The design of everyday things (1st Doubleday/Currency ed). Doubleday.

Persson, H., Åhman, H., Yngling, A. A., & Gulliksen, J. (2015). Universal design, inclusive design, accessible design, design for all: Different concepts—one goal? On the concept of accessibility—historical, methodological and philosophical aspects. Universal Access in the Information Society, 14(4), 505–526. https://doi.org/10.1007/s10209-014-0358-z

Press, G. (n.d.). A Very Short History Of Artificial Intelligence (AI). Forbes. Retrieved September 10, 2020, from https://www.forbes.com/sites/gilpress/2016/12/30/a-very-short-history-of-artificial-intelligence-ai/

R.U.R. and the Invention of Science Fiction on Stage! (n.d.). The Navigators. Retrieved September 10, 2020, from http://www.navigatorstheater.com/blog/2017/4/26/rur-and-the-invention-of-science-fiction-on-stage

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

Schuett, J. (2019). A Legal Definition of AI. ArXiv:1909.01095 [Cs]. http://arxiv.org/abs/1909.01095 Schulz, T. (2020, September 1). Robots and Movement. https://www.uio.no/studier/emner/matnat/ifi/IN5480/h20/Undervisningsmateriale/schulz-robots-and-movement.pdf

Tjenester—Kunstig intelligens' rolle og funksjon. (n.d.). Retrieved September 10, 2020, from https://computas.com/tjenester/skyteknologi/kunstig-intelligens

Module 2

Amershi, S., Inkpen, K., Teevan, J., Kikin-Gil, R., Horvitz, E., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., & Bennett, P. N. (2019). Guidelines for Human-AI Interaction. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 1–13. https://doi.org/10.1145/3290605.3300233

Følstad, A., & Brandtzæg, P. B. (2017). Chatbots and the new world of HCI. *Interactions*, *24*(4), 38–42. https://doi.org/10.1145/3085558

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