# Module 1

### Concepts, definition and history

Alan Turing lead the code breaking team that managed to decrypt the encrypted codes used by the Germans during World War II. This accomplishment made people wonder what else computers could do in the future and whether it was possible that computers could be able to compete with the human intellect (Grudin, 2009). Grudin said that the term “artificial intelligence” was first used in a workshop by John McCarthy in 1956 (Grudin, 2009), but actually the term was first used in the project proposal in 1955 (McCarthy et al, 1955).

### Definitions of AI

**One definition of AI:** “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, p. 1-2).

Bratteteig and Verne’s definition is focused on the fact that computers is not able to reason and act in a context-based manner like humans, which means they can only mimic our intelligence.

**Another definition of AI:** “The ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings.” (Copeland, 2006).

This definition doesn’t limit the artificial intelligence to only cover human intellect or behaviour. One can argument that for instance some animals can be viewed as intelligent in comparison with other animals.

**Third definition of AI:** "AI … is concerned with intelligent behavior in artifacts" (Nilsson, 1998)

In order for this definition to be understood you need to know the definition of intelligence, this relates to the definition by Copeland, and then add this definition into a thing or an artifact.

**My definition of AI:** "Learning, improving and understanding"

I am not that pleased with this definition. By learning I mean that the system should learn from training on data sets and use. By improving I mean that the system should improve due to adapting what it has learned, which leads to some sort of understanding how to adapt and being able to see contexts.

## Google and AI

Google is one of many contemporary companies that work with AI. They describe AI as a service that can make information more accessible to everyone. They use AI for their own products such as search engines, image-search, speech-based systems and translate suggestions (Google, nd).

## AI in Westworld

Westworld is a series where humans or "guests" meet lifelike robots or "hosts" in a fictional wild-western themed world. In Westworld the guests can get different experiences by picking from a set of predefined storylines, where the hosts play characters in the story. It can be compared to a roleplay or an interactive game. The interaction between the guests and hosts is the same as human to human interaction, but since the guests know that the hosts are robots, they don't treat them with the same respect as other humans.

## Robots and AI systems

The word robot come from the Church Slavonic word “robota” which means forced labor or servitude (Markel, 2011). Karel Capek, the director of the play R.U.R, which is short for Rossum’s Universal Robots, also introduced the word robot (Markel, 2011).

Definitions of Robot

**One definition of Robot:** “Robot, any automatically operated machine that replaces human effort, though it may not resemble human beings in appearance or perform functions in a humanlike manner.” (Moravec, 2005)

**Another definition of Robot:** “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” (Thrun, 2004, p.11)

In both of these definitions the robots are described as an artifact that is able to do physical tasks. A robot does not need to look humanlike or perform human-like tasks. Often a robot is a replacement for a machine which usually needed to be operated or handled by a human to work, for example a lawn mower or a vacuum cleaner or more old school industrial robots.

**My definition of Robot:** “A robot is a machine that can perform human tasks without needing to look like a human or perform the task in a humanlike manner”

My definition is inspired by Moravec (2005) since I find this definition highly representative of my own thoughts about robots. Robots are often specialized in one task, but I believe that this can change in the future of robotics and that is why I did not mention it in my definition.

**Robot vacuum cleaner**

Robot vacuum cleaners is more and more common in Norwegian homes. The vacuum cleaner robots moves across the floor in a household and if it meets an obstacle it turns around and finds a new path. Humans can interact with the robot by choosing time intervals which they find convenient for the robot to do it's cleaning job. The robot vacuum cleaner is an example of a robot with one specialized task.

## Universal Design and AI systems

**Definition: "**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." (NDA, nd)

Universal design is design for all, this means that everybody should be taken in consideration both in the digital and physical world. When designing for people with disabilities either it's cognitive, motorical or perceptual the outcome often leads to a better design for everybody. An example of physical universal design is wheelchair ramps, the intention of building these ramps was to give people that sit in wheelchairs access into buildings. Now the wheelchair ramps are used by everyone, it's convenient for people with or without disabilities who are for instance pushing their bicycle or wagon.

Universal design is supposed to be inclusive, but some AI systems has met some challenges when it comes to include everyone. One example of AI being exclusive is of facial recognition. Some AI builds on machine learning algorithms which uses huge datasets for training the AI. Many examples have been given where AI has failed to recognise people with darker skin color. One person who has been addressing this problem is Joy bla bla, she showed the world that AI facial recognition tools failed to determine the gender and names of some of the most historical females. The problem within AI facial recognition is that the datasets used to train the AI has not represented the entire population. This problem doesn't only relate to AI facial recognition but to AI in general. When the datasets used to train the AI systems is based on averages it can potentially lead to excluding edge cases like people with impairments. This is a huge problem because users in the edges of the normal distribution bell curve might be the ones who need universal design the most in the first place.

# Module 2

**AI-infused systems**

When designing an AI-infused system it's important that the system is dynamic and has the ability to change. Of course there either are or should be restrictions to how much the system can change. The AI-infused system should make clear to the user what it is able to do and how well it can do it (Amershi et. al, 2019). When interacting with for instance Google Assistant you are provided with some examples of what you can ask for. The AI can then learn from your searches or context of use and improve and personalize it's services.

Sometimes when the user is interacting with an AI-infused system the system can't detect or understand what the user is asking for. When this happens the system should make it easy for the user to recover or refine the input. If we still use the Google Assistant as an example it could be that the user asks for a complicated question like why did the tram-traffic stop in Oslo on Saturday. The Google Assistant may not be able to answer the question, but it could degrade it's services and provide the user with an Google search on some or all of the input data provided by the user.

From a user perspective it may be hard to understand why the service provides this kind of output, given the user input. This leads us to the third characteristic which is black box. Black box is a metaphor for the process between providing input and being provided with output. When the system has trouble with validating output, or make understandable output, it should make clear why it didn’t manage to do so and what it did instead. The fourth characteristic is that AI-infused systems are fuelled by big data sets. Because the data sets grow as we provide user input the loop kinda closes because the AI-infused system is then learning from the new input and how the user behaves. (Fløgstad, 2019)

I used Google Assistant on my mobile phone as an example of an AI-infused system. With Google Assistant you can search through over one million actions either with voice or text entry (Google, nd). The Google Assistant can help you with anything from remembering where you parked your car, tell you how the weather is going to be or provide shortcuts to Google searches. It’s probably trained on Google’s database, this database is growing everyday which provides the Assistant with new up-to-date training material. It’s probably also improving from use, personalizing and optimizing it’s services to the users. I’ve used the assistant for setting alarm clocks and reminding me what’s on my schedule for the day, but it makes some mistakes if I ask more complicated questions or make longer sentences. Sometimes the assistant doesn’t understand what I say when I’m using voice entry, and it don’t quite understand when I try to correct the input. When this happens I often try to either type text entry or use a Google search instead. I think this makes the user avoid asking difficult questions to the assistant, which makes it gather less user data to improve on.

**Human-AI interaction design**

The principle of consistency is one of the existing guidelines for user interface design. AI-infused systems change from learning from use over time and adjusts to the user making it behave differently from user to user. This conflicts with the principle of consistency in user interface design where we want to minimise unexpected changes in the user interface, making it easier and more intuitive to use different systems due to standardized mechanisms for interaction and layouts. (Amershi et. al., 2019)

Amershi et. al. is addressing the need for a set of shared guidelines for AI-infused systems and together with a team of researchers from Microsoft they have made 18 guidelines for AI design. The guidelines are based on a total of 168 potential AI design guidelines gathered through reviews of existing literature, with the goal of making a set of relevant and clear guidelines. (Amershi et. al., 2019)

Expectations to the capabilities of AI-infused systems vary among different users which can result in users being disappointed with the imperfections in the AI systems. Kocielnik et. al. (2019) studied how methods for setting expectations could impact the users acceptance of AI-infused systems, using a Scheduling Assistant to test different techniques and how these are perceived by the users.  (Kocielnik et. al., 2019)

One guideline by Amershi et. al. (2019) is “Make clear of what the system can do”. When opening the Google Assistant it gives you the choice to learn more about the services it can offer to the user along with some suggestions based on earlier use. When it makes suggestions based on earlier use it also follows guideline number 13 “Learn from user behavior”.

I find that the Google Assistant can be quite dumb sometimes and don’t follow the guideline 9 “Support efficient correction”. I will provide an example to help explain one situation where I found it difficult to use the assistant to edit my entry.


Fig 1) So here you can see that I was trying to create a reminder for my meeting tomorrow at 12 pm. The assistant didn’t perceive the information that I provided in my voice entry, and only manage to recognise 12, but thought I meant 12 am. This gave me the opportunity to correct the information, testing if it followed guideline 9. See figure 2

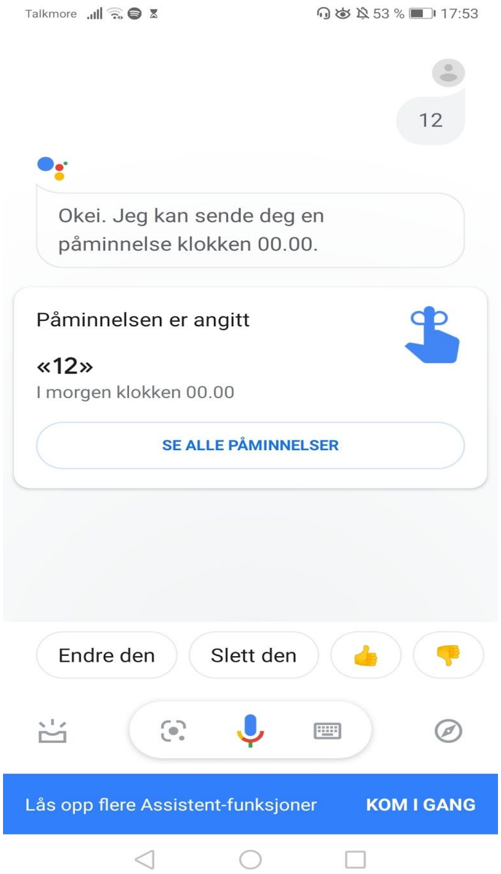


Fig 2) I then tried to correct the time for the reminder and repeated “12”. Google Assistant then thought I would like a reminder at 12 am reminding me about 12. Because this makes no sense I then decided I had to correct it, clicking the “Endre den”-button. Figure 3 shows what happened next.

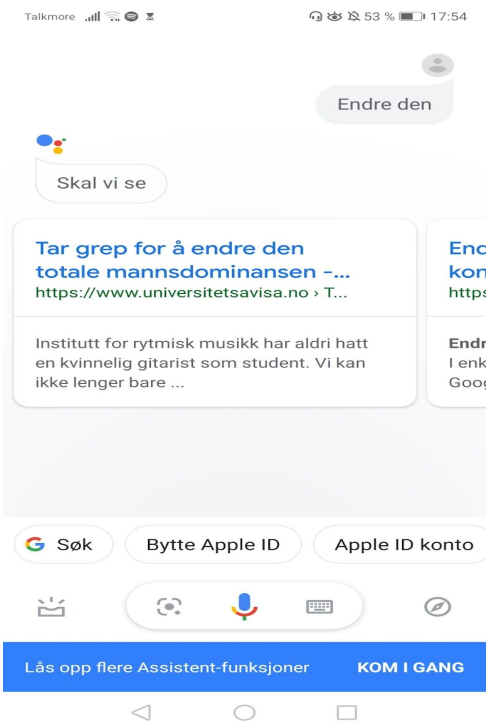


Fig 3) I clicked the “Endre den”-button expecting the Google Assistant to understand that I wanted to change the reminder I had set saying “12” at 12 am. Instead the assistant thought I wanted to do a Google search on the words “endre den”. The assistant then suggested an interesting article in a student newspaper about the first female guitarist student at the institute of rhythmic music.

Based on these figures I believe that I have managed to demonstrate that Google Assistant does not always support guideline 9 “Support efficient correction”. This also relates to the article by Kocielnik et. al. about setting expectations for the AI system. I can understand why the assistant thought I wanted the reminder to say “12” since it asked me what the reminder should be about, but when the system provides a choice to change it or “endre det” I expected it to change the reminder. I think guideline 9 by Amershi et. al. (2019) could inspire improvements in Google Assistant. I believe it is a good solution to offer the user an opportunity to change or delete a reminder or other kind of entry, but the system should then understand that the user wants to make changes or correction on the previous input or output.

**Chatbots / conversational user interfaces**

Key challenges:
Design of chatbots and natural language user interfaces is expanding the field of HCI. One key challenge of conversational user interfaces is “Seeing conversations as the object of design” (Følstad & Brandtzæg, 2017, p.41). Følstad and Brandtzæg suggest that we move away from viewing design as an *explanatory task* to viewing design as an *interpretational task*. This means that we need to step away from the task of explaining what is or is not available, and step into the task of understanding what the user needs and how she wants the services presented (Følstad & Brandtzæg, 2017). Another challenge is that the conversation breaks down fast (Følstad & Brandtzæg, 2017) and that it often gets overshadowed by errors (Luger & Sellen, 2016). Luger and Sellen also mention the lack of conceptual understanding in conversational user interfaces as a challenge (Luger & Sellen, 2016).

**Chatbots in relation to guideline 1 and 2 by Amershi et. al. (2019):**

G1: Make clear what the system can do
This guideline is relevant for all systems, including chatbots. A user would want to know what the chatbot can do in order to make sure it’s actually useful for the desired task that the user want to accomplish. It’s important that the user know

sooner rather than later if the system is actually relevant to the task at hand so the user don't get frustrated or end up with the feeling that the chatbot has wasted the user's time.

G2: Make clear how well the system can do what it can do.
If the user knows what the system can do (G1) then the system should also provide information about how well it can do it. This relates to the article by Kocielnik (2019) about setting the user's expectation. When the user is provided with information about how well the system works it may be more understanding to why the chatbot is not successful in providing the desired information or output.

# Module 3

**Collaboration and levels of automation**

Philips et al. (2016) argues that the general mental model of robots is not representative for real-world robots, because most of these mental models are developed through fictional media. The gap between the user's expectations of robots and the reality leads to distrust (Philips et. al., 2016). In a team with humans and animals there is an interaction hierarchy where the humans in general are have a higher degree of authority than animals. People have less experience interacting with robots than animals in general and therefore people know more about the animals ability to perform different tasks. Philips et al. (2016) want to transfer the mental model and insight of the hierarchy and abilities between teams of human and animals onto the teammate relationship between human and robot. Transfering or refining the mental model of the human-animal relationships so that it fits also with the human-robot relationship can help in forming a trusting relationship.

Endsley (2004) have created a list of different levels of automation for computers ranging from manual control, where the human performs all aspects of the tasks, to full automation where the computer performs all the aspects of the tasks. I will draw out two examples of human-robot collaboration presented in Philips et al (2016) and describe which level of automation I believe they fall into.

Animals has been used, and are still being used in some situations, to “multiply human physical capabilities” (Philips et al., 2016, p. 104). This teaming has been transferred to the field of robotics and can for instance be used for lifting heavy objects or people.

**RIBA II**

RIBA II is a robot nurse that can help human nurses with lifting patients out of their beds or chairs. RIBA II is designed to lift humans up to 61 kg (Salton, 2009). RIBA II fits into the Endsley’s (2004) fourth level of automation “Computer aids in doing each action as instructed” (Endsley, 2004, p. 185) because it needs instructions and guidance by a nurse in order to complete a single tasks (Ikinamo, 2011, 0:05).

One advantage of increasing RIBA II’s level of autonomy could be that it would be able to complete its specialized tasks without guidance or presence of a nurse. Increasing the autonomy could probably also lead to increased effectiveness because RIBA II moves very slowly. I imagine that increasing RIBA II’s level of autonomy even more, making it capable to detecting if people need help getting out of their beds or chairs, or help them up from the ground after a fall it could increase patients level of autonomy as well. This is a huge advantage because being hospitalised and unable to move without help from others is, in my belief, an unwanted situation for most people. One disadvantage with increasing the level of autonomy could be patient discomfort because they don’t trust the robot enough to carry them carefully without supervision by a human.

Decreasing the current level of autonomy would make RIBA II quite difficult to handle by the nurses because it’s already highly dependent of human interaction and guidance to complete tasks. I don’t believe that it would be any advantages with decreasing the level of autonomy any lower than at its current state.

**PARO**

Paro is an example of adapting human-animal teams in the field of robotics. Paro is a robot seal companion that's designed for elderly to alleviate depression (Philips et al, 2016). Paro has five sensors that makes it able to perceive its environment, learn to behave and respond as if it was alive (PARO Therapeutic Robot, nd). Paro is used for animal therapy within hospitals or care facilities.

Paro doesn't fit very well with the levels of automation defined by Philips et al (2016) because it's designed to be a companion rather than completing tasks. But if we view it's task as being a companion to elderly individuals I believe that it would fit into the level "Shared control". The reason I believe this is the best fit is because Paro is able to remember previous actions, which means that the human is in control of the decision making as to how Paro behaves given different stimuli by humans.

I imagine that an increase in the level of Paros autonomy would lead to Paro being able to walk around by itself, detect who's in need for it's companionship and behave differently with all patients based on their preferences. This could be an advantage when Paro is shared by some or many people in a healthcare facility because more people would be able to interact with Paro during the day. I imagine that Paro is an expensive robot, but if it could walk between different patients by its own then the hospitals could get one or two to share for a whole section. Paro is designed to alleviate depression (Philips et al, 2016), and one disadvantage with increasing the level of autonomy could be that patients might feel sad or more depressed if Paro were to leave them because it needed to company someone else.

If we were to decrease Paro's current level of autonomy I would imagine that it would not remember previous actions or the user's preferences. Paro would then become more like a toy that makes noises when touched by humans. One advantage with this decrease in autonomy could be that patients that's not able to take care of or interact with Paro at its current level, because it requires attention, could be able to handle it when it's less cognitive demanding. The disadvantage with this decrease could be that the patients don't connect with or view Paro like a companion and it could affect its therapeutic capabilities, not helping the patients heal their depression.

*References:*

Amershi et al. (2019) Guidelines for Human-AI Interaction

Bratteteig, T., & Verne, G. (2018, August). Does AI make PD obsolete?: exploring challenges from artificial intelligence to participatory design. In Proceedings of the 15th Participatory Design Conference: Short Papers, Situated Actions, Workshops and Tutorial-Volume 2 (p. 8). ACM.

Copeland, B.J. 2006. “Artificial Intelligence”. Encyclopedia Britannica. https://www.britannica.com/technology/artificial-intelligence/The-Turing-test (Accessed 27.09.2019)

Følstad, A., & Brandtzæg, P.B. (2007). Chatbots and the new world of HCI interactions

Google, nd. Advancing AI for everyone. https://ai.google (Accessed 27.09.2019)

Grudin, Jonathan. AI and HCI: Two Fields Divided by a Common Focus. AI magazine 30, no 4 (Accessed September 18, 2009)
Ikinamo, October 24th. 2011. RIBA II Care Support Robot For Lifting Patients #DigInfo [Video]. Retrieved from: https://www.youtube.com/watch?v=wOzw71j4b78

Jeff Salton, September 3. 2003. RIBA the friendly robot nurse. Retrieved November 15 from: https://newatlas.com/riba-robot-nurse/12693/

Kocielnik et al. (2019). Will you Accept an Imperfect AI? Exploring Designs for Adjusting End-user Expectations of AI Systems

Luger, E., & Sellen, A. (2016). “Like having a really bad PA”: The Gulf Between User Expectation and Experience of Conversational Agents

Moravec, Hans Peter. 2005. “Robot”. Encyclopedia Britannica. https://www.britannica.com/technology/robot-technology (Accessed 27.09.2019)

McCarthy, John, et al. "A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955." AI magazine 27.4 (2006): 12-12.

Markel, Howard. 2011. The Origin Of The Word Robot. https://www.sciencefriday.com/segments/the-origin-of-the-word-robot/?fbclid=IwAR0caNXIZpP05wSPDyDkP3CmADYo0I9xtbZx-AAHZ95uiNkz3dhF_b2pNOU (Accessed 27.09.2019)

Nilsson, N. J. (1998). Artificial Intelligence: A New Synthesis. Morgan Kaufmann

National Disability Authority. nd. What is Universal Design. http://universaldesign.ie/What-is-Universal-Design/ (Accessed 27.09.2019).

PARO Robots, nd. PARO Therapeutic Robot. Retrieved November 15 from: http://www.parorobots.com/

Russell Stuart, J., & Norvig, P. (2009). *Artificial intelligence: A modern approach*. Prentice Hall.

Thrun, S., 2004. Toward a Framework for Human-robot Interaction. Hum.-Comput. Interact. 19, 9–24

**Feedback iteration 1:**

My first iteration wasn’t complete when I delivered it, so in this iteration I have focused on completing the first iteration. In this iteration, module 2, I have also focused on making shorter and more understandable sentences.

**Feedback iteration 2:**

I got feedback that I should include some more references earlier on in the second module. So for the last iteration I have added some more references in the beginning of module 2. The positive feedback I got was about providing pictures and description of how the Google Assistant failed to complete the task I wanted it to do.

In my opinion getting feedback from a fellow student has been very helpful for rewriting and improving the “finished” modules. I also thought of the feedback I got on earlier modules when I wrote the last one.