

# IN5480 Individual assignment - Iteration 2

---

## Table of contents

Concepts, definition and history of interaction with AI	1
Robots and AI systems	4
Universal Design and AI systems	5
Characteristics of AI-infused systems	7
Human-AI interaction design	9
Chatbots/conversational user interfaces	10
References	11
Appendix	12

## Concepts, definition and history of interaction with AI

### **The history of AI**

The term “artificial intelligence” was first coined by John McCarthy in 1956 when he called for participation in a workshop (Grudin, 2009).

Despite McCarthy being the first person to use the term, people have been talking about AI before that. Alan Turing, recognized as a leading code breaker, mathematician and logician, wrote in the London Times in 1949 that he believed machines eventually would take over some of the fields normally covered by human intellect (Grudin, 2009). Before that again, In the 18th century Mary Shelly speculated about autonomous machines when writing Frankenstein.

After World War II several conferences were held about AI, gathering participants with backgrounds such as logicians, mathematicians and psychologists (Grudin, 2009). In the 60's the US government financed several projects in the field of AI during the moon race with the Soviet Union, enabling AI to become a field of its own, both in the US and later abroad (Grudin,

2009). One notable project from this era was the General Problem Solver (GPS), a program developed to simulate human problem-solving methods (Newell & Simon, 1961). In the 70's, the UK government withdrew funding for more AI research as "they did not find any major or even significant results from AI research" (Negnevitsky, 2005). This was during what Grudin (2009) called an "AI winter". The 80's saw a boom in expert systems and AI research, while during the 90's the field experienced another lull (Grudin, 2009). In our modern times AI is on an all time high, and a lot of work and funding is dedicated to research of robotics, machine learning and speech interaction, to mention a few areas.

### **Definitions of AI**

*It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable (McCarthy, 1998).*

This definition says something about how AI systems do not necessarily have to imitate human (or animals) intelligence. This definition is from McCarthy, the man that coined the term AI in 1956, but this definition is from 1998.

*"The study of how to make computers do things at which, at the moment, people are better" (Rich & Knight, 1991).*

Rich and Knight (1991) focus on the fact that Artificial Intelligence is developed with the goal to become more like human beings.

*"In computer science AI research is defined as the study of "intelligent agents": any device that perceives its environment and takes actions that maximize its chance of successfully achieving its goals. [1] Colloquially, the term "artificial intelligence" is applied when a machine mimics "cognitive" functions that humans associate with other human minds, such as "learning" and "problem solving" (Russel & Norvig, 2009).*

This definition describes AI as capable of mimicking certain ways humans think or perceive. It also defines AI as an 'agent' capable of acting on what it perceives, which requires a certain degree of autonomy. This quote is from a textbook used in computer science studies.

My own definition:

Artificial intelligence is a branch of computer science that focuses on making systems that are able to perform autonomous or semi-autonomous calculations and/or tasks, where these calculations or tasks traditionally are executed by human beings.

### **Company**

I want to talk about how Telenor, one of Norway's biggest mobile and web providers, present AI. Their webpage about AI stresses how the technology will change our lives and for the better. They continue talking about how their company support innovation in the AI field in Norway and the place they imagine AI will have in our future.

The text present AI in a very optimistic light, and as an idea or tool that can help and support creation of value in the future.

### **Film**

The film I will bring up is "The Hitchhiker's Guide to the Galaxy". There are several AI systems in this science-fiction comedy, among them the robot Marvin and Eddie, an AI system that runs a spaceship.

Marvin is a 'manically' depressed and very intelligent robot, something he likes to bring up. He is portrayed as very human-like in many respects, in the way he moves and the way he holds conversations. The humans around him talk to him as a human being, although the first time the main character Arthur sees him, he seems impressed and intrigued because of Marvin's capabilities. Most of the spaceship's crew treat Marvin as a helper or servant.

Eddie is an AI system that controls the spaceship. The people inside talk to him, much in the way one might talk to a Google Home or Alexa. The interaction starts by addressing his name, and then give him a command or task. He usually replies in a very casual and happy manner

before doing what he is told. Because the film was originally a radio-play from 1978, I find it interesting that Eddie resembles something like Google Home.

## Robots and AI systems

### **Robots**

According to an article from 'Wired' the word 'robot' first appeared in 1921, in Karel Capek's play R.U.R. (Rossum's Universal Robots) (Simon, 2018). The word 'robot' has roots in the Czech, and means forced labour (Simon, 2018).

Thrun (2004) talk about how the field of robotics is very broad, and explains that robots can be categorized in three categories based on their purpose or application - industrial robotics, professional service robotics and personal service robotics. Of these three, the first that came about is industrial robotics, robots that work in an industrial setting (Thrun, 2004). Industrial robots became available on the market in the 60's and have since become popular in several industries, in fact, when Thrun wrote his article in 2004 there where an estimate of 780 600 operational robots in use (Thrun, 2004).

### **Definitions of Robot**

*"From a technological perspective, robotics integrates ideas from information technology with physical embodiment. Robots share with many other physical devices, such as household appliances or cars, the fact that they "inhabit" the same physical spaces as people do in which they manipulate some of the very same objects" (Thrun, 2004).*

This definition underlines the fact that robots are physical artifacts, existing in the same spaces as humans. Later in the article, Thrun also explains that the biggest difference between household appliances and cars are that robots are to a certain degree autonomous. Note that this definition does not mention artificial intelligence.

*"A robot is an intelligent, physically embodied machine. A robot can perform tasks autonomously. And a robot can sense and manipulate its environment." (Simon, 2018)*

This definition emphasizes the autonomy of robots, and assumes they have a certain type of intelligence that they use to determine how they operate. This quote is from a magazine called “Wired” and was placed in an infobox on the side of the article.

## **Robots and AI**

I understand AI as an umbrella term where robotics can be a *type* of AI. Robots might have many qualities that are common in AI programs, for example autonomy, facial or voice recognition and so on. On that note, there are also examples of robots that do not have much to do with AI, for example some industrial robots, as mentioned in Thrun’s (2004) article. An assembly robot cannot be said to be artificially intelligent if it only makes the same motions over and over again, without perceiving or changing its behaviour based on its surroundings.

The main difference between AI and robots with AI capabilities, as I see it, is the physical embodiment that the robot is. Both of the definitions I have brought up earlier defined robots as ‘embodied’. AI might be used in a program that is not physical, Siri on Iphone is an example of this. When we put AI in a physical artifact, it becomes a robot.

## **An example**

The robot AV1 from No Isolation (2018) is a robot made to enable children to take part in lessons in school when they cannot be there, for example because of long-term illness. AV1 enables the children who use it to remotely communicate with their classmates, stream the lesson on their phone and show emotions they are feeling through the robot that is placed in the classroom. The robot is controlled by the child using their phone or tablet from home. AV1 does not move much, but is able to rotate the headpiece when the child controlling it swipes their screen.

Children use the robot by navigating a menu through an app. The options available include sound-control, ‘raise hand’, which gives the robot a blue light around the head, and several more.

## Universal Design and AI systems

### **Universal design**

*“Universell utforming handlar om å utforme omgjevnadane slik at vi tek omsyn til variasjonen i funksjonsevne hos innbyggjarane, inkludert personar med nedsett funksjonsevne. Når du lagar noko som er universelt utforma, når du alle målgruppene gjennom éi og same løysing.” (Difi, 2019)*

This definition is from Difi, the directory with power to warn and punish in case of breach on universal design rules in Norway. The definition does not mention technology, but “surroundings”. Despite this, digital systems is without a doubt also required to be designed in a way as to not exclude people with disabilities or other groups of people. The definition also adds that one service should not have several solutions to achieve the same goals, on the basis of who uses them.

I understand universal design in respect to inclusion as the act of making solutions that are possible to use for as many as possible. By designing for the whole population, we democratize the solution by not restricting use or information to one or some groups of people, and we enable them to take part in our service, product or whatever we are making.

### **AI og human perception**

AI is often implemented in systems that one might use in the home or at work. By appealing to what is familiar, mimicking human perception, movement or feelings, AI systems might be more approachable and easier to use by humans with little experience in similar systems.

Schulz et al. (2018) wrote that animation can improve how people interact with robots, and used animation techniques to give the robots style in their research paper. Making robots appear to have personality might give the users interacting with them a feeling they are more predictable and easier to understand (Schulz et al, 2018).

I think the same principle can be applied for a chatbot for example. If a chatbot is made to simulate a certain type of person, for example a human working in customer service, it might be easier for users of customer support to relate or communicate with the chatbot than if it was made to simulate a parrot or something else unrelated.

### **Inclusion and exclusion**

It is quite possible to exclude people in AI systems, which is why it is of great importance to think about design choices in making them. Verne and Bratteteig (2018) pointed out in their paper that AI is unpredictable, and it can be difficult to foresee what AI systems might 'learn' with machine learning. Shein (2018) quote in her paper that machines have no common sense, so if a person inputs errors, the machine will continue to apply them. The data fed into a system might be biased or faulty, and this might lead to exclusion of some groups of people. There are several examples of facial recognition that has problems recognizing faces of women or people of non-white ethnicity, probably because these groups were not heavily involved in developing or testing of the software. Similar issues have been raised in predicting policing systems, where data that have been fed to the system was biased, and therefore targeted certain groups or neighbourhoods (Shein, 2018). In addition to this, some AI technologies use a different way of interaction, for example by voice or movement, and this might exclude, or in some cases, include people with different types of disabilities.

On the other hand, AI has great potential to include people. Not only through different ways to interact with the systems, but also because many AI technologies are made to be used for assistance, in the house, as entertainment or for doing tasks that might be difficult for some or many people to do (Schulz et al, 2018).

To ensure that AI include rather than exclude people in the future, Shein (2018) raises arguments that what is being used to train machine learning programs should be transparent, and that it is important that users of AI systems understand the limitations of their programs, and not trust them blindly.

## Characteristics of AI-infused systems

### **Key characteristics of AI-infused systems**

#### Learning

The first characteristic of AI-infused systems are the ability to learn. As well as showing how the system is evolving, the system should make it clear what it can do, and guide users in how they can use it. It should also show how well it can do what it does.

#### Improving

AI-infused systems should be designed for the times it makes mistakes because they are going to happen. The system should support correction to help the system work better, and when it is in doubt, the system should express this. Kocielnik et al (2019) describe that in their system it was easier for a user to recover from a False Positive rather than a False Negative, and it should therefore be designed after this.

#### Black box

Sometimes it can be difficult to understand why a certain output was generated, and it is therefore important for AI-infused systems to show why the system did what it did.

#### Fuelled by large data sets

To improve AI-infused systems they should be designed for data capture. This can help the system in evolving and personalize the service provided to the users. The data should be private, and it should benefit the users not be taken for other purposes.

#### **An example**

An AI-infused system I have used is the photo gallery that comes with Iphone. The gallery groups up the photos of people it recognizes as the same people, making a small album with all the photos where that person shows up. The application makes a little profile of the different people in my gallery, where it is shown who shows up in photos together with that person. It also shows the places where the photos of them where taken, as well as the albums where they are present, for example from holidays or trips where a lot of photos where taken.

The gallery is learning because when I first got this application the phone defined my face as 4 different people. At the moment, there are two profiles on my face, one where the photos of me look normal, and another profile where I make silly faces, and it therefore thinks I am someone else. It also gives options to the profiles of people in my gallery so that I can help when the application grouped up people wrong.

The gallery supports improvement, because it is possible to delete profiles or add photos. However, it is not possible to merge two profiles together in an easy way. The system is probably using large data sets to be able to perform the face recognition. I have seen that it has become more accurate over the last years, and I know it's not because of my own photos,

because a lot of the people in my gallery are people I knew four years ago when I lived abroad. This makes me think the application is fueled by large data sets from other users as well.

## Human-AI interaction design

### **Main points from Amershi et al. (2019) and Kocielnik et al. (2019)**

Amershi et al. (2019) present 18 design guidelines when designing applications for human-AI interaction. They state that these guidelines can be used as a resource when creating and researching human-AI interaction design (Amershi et al, 2019).

Kocielnik et al (2019) describe a project where they have used a scheduling assistant to study the impact of several methods of expectation setting. In the study they set the AI to have 50% accuracy, but in the first scenario it gives a False Positive, and in the second a False Negative (Kocielnik et al 2019). Even though the results were equally “accurate” the participants reacted differently to the two types of results, with different perceptions of accuracy and acceptance (Kocielnik et al 2019). Lastly, Kocielnik et al (2019) present design expectation adjustment techniques that prepare users for imperfections in AI systems.

### **Design guidelines**

G3: Time services based on context (Amershi et. al, 2019)

The gallery will send me a notification when it generates a new album based on the places and people in the photo collection. The notifications are unobtrusive and do not seem to happen when I am currently on my phone, but while it is locked. I am not sure however if this is a coincidence or not, but it has never happened to me the 4 or 5 years I have had this feature.

G9: Support efficient correction (Amershi et. al, 2019)

I do not think the gallery supports correction in a large degree. It is possible to merge two profiles together, but there are a lot of photos that have not ended up in anyone’s profile on my phone, and these photos do not show up in this part of the application making it very tedious to assign them. If there is a profile with several different people’s faces, it is not possible to add these photos into different profiles, they are attached only where the application assigned them. I think that if there was an easier solution to do this, I would have more use of my gallery, as it currently makes some mistakes.

## Chatbots/conversational user interfaces

### **Key challenges in the design of chatbots or conversational agents**

Følstad & Brandtzaeg (2017) bring up several challenges when designing chatbots or conversational agents. One of them is the transition of designing traditional interfaces versus designing conversations (Følstad & Brandtzaeg, 2017). The user interface will be blank, and the user will have much more control over how they navigate the system, but it will also be less obvious what the system can do because the options are not visible (Følstad & Brandtzaeg, 2017). Another challenge that is brought up is the fact that new technologies often create new technical divides and bias, and chatbots and conversational agents are not immune to this (Følstad & Brandtzaeg, 2017). It is important that chatbots and conversational agents should be usable for all ages, gender and societal status (Følstad & Brandtzaeg, 2017). This can pose a challenge as chatbots and conversational agents often are set up in a one-size-fits-all approach, and uses data from one dataset regardless of the user's preferences or requirements (Følstad & Brandtzaeg, 2017).

Luger & Sellen (2016) describe some challenges in their paper where they interviewed several participants about the use of their CA's (conversational agents). The challenges they found were that the users had a mental models of how their CA's worked because of poor feedback from the system about what it could do and how it did it (Luger & Sellen, 2016). The paper describes several ways to create a better user experience, for example that it should be possible to reveal the system intelligence and the system feedback should be redone to fit with the dominant use cases (Luger & Sellen, 2016).

### **Resolving challenges**

G1: Make clear what the system can do

I think a lot of conversational agents and chatbots have this problem, as the possibilities of the system can be obscure. It can be hard to find features one is not looking for, or know they exist without being told first. I think many challenges that were mentioned earlier would be addressed if more conversational agents managed to implement this feature. For example, the participants that were interviewed in Luger & Sellen (2016) would have a better user experience if this guideline was implemented in a larger degree.

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

Luger & Sellen (2016) found that deviation from this principle was a problem among the participants they interviewed. To be able to use a technology efficiently, it is important to know how it works, and this is a challenge for CA's as they do not rely very much on visualization. I think that if CA's were to implement this guideline in a good way, it would be easier to use for many people, as they would know more of what to expect. In this way, one can interpret the output with the knowledge of how accurate they are.

## References

Amershi, S., Weld, D., Vorvoreanu, M., Fournery, A., Nushi, B., Collisson, P., ... & Teevan, J. (2019). Guidelines for human-AI interaction. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (paper no. 3). ACM.

Følstad, A., & Brandtzæg, P. B. (2017). Chatbots and the new world of HCI. *interactions*, 24(4), 38-42.

Luger, E., & Sellen, A. (2016). Like having a really bad PA: the gulf between user expectation and experience of conversational agents. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (pp. 5286-5297). ACM.

Kocielnik, R., Amershi, S., & Bennett, P. N. (2019). Will You Accept an Imperfect AI?: Exploring Designs for Adjusting End-user Expectations of AI Systems. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (paper no. 411). ACM.

McCarthy, J. (1998). What is artificial intelligence?.

Newell, A., & Simon, H. (1961). Computer simulation of human thinking. *Science* (New York, N.Y.), 134(3495), 2011-2017.

Negnevitsky, M. (2005). *Artificial intelligence: a guide to intelligent systems*. Pearson education.

No Isolation. (2018, April 26). The impact of AV1 on children with long-term illness and absence. Retrieved from [www.noisolation.com/global/research/the-impact-of-av1-on-children-with-long-term-illness-and-absence/](http://www.noisolation.com/global/research/the-impact-of-av1-on-children-with-long-term-illness-and-absence/)

Rich, E., & Knight, K. (1991). Artificial intelligence (2nd ed.). New York: McGraw-Hill.

Russel, S. J., & Norvig, P. (2009). Artificial Intelligence. A Modern Approach (3rd ed.). Prentice Hall

Schulz, T., Herstad, J., & Torresen, J. (2018). Classifying Human and Robot Movement at Home and Implementing Robot Movement Using the Slow In, Slow Out Animation Principle. International Journal on Advances in Intelligent Systems, 11, 234–244.

Shein, E. (2018). The dangers of automating social programs. Communications of the ACM, 61(10), 17-19.

Simon, Matt. (2018, May 17). The Wired guide to robots. Retrieved from <https://www.wired.com/story/wired-guide-to-robots/>

Telenor (18. august 2019 kl. 20:39) Den viktigste teknologien i det 21. Århundre. Telenor. Retrieved from: <https://www.telenor.no/om/teknologi-norge/kunstig-intelligens.jsp>

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

Verne, G, Bratteteig, 2018, Does AI make PD obsolete?; exploring challenges from Artificial Intelligence to Participatory design,

## Appendix

### **Changes made based on feedback**

Based on my feedback I added some explanation of where my quotes are from and the context they were mentioned. I was also asked to put in how they came about, but I was unable to find that in most of the cases, unfortunately.