IN5480 – Specialization in Research in Design of IT
All modules

Module 1 – Concepts, definition and history of AI
1.1 Definition and history of AI

1.1.1 The history of AI

The term 'Artificial Intelligence' was formally founded at a Dartmouth College, New Hampshire, at a conference regarding the 'Dartmouth Summer Research Project on Artificial Intelligence' in 1956 (Anyoha 2017), (Grudin 2009, p. 49). But the story of AI begins longer ago than one would think. The earliest influences can be traced back to Greek philosophy about 800 B.C. – were some of the philosophers had ideas about inanimate objects 'coming to life'. The Greek writer Homer (author of the Illiad and the Odyssey) wrote about mechanical 'tripods' (a small table or stool with three legs) who were imagined 'mechanical assistants' able to 'enter the gathering of the Gods' – with their 'golden wheels' (Kraggerud 2019), (Buchanan 2006, p. 53). These and similar thoughts where one gives human characteristics to non-humans such as objects, animals, fantasy figures etc. is called 'Antomorphism', is something that influenced both science, religion, and human imagination for decades to come. Examples of 'Antomorphism' can be found in fantasy stories like Wizard of Oz, Beauty and the Beast, Frankenstein and Star Wars - but also real-world AI innovations like Apple's 'Siri', Boston Dynamics 'Spot' or Japan's 'Weird Hotel' staffed with robots (Løøv, 2019), (Buchanan, 2006, p. 53-54).

Modern day AI is a field influenced by many different disciplines; engineering (cybernetics), biology (neural networks), psychology, communication theory, game theory, mathematics, statistics, logics, linguistics and like mentioned before – philosophy (Buchanan, 2006, p. 54-55). But one of the most pivotal events in the history of AI was the code breaking machine 'Bombe' – designed by Alan Turing (amongst others) in 1939. Turing developed the first systematic method for breaking encrypted messages, which successfully decrypted German messages in World War II (Copeland, 2021). It is said that this effort shortened the war by several years (Goodwin and Thormodsæter, 2019). Further, Turing wrote a paper in 1950 about the possibility of an electronic computer who could behave intelligently. The paper also included some criteria for differentiating between an intelligent machine and an unintelligent machine – known as the 'Turing Test' (or 'Imitation Game'). The test is; if a person is communicating with a computer or another person just by using a keyboard and a screen and he/she cannot decide if they are communicating with a machine or not – the machine is in fact 'intelligent' (Goodwin and Thormodsæter, 2019).

Despite the 1950s innovations, AI does not become a major research field until Mid-1960s to the Mid-1970s. AI pioneer Herbert Simon (one of the founding fathers of AI) wrote in 1960, 'Machines will be capable, within twenty years of doing any work that a man can do" (Simon in Grudin, 2009, p. 50). Even today, with machines more powerful an able than one might could have imagined at the time, we have yet to develop AI that can do 'any work' that a human can do. Still, in the last twenty years our modern computers have had enough memory and processor power to be able to test some of the grand ideas proposed in the 20th century (Buchanan, 2006, p. 56). The 21st century is so far known for multiple AI achievements – for instance AI technology being used to remotely control robots (The Mars Quests), develop self-driving cars (Tesla Autopilot), be used as virtual assistants/chatbots, in the health care sector, security and surveillance sector and much more.

1.1.2. Definitions of AI

There are many definitions of AI, and I have selected three different definitions to present in this section. One from the author/AI expert Lasse Rouhiainen, another one from the computer/cognitive scientist John McCarthy, and one from the Oxford Dictionary.

Rouhiainen defines AI as; '[…] using computers **to do things that normally require human intelligence"** (Rouhianen 2019, p. 3). McCarty's definition of AI is:

'*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 in IBM, 2020).

The Oxford definition is; *'(AI is) The theory and development of **computers systems able to perform tasks normally requiring human intelligence**, such as visual perception, speech recognition, decision-making, and translation between languages'* (Oxford University Press, 2021).

It is clear that McCarty emphasizes something different than Rouhiainen, by this I mean that Rouhiainen's definition is computers **doing things that requires human intelligence**, and McCarthy's definition is **using computers to understand human intelligence**. For me I would say that the first is more action-focused than the latter. McCarthy also states that AI does not have to **confine itself** in relation to things **that are biologically observable**. In other words – AI computers can go beyond the scope of human biology (even human intelligence). I would also argue that Rouhiainen's definition of **'doing things'** is quite broad – and gives the reader the impression that AI has a wide variety of possibilities, which we know is true. At the same time, I start to wonder; what aspect of human intelligence does Rouhiainen mean? Is it the ability to speak, to read, to make sense of the world or being able to communicate emotions? Is it the ability to understand reality in form of past, present and the future or to be able to solve complex problems within a variety of contexts? Maybe all the above? From my point of view, I am leaning a bit towards the Oxford definition by claiming; **AI is developing computer systems that are able to perform a variety of human tasks and imitate human behavior and intelligence.** I say 'a variety of tasks' and imitate' because my belief is that machines (however 'seamlessly' the technology) will never fully understand human intelligence. Compared to the other definitions. who comes from a cognitive and mathematical/logical perspective, I put more emphasis on the social, biological and contextual perspective, stating the difference between what is artificial, and what is biological.

1.1.3 Review of the article "AI and HCI: Two Fields Divided by a Common Focus"

The article mentioned in the title by Jonathan Grudin tries to outline the history of AI and HCI, and the tension between the two fields. The main focus is to give the reader an overview of the background and the forces which historically kept AI and HCI apart – but also explain the current situation of 'Usable AI' (in other words; human-robot interaction). My main objection with this article is that it seems to be lacking some in-depth context (being that it is a bit short), thus being a bit fragmented. I was also left with some questions after reading this sentence; 'conferences are increasingly populated by students, who bring wonderful energy but uncertain commitment. Many prominent HCI researchers have moved from computer science departments to information schools, introducing new identity issues and distancing them from AI colleagues' (Grudin 2009, s. 55). Why is there a lack of commitment, and why is there a distance from AI colleagues? In my view, HCI and AI can learn much from each other's domain, and seems like a 'perfect partnership' on paper, especially when it comes to solve complex world issues like AI-ethics, participatory AI, health-care robotics, and the threat of unemployment (due to AI and robots).

1.1.4 **'Emotech' - a contemporary company that work with AI**

'Emotech' brands itself with the slogan 'Reimagining the relationship between humans and technology' and is (in their words) 'emotion technology'. The company makes multi-modal home solutions, which they claim makes a new way of interaction (Emotech 2021). The company talks about AI as a series of service products and frequently uses the term 'proactive interaction' – and sells products that do text to video, event- and behavior detection, stress-, emotion- and speech analysis/recognition etc. They can also provide tailored solutions for both homes and offices. It seems that their 'proactive interaction' is an idea they hope would change how humans interact with AI. Olly, branded as 'the first home robot with personality' anticipates the user's needs, and won several CES (Consumer Technology Association) Innovation Awards in 2017 (Emotech 2017). Olly uses a 'character engine' to create an unique personality, so that it will relate to the user in a distinctive way.

1.1.5 **The fictional film "I, Robot" starring Will Smith**

This is an American Sci-fi movie from 2004. The plot is set to 2035 in a dystopian world, where the main character Del Spooner (Will Smith) investigates the alleged suicide of Alfred Lanning – the founder of a company called U.S. Robotics. US Robotics has successfully filled the world with highly intelligent humanoid robots - which all have the same appearance and has an assisting role for humans. The movie is based on some elements from the novel 'I, Robot' written in 1950 by Isaac Asimov. The unique feature of Asimov's book series about robots are 'The Three Laws of Robotics' – rules that ensures that the robots does not turn against their creators (Øverland 2019). These rules sometimes result in ethical dilemmas, for instance the car accident where a robot decides to save Spooner instead of a little girl (which is why Spooner does not like robots). Spooner suspects that a humanoid robot called 'Sonny' killed Lanning and somehow was able to break (hack) one of The Three Laws in his code. The interaction between robots and humans in this movie mostly being done by voice, tangible interaction - but also through behavior- and emotion recognition. These robots manipulate the same objects as humans and seem very 'evolved' in terms of technology and human 'embodiment'. Still, there is a significant difference between the robots and humans as their speech, manor and looks make them easy to differentiate from humans.

1.2 Robots and AI systems

1.2.1 **The word 'Robot'**

The word 'robot' was first introduced as the word 'robota' in the Czech play 'Rossum's Universal Robots' (R.U.R.) written by Karel Čapek in 1921. The meaning of the word is 'mandatory work', so the robots in the play served humans by doing their jobs. The robots in this play did not have feelings or intelligent life, but after some time they revolt and destroy humankind (Wallén 2008, p. 4), (Blekastad, 2020). However, the term 'robotics' (and the Three Laws of Robots mentioned in the previous section) were introduced by Isaac Asimov (Wallén 2008, p. 5).

1.2.2 **‘Robot’ definitions**

I have selected two different ‘robot’ definitions to present in this section. One from the The Robot Institute of America and another from the Merriam Webster’s collegiate dictionary. The Robot Institute of America defines a robot as: ‘*A reprogrammable, multifunctional manipulator designed to move materials, parts, tools, or specialized devices through various programmed motions for the performance of a variety of tasks’* (The Robot Institute of America in Thrun, 2004, p. 3), and Merriam Webster’s dictionary defines it as; ‘*An automatic device that performs functions normally ascribed to humans or a machine in the form of a human’* (Merriam Webster’s collegiate dictionary (1993) in Thrun, 2004, p. 3). I would say that both definitions make it clear that we are talking about a physical device (‘manipulator’) that can do tasks. This separates it from being a mere computer program, but I still think that the Robot Institute focuses more on the program and the mechanics behind the ‘motions’ than Webster’s – who talks about the human aspect of it (the robots so-called imitating human behavior). In my eyes, a robot is a programmed device that can perform functions normally ascribed humans and/or other machines by using their own built-in program and/or continually reprogramming their own system through learning. I think that the ‘learning’ aspect is important to incorporate as more and more of our systems are able to learn new skills through learning (making them even more ‘like us’).

1.2.3 **The relation between AI and robots**

The relation between AI and robots is that robotics involves building mechanical devices, and AI is the program that can be used to make the robot ‘intelligent’. If I were to compare it to something known I would say that the robot is the ‘body’ and the AI is the ‘brain’.

1.2.4 **The iRobot Roomba**

Roomba is a robot vacuum cleaner. The company markets it by stating ‘[…] your Roomba robot vacuum will make sure you come back to a clean home. Just the way you like it’ (iRobot, 2021). In other words, it moves from room to room, floor to floor – vacuuming your dust. Humans mostly interact with Roomba the first time it is set up or when they empty it for dust. In the set-up process the owner chooses the settings in the accompanying app (which rooms it should clean, at what time, how many times a day etc.) before Roomba begins it’s initial ‘room scan’ which will become the default pattern it follows for every vacuum session – unless chosen otherwise by the owner. The robot moves slowly in a pattern movement, only vacuuming places it can access with its round shape. After the session is done, the robot returns to it’s designated charging place to ‘rest’ between vacuum sessions. I would say that the ‘normal’ interaction between a Roomba and its owner is nearly non-existent if used like intended – e.g., the owner relaxing in the couch or making dinner while Roomba is cleaning the house automatically.

1.3 Universal Design and AI systems

1.3.1 **Universal Design**

The Norwegian Digitalization Agency defines Universal Design as this; ‘Universal design is about providing good technological solutions for everyone’ (The Norwegian Digitalization Agency, 2021). The definition specifically states, ‘good technological solutions **for everyone’**, which involves making solutions that is accessible to all people – regardless of their physical and/or cognitive ability. Personally, I think that Universal Design should strive to be more than accessible – it should also be *used* and *understood* – have a real-life value for the people that uses it. Only when the people of representation can use the design seamlessly, I would say that it is inclusive.

1.3.2 AI and Universal Design

The potential of AI with respect for human perception, movement, cognition and emotions is in my opinion huge. The question is, how can we make sure that AI is including? There are some known examples of more inclusive innovations like speech recognition technology – for instance known from voice-assistants like 'Siri' and 'Alexa'. Auto-generated captions on YouTube-videos or real-time transcriptions of conversations using 'Google Live Transcribe' are also some other known examples (Cahalane 2019). The first two can be helpful for people with visual impairment, and the other two can be used by people with hearing loss and/or in situations where the context makes it is hard to hear or understand sounds. Still, there is a big 'but' here – because the potential of AI becoming including or excluding depends on *who* trains the AI and *how* it is trained. The more diversity is represented, and more diverse datasets are being used – the better. This is very important since recent reports has shown that AI can become biased, even racist, if the team who builds it aren't aware of ethics, what data they use to train it, or are transparent about how the decisions are being made (Borgan 2019). The thing is, there is a difference between lab-settings, and real-life and observed effects that can have many different causes (more about this in the next part) (Heaven, 2020).

1.3.3 The concept of 'Understanding'

In the WCAG 2.1 principles and in many of the Human AI-Interaction guidelines the concept 'understand' and "understanding" is used. So, what does it mean to 'understand' and what is 'understanding' really? According to Guideline 3.1 in WCAG 2.1, 'the information and operations of user interface must be understandable', meaning; the text content and the operations must be readable and understandable. Then, when it comes to machines – do they understand? The answer is both yes and no. Yes, in terms of that they can read program code and execute operations, but also no because they do not understand 'context' or cause-and-effect like we humans do. To exemplify this, there is a difference between correlation (if this happens, then this happens, e.g., if I pull down this handle, the door will open) and causality (one thing affects something else, e.g., if I pull down this handle, an alarm might go off). In other words, humans can use knowledge they've learned in one context, in another context, using their intuition. What I mention here, is defined as the 'frame problem' or 'alignment problem' by John McCarthy (Goodwin and Thormodsæter 2021).

1.4 Guideline for Human-AI Interaction

1.4.1 Human-AI interaction Guidelines

One of the 18 guidelines for human-AI interaction from Microsoft states this; 13. 'Learn from user behavior. Personalize the user's experience by learning from their actions over time.' If I were to describe this guideline, I would say; Learn the user's actions continually over time and reuse this information to make the user experience more accurate and valuable for the specific user. I use the word 'continually' here because humans are both static and dynamic. Our habits might be static, and never change – but we can also change them or gain new knowledge throughout life.

1.4.2 HCI design guidelines

Some of the HCI design guidelines have similarities with the Human-AI Interaction Design Guidelines (IDC from here) – but the main difference in my view, is that the Human-AI IDC has their guidelines inside categories for the different steps in the interaction. For example, they have a category called 'initially', then 'during interaction', 'when wrong' and 'over time'. Compared to the HCI design guidelines, these are more 'general'. Consequently, most of them apply to all parts

of the design, the 'entire time'. For example, 'strive for consistency', 'cater to universal usability', 'reduce short term memory load'. At the same time the HCI guidelines also have some principles that apply when something that halts the interaction (like the principles under 'when wrong' for Human-AI IDC); 'offer informative feedback', 'design dialogs to yield closure'. Personally, I favor the Human-AI IDC because they appear more detailed, with more specific examples than for instance 'cater to universal usability'.

Module 2: AI-infused systems, Human-AI Interaction Design, Chatbots

1. Characteristics of AI-infused systems

According to Amershi et al. (2019), 'AI-infused systems are systems that have features harnessing AI capabilities that are directly exposed to the end user' (Amershi et al., 2019, p. 1). More specifically, we are talking about systems such as intelligent assistants, navigation systems, e-commerce websites and social networks (like 'Siri', chatbots, Google Maps etc.) (p. 5).

AI technology uses algorithms such as natural language understanding, behaviour prediction or object recognition to give the user recommendations (Kocielnick et al., 2019, p. 2). In other words, these systems gain knowledge using large datasets about users, and are therefore characterized as 'dynamic' or 'improving' systems. Since the knowledge is based on a huge amount of content – some information is filtered out for the end-users. So, information about how they operate and/or contextual information can be hidden, making them unprecise and error prone or 'opaque' if you may (Følstad, 2021, presentation 22-23). The learning aspect also means that the systems are impacted by user's actions as they make recommendations based on earlier interactions. Consequently, Kocielnik et al., (2019), describes AI-infused systems as 'probalistic' and that they have 'transparency issues' (Kocielnik et al., 2019 in Følstad, 2021, presentation p. 24). The transparency issue is a known debate when it comes to AI systems, since many users might not be aware of how their information is collected, processed, and presented.

Yang et al., (2020), characterizes AI-infused systems as 'computational systems that interpret external data, learn from such data, and use those learnings to achieve specific goals and tasks through flexible adaptation' (p. 5). It should be noted that Yang et al., (2020) focus on the technicalities of AI systems for the research in their paper. Even though they admit that their definition does not specify 'what counts as "learning"', they empathize the learning aspect, stating that these systems learn to achieve certain tasks and goals (Yang et al., 2020, p. 5). Based on module two's papers, I would argue that the main characteristics of AI-infused systems are evolving, inconsistent, user-affected, and fallible. It is also important to state that we are specifically talking about systems which are a mix of non-learning and learning (AI) software.

1.1 Instagram

The social media platform Instagram process user-generated content (photos, videos, likes) to present recommendations to its 1,4 billion users (Business of Apps 2021). I presume that the platform uses object- and behavior recognition to process ('interpret') and use this information to achieve specific goals – getting the user to stay on the platform, interact with other user and generate more content (Yang et al., 2020, p. 5). In other words, it uses AI capabilities that are directly exposed to the end user (Amershi et al., 2019, p. 1). Like the characteristics mentioned in the last section, I think it is fair to say that Instagram is a dynamic and user-affected AI-infused system with mixes non-automatic and automatic software.

This (of course) does not come without its challenges, as it is also a platform that has received critique for its transparency issues. Since the main content of the application are pictures, and the most popular pictures are shown (forwarded) by the algorithm – a lot of information are filtered out. The users do not get an explanation as to why certain pictures or accounts are recommended (explicitly), but it is not difficult to see that the accounts with a big following and many likes are more often recommended. I will not get into the discussion about how this digital culture impacts the users mentally and socially, but I would argue that the transparency issue affects their ability to understand.

Research shows that young people struggle to grasp the commercial mechanisms and interests behind the 'idol' accounts they follow. When they see advertising, they are quick to reveal the rhetoric, but they are naïve when it comes to the intentions of the sender. It should also be noted that they claim that they are 'not affected by Instagram' (Elvebakk et al., 2018 in Aamli 2018). If the AI-capabilities of Instagram were presented – the goals, how they work, what they do, why it is uncertain – the users might make different choices, which in turn might change other challenges such as negative body image, mental health issues etc.

2. Human-AI interaction

The main take-away from Amershi et al. (2019), is that the paper presents 18 general design guidelines for human AI-interaction. The guidelines are a result of user testing proposed guidelines on 20 popular AI-infused systems. Consequently, they can serve as a resource for more success when it comes to human-AI interaction (p. 1-2). This paper also explains why these guidelines are needed – arguing that increased pattern recognition and automated inferences creates unpredicted behaviour, uncertainty and the possibility of violating established usability guidelines such as consistency, error prevention and understanding (p. 1-3). Personally, the biggest take-away for me is how this paper exemplifies the complexity of AI and how difficult it is to design systems which can take the technical, ethical, moral and human aspects into account. As Amershi et al. (2019) states: 'AI technologies […] can create unforeseen consequences and actions at odds with user goals and expectations' (Amershi et al., 2019, p. 1).

The main take-aways from Kocielnik et al. (2019) is that they explore techniques for shaping end-user expectations by studying how shaping impacts user acceptance (Kocielnik et al., 2019, p. 2). The researchers used two versions of a Scheduling Assistant (an AI system for automated meeting request detection in an email) to study how the users' perceptions of accuracy and acceptance were impacted by the different types of AI imperfections (also called 'False Positives' vs. False Negatives') (p. 1). By using three 'setting expectation' techniques (accuracy indicator, examples-based explanation, performance control), Kocielnik et al., (2019) demonstrated their ability to keep user satisfaction and acceptance in an inconsistent Scheduling Assistant (p. 2). For me, the biggest take-away is how knowledge about standard usability design principles (e.g., 'strive for simplicity', 'stay consistent') should be used in combination with universal design principles *and* Human-AI interaction design principles to make future designs a success.

2.1 Two Design Guidelines

How Instagram adheres to or deviates from the following to design guidelines by Amershi et al., (2019, p. 3):

	Guideline	Adheres and why	Deviates and why
G5	**Match relevant social norms** – Ensure the experience is delivered in a way that users would expect, given their social and cultural context	Depends on how and where the platform is used – does the user actively search for similar accounts, use the feed and/or both? Do they post pictures themselves, comment, like or are they 'silent' watchers? What is their context?	Depends on how the platform is used – does the user actively search for similar accounts, use the feed and/or both? Do they post pictures themselves, comment, like or are they 'silent' watchers? What is their context?
G8	**Support efficient dismissal** – Make it easy to dismiss or ignore undesired AI system's services when needed.	-	The users do not control the feed directly. They might be able to change what is recommended over time by cleaning up their 'follows' and by changing what they like, but in my opinion, this is not sufficient.

I think that when it comes to guidelines it is difficult to answer plainly yes or no. Guideline G5 is both a yes and no question because your feed will most lightly match your relevant social norms (getting recommendation based on your interest, demography, previous likes and so forth) – at the same time we must remember that Instagram is a western platform, meaning that western social and cultural norms (western lens) might be the majority of the content even though they have users across the world. Hopefully, other cultures and norms might be better represented in the future. G8, in my opinion, could be improved if Instagram offered an easy accessed function to dismiss or ignore their AI-system services. One way they could do it is to add it to their 'settings' with a listing of features the user can 'toggle' off. Alternatively, they could mark/inform the users when a specific feature is being used. That way the user can read about the AI specification, and maybe be offered a way to turn it off.

2.2 Summary of Bender et al. (2021) article

Bender et al., (2021) argues that deep learning systems have problematic aspects when it comes to environmental- and financial issues, as well as textual content. This needs to be addressed, especially since training just one of these big language models is estimated to require the same energy as a trans-American flight (p. 611-612). First and foremost, they describe some of the risks, for instance that having a language model of a large size does not equal diversity. This is because internet access, gender and age representation are not evenly distributed. Also, a small set of subpopulations share their thoughts and develops platforms or forums that present their worldviews – possibly creating an echo chamber (p. 613). When you have too much data, it is difficult to both understand and filter out what is in the data, how to categorize it, process it and so forth. The results of this are 'encode bias', discrimination and errors in the textual content. Furthermore, datasets that are too large, cannot be thoroughly documented (p. 1). Bender et al., (2021) also discuss how humans tend to give meaning to text. This combined with language models' ability to learn patterns, leads to risks of real-world harm (p. 618). For instance, harmful and extremist ideologies can mislead the general public.

The solutions suggested in the paper are; the use of careful data collection practises (planning), careful selection of datasets (those suited for the tasks), to consider the financial and environmental costs up front (reporting) and provide thorough documentation (p. 617-618). Another important note from the paper is shifting the focus from 'making models' to learn about how machines achieve goals and in which way they take part in socio-technical systems (p. 618). In other words, recommending that the researchers and builders of the language models become more proactive and shift their mindset. After all, both humans and the models have limitations, and we must be aware of them to be able to do something about it.

3. Chatbots / Conversational User Interfaces

Since the slides from the lectures was not published before the delivery date of this assignment, I will use my own reflection and other resources to write about some of the key challenges in the design of chatbots. Chatbots are machines, but they try to mimic human conversations often through a graphical user interface (Luger and Sellen, 2016, p. 2). The key word here is *conversations,* a complex human phenomenon which is difficult to adapt dynamically in a program. This includes both the language itself (phrases, words), but also the context, the meaning behind the words, body language, previous knowledge etc. Since humans conversate to gain a *mutual understanding,* one must design a chatbot which is able to gain a mutual understanding with the user.

Moore (2018) states that none of the natural language processing tools or conventions of GUIs help designers decide how to combine bits of natural language together into naturalistic conversational sequences (p. 182). It is therefore clear that the textual, contextual and conversational competence of both the designer and the chatbot is important in order to address to satisfy users. Challenges like misinterpretation, limited responses, security and privacy, how they are trained, bias, complexity etc needs solutions.

Several suggestions on how to 'resolve' some of these challenges have been made. Moore (2017) suggest a Natural Conversation Framework, Amershi et al. (2019) suggest Human-AI Guidelines, Kocielnik et al. (2019) suggest three 'setting expectations' techniques. For instance, Amershi et al.'s (2019) guidelines G1 (Make clear what the system can do) and G2 (Make clear how well the system can do what it can do) could possible resolve some of these challenges because implementing these guidelines in the chatbots can create a more accurate expectations for the users. In the chatbots' UI this can be done by displaying important metrics, informing the user about this during the interaction, and using language that states when the chatbot does not know or think it know something (p. 3). Personally, I think that the suggestion from Følstad and Brandtzæg (2017) in combination with the other mentioned is the best solution. They argue that designers need to move from viewing design of chatbots as an explanatory task – to an interpretational task – meaning an understanding of the *users' needs and how to address them* (p. 41).

1.5 References:

Aamli, K. (2018, 16th november). *Ungdom er naive i møtet med Instagram.* *https://forskning.no/barn-og-ungdom-media-oslomet/ungdom-er-naive-i-motet-med-instagram/1257202*

Anyoha, R. (2017, 28th August) *The History of Artificial Intelligence.* Harvard University. https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, 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. (https://www.microsoft.com/en-us/research/uploads/prod/2019/01/Guidelines-for-Human-AI-Interaction-camera-ready.pdf)

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

Blekastad, M (2021, 6th september). Karel Čapek. *Store norske leksikon.* https://snl.no/Karel_%C4%8Capek

Borgan, E. (2019, 29th November). *Kunstig intelligens blir mannssjåvinistiske rasister. Hva kan vi gjøre for å stoppe det?* Forskning.no. https://forskning.no/arbeid-it-juridiske-fag/kunstig-intelligens-blir-mannssjavinistiske-rasister-hva-kan-vi-gjore-for-a-stoppe-det/1599018

Buchanan, B. G. (2005). A (Very) Brief History of Artificial Intelligence. *AI Magazine*, *26*(4), 53. https://doi.org/10.1609/aimag.v26i4.1848

Cahalane, C. (2019, 17th April). *9 Useful apps for people who are Deaf or have hearing loss.* AbilityNet. https://abilitynet.org.uk/news-blogs/9-useful-apps-people-who-are-deaf-or-have-hearing-loss

Copeland, B.J. (2021, 19th June). Alan Turing. British mathematician and logician. *Britannica.* https://www.britannica.com/biography/Alan-Turing#ref214877

Dautenhahn, K., 2018. Some Brief Thoughts on the Past and Future of Human-Robot Interaction. ACM Trans. Hum.-Robot Interact. 7, 4:1–4:3.

Dennis, M. A. (2021, 15th July). Allen Newell. American computer scientist. *Britannica.* https://www.britannica.com/technology/artificial-intelligence-programming-language

Emotech (2021). *Reimagining the relationship between humans and technology.* Emotech. https://www.emotech.ai/#company

Emotech (2017, 27th September). *Olly – The First Home Robot With Personality* [Video]. https://www.youtube.com/watch?v=xCS8YXhT7j0&ab_channel=Emotech

Følstad, A. (2021, 22th September). *Interacting with AI* [Lecture presentation]. https://www.uio.no/studier/emner/matnat/ifi/IN5480/h21/interacting-with-ai-2021---module-2---session-1---handout.pdf

Følstad, A., & Brandtzæg, P. B. (2017). Chatbots and the new world of HCI. interactions, 24(4), 38-42. (https://dl.acm.org/citation.cfm?id=3085558)

Goodwin, M & Thormodsæter, M. (Programledere) (Universitetet i Agder). (22.04.2019). *Game Over? «De tidlige algoritmene – Ada Lovelace og Alan Turing».* Universitetet i Agder. https://open.spotify.com/episode/4xVFoFC0o5kME11TybF9pc?si=0c8a8912d6de4f8f

Goodwin, M & Thormodsæter, M. (Programledere) (Universitetet i Agder). (06.08.2021). *Game Over? «Rammeproblemet».* Universitetet i Agder. https://open.spotify.com/episode/2fRnvfR05fo4Fs10ZiwWIW?si=EIrkjrR_TYautXmrcbkduQ&dl_branch=1

Grudin, Jonathan. AI and HCI: Two Fields Divided by a Common Focus. AI magazine 30, no 4 (September 18, 2009).

Heaven, W. (2020, 18th November). *The way we train AI is fundamentally flawed*. MIT Technology Review. https://www.technologyreview.com/2020/11/18/1012234/training-machine-learning-broken-real-world-heath-nlp-computer-vision/

IBM Cloud Education. (2020, 3th June). *Artificial Intelligence (AI).* IBM. https://www.ibm.com/cloud/learn/what-is-artificial-intelligence

Illustrated Fiction. (2018, 25th December). *Sonny's Interrogation in I, Robot* [Video]. Youtube. https://www.youtube.com/watch?v=eI9IlAQGgZM&ab_channel=IllustratedFiction

IMDB (2021). *I, Robot.* https://www.imdb.com/title/tt0343818/

iRobot (2021). *We'll keep cleaning while you're away.* https://www.irobot.com/roomba

Iqbal, M., (2021, 8th October). *Instagram Revenue and Usage Statistics.* Business of Apps. https://www.businessofapps.com/data/instagram-statistics/

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. (https://www.microsoft.com/en-us/research/uploads/prod/2019/01/chi19_kocielnik_et_al.pdf)

Kraggerud, E. (2019, 30th January). Homer. *Store norske leksikon.* https://snl.no/Homer

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. https://www.microsoft.com/en-us/research/wp-content/uploads/2016/08/p5286-luger.pdf

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

Moore, J. R. (2018). Chapter 9 - A Natural Conversation Framework for Conversational UX Design. IBM Research. *Studies in Conversational UX Design,* Human-Computer Interaction Series. Springfield International Publishing. DOI:10.1145/3304087

Oxford University Press (2021). https://www.oxfordreference.com/view/10.1093/oi/authority.20110803095426960

Rouhiainen, L. (2019). *Artificial Intelligence: 101 Things You Must Know Today About Our Future*. https://books.google.no/books?hl=no&lr=&id=P3fSDwAAQBAJ&oi=fnd&pg=PP1&dq=artificial+intelligence+today&ots=TZ-fEBpXrp&sig=mlLxDxFH6cz-ynNYCeyr5yo63H8&redir_esc=y#v=onepage&q=artificial%20intelligence%20today&f=false

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.

The Norwegian Digitialisation Agency (2021), *Interpretation and overview of test procedures for WCAG 2.0 A and AA.* Uutilsynet. https://www.uutilsynet.no/english/interpretation-and-overview-test-procedures-wcag-20-and-aa/138

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

Yang, Q., Steinfeld, A., Rosé, C., & Zimmerman, J. (2020). Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. In Proceedings of the 2020 CHI conference on human factors in computing systems (Paper no. 164). (https://dl.acm.org/doi/abs/10.1145/3313831.3376301)

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

Wallén, J. (8th may, 2008). The history of the industrial robot. Technical report from Automatic Control at Linköpings universitet, 1-18.

Øverland, O. (2019, 17th October). *Isaac Aminov.* Store norske leksikon. https://snl.no/Isaac_Asimov

Module 3 – use the feedback to edit your last two modules

1.6 Appendix 1 – Feedback 1

The feedback I got on module 1 was “good and relevant quotes – I especially like the discussion about the definitions on AI” and “you have answered thoroughly on all assignments”. The reviewer also wrote that I have some language and formatting errors. For instance, I had made mistakes such as using ‘an’ instead of ‘and’. I edited the text again, so these things are now corrected.

1.7 Appendix 2 – Feedback 2

The feedback I got on module 2 was that I have good control of the literature and that I use good examples in my text. The reviewer also liked the table in module 2 task 2.1 «Two Design Guidelines» as it “gave a good overview over the arguments and thoughts you are presenting”. One thing that my reviewer wanted me to edit, was the formatting of the text – specifically to split up some of the sections to give the reader more “pauses”. This has been corrected in module 3 (this paper).

Appendix 1 & 2 (feedback) is located after the references in this paper