Nadia Møller, nhmoller
IN5480, Autumn 2021
22/10 2021

Individual Assignment

IN5480

Iteration 1

Module 1: History, Concepts and Universal Design

+

Iteration 2

Module 2: Design of interaction with AI

Iteration 1

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

A brief account of AI

Grudin (2009) offers a brief walk-tough of the main events taking place in the field of AI. In his article, Grudin (2009) presents how AI and HCI/CHI came to be fields of their own, alongside and in competition of each other in terms of academic and economic resources (Grudin 2009, p. 49). The term "Artificial Intelligence" was used for the first time in relation to a workshop in 1956 by John McCarthy, while the line of thinking of computers as (in the future) compared to human intellect was already presented in 1949 by Alan Turing (Grudin 2009, p. 49). The term and thought of computers and AI were primarily forwarded by mathematicians, imagining the potential for computers and systems to become autonomous and intelligent. This fascination and acknowledgement of potential spread to different institutions and made in AI-research target for investments (Grudin 2009, p. 50). AI evolved alongside HCI, but with a distinct focus on intelligence and a base in the field of mathematics and later entering computer science, not necessarily grounded in psychology like HCI (Grudin 2009, pp.50-51).

Definitions of AI

Bratteteig & Verne (2018) are with their article exploring the intersection, potential, and future of participatory design (PD) and Artificial Intelligence (AI). To frame AI in this context, they are using the following definition:

"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. 2)

This definition is tied to human intelligence and portrait an AI definition where the processing of data and pattern recognition is more developed (for the action to also demand human intelligence when executed) and that this potentially could make decisions that would erode the need for PD in design.

A second definition on AI can be found from one of the early scientists working with AI. In the 1950'ies Marvin Minsky defined AI as following:

"the science of making machines do things that would require intelligence if done by men." (Dennis 2021).

After this, Minsky continued to define the concept of "frames" to specify the information that is needed for machines to be able to move in space. Already in its early years, the potential for AI was compared with humans.

A third look at a definition can be found by the European Commission and presented by the Norwegian Government as:

”AI systems act in the physical or digital dimension by perceiving their environment, processing, and interpreting information and deciding the best action(s) to take to achieve the given goal. Some AI systems adapt their behaviour by analysing how the environment is affected by their previous actions.” (Norwegian Ministry of Local Government and Modernisation, 2020).

This definition is presented by the Norwegian Government as part of the National Strategy for Artificial Intelligence from 2020. AI is here presented as almost actors, as it has to ability to “perceive” their environment and make decisions based on the interpretation and processing of information – all while having a goal with this. The way AI is presented here, can again be interpreted as human-like (perceiving and interpretating) but it also makes room for the learning component, that the systems will adjust behaviour based on former behaviour patterns. This definition is very implicit on the input data and pattern recognition.

A definition of AI of my own would need to include how data is processed and applied unto known and new unknown situations – much like how a human would meet the world. It could we worded as “Artificial Intelligence describe systems being able to handle and process large quantitative amounts of data and being able to act upon old and new information.

Article review

A review of the article: “**The ‘Problem’ with Automation: Inappropriate Feedback and Interaction, not ‘Over-Automation**” by Norman (1990).

As the development an application of technology is becoming part of our work routines, it becomes ever so relevant to discuss how such advances are affecting how we work. Through three case studies of the aviation industry, Norman (1990) offers a discussion on how automation both have meant a great leap in terms of aircraft security and an increasing disconnect between the systems and the staff. On one hand, many human errors are being erased due to systems having the control of the operation of flying an aircraft. On the other hand, Norman argues that the automation of the operations is leading to a disconnect and isolation of the staff from the systems, causing huge disadvantages, mentally and physically, as they are left ‘out of the control loop’, should a situation arise (Norman 1990, p. 140). Norman shows throughout his cases how this disconnection from the system can lead to incontrollable situations for the staff if they are indeed ‘out of the loop’. But the answer is not only in the automation. Norman argues throughout the paper that the problem is mainly to be found in the design of the automation; in the way feedback is provided and how interaction between humans and computers/systems are designed.

The article was published in 1990. Here 31 years later our society are arguably even more automated, and thus the questions arise: How does Norman’s arguments and cases translate to the implementation and usage of automation in today’s day and age? Are we still facing the same problems regarding disconnection and being in/out of the loop?

AI as a feature
An example of AI as part of a company's business, can be the autopilot of Tesla's cars (Tesla n.d.a; Tesla n.d.b). The autopilot is a feature of the Tesla, and the potentials for the autopilot and the self-driving possibilities are presented here. On the webpage of the autopilot/AI, several aspects of the AI behind the autopilot are described: the hardware, AI as neural networks, algorithms for autonomy, the foundational coding, the infrastructure for evaluation, and lastly the Tesla bot (Tesla n.d.b.).
The AI works as a foundation for the autopilot, as it should be founded on "Advanced AI for vision and planning", alongside with the other components. Each component is both described on the page. But the language is ambivalent – for some of the components, it is written strangely with imperatively wording, almost like a suggestion to how future employees or other people of interest can work with said component.
AI is the core of the systems that enable potential self-driving and is not presented with the range of concerns that can be with an automated autopilot and self-driving cars.

AI depiction in (pop-)culture
The possibilities and potential danger of AI is a popular theme in pop-culture. One series where AI and the future of technology is addressed is the show Black Mirror on Netflix. One episode is particularly relevant here: "*Rachel, Jack, and Ashley Too*" (Black Mirror 2019). In this episode a young girl receives the AI robot doll, Ashley Too. This robot doll is modelled on a popular singer and functions as a 'friend' for our main character. The robot encourages the main character to engage and come out of her comfort zone, which crosses a boundary between human and robot according to her father and is therefore shot down. Here, in the first half of the episode, the interaction at first is portrayed very innocently as she receives the robot. Up until the shot-down, the 'friendship' becomes more and more intense, and the robot seemingly becomes more autonomous in its interaction with our main character. The relationship and form of interaction changes greatly throughout the episode, which emphasises that the systems behind Ashley Too are not linear and constant, but ever evolving – much like we see in our main character.

1.2 Robots and AI systems

How 'Robot' came about
The word Robot comes from the Czech word "Robota", meaning slave labour. Robot as a word was used for the first time in a play written by Karel Capek in 1920. The play portraited a dystopian future where robots would take over all labour and, in the end, exterminate humanity (De Etiske Råd 2007).

Definitions of "Robot"

Two definitions on robots are presented in Thrun (2004). The definitions provided are from respectively the Robot Institute of America, 1979, and the Merriam Webster's Collegiate Dictionary, 1993, and are as following:

1. "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)
2. "An automatic device that performs functions normally ascribed to humans or a machine in the form of a human." (Thrun 2004, p. 11)

The first definition puts a lot of emphasis on the programmable background to robots and a lot on the possibilities that the programming can lead to. The second definition on the other hand "reduces" the robot to a "automatic device" where the primary characteristic is that it can imitate human functions.

A definition of robots of my own would be something like – "*a programmed system having a physical form and able to perform specific tasks with or without human inputs*". This definition is and cannot totally be accurate due to my limited knowledge of the full possibilities and limitations of robots. But I do think it is important to include the core elements of robots, such as the underlaying programming, some sort of physical form and lastly that it holds *some* degree of autonomy.

AI and Robots

Given the definitions above, AI and robots do have some overlap. Both can be described in relation to humans and humans' actions. They both bear potential for developing even more abilities to simulate (or surpass) the humans building them. On the other hand, they have two different forms – AI is described as large systems and networks of data and algorithms. Robots are generally ascribed a physical form. The similarities and differences are somewhat connected, when we see AI technology "inside" robots, as part of their coding, and here the boundaries are becoming blurrier.

Robot movement

An example of a commercial robot could be that of robot vacuum cleaners. The robot vacuum is one that uses sensors to measure out the dimensions of the room or potentially whole house and moves around in patterns to ensure it is vacuuming the whole space. The robot moves from a base station where it is charged and resting before its next rounds. One would think that such a robot is easy breezy to put into the daily routine. The problems that arise are both that the robot requires all/most things in the room lifted or removed from the floor and that it makes a relative amount of noise. Lifting chairs off the floor and unto the table might be easy for many, but more difficult for users having such a robot because vacuuming is too demanding. On the other hand, having a robot to assist you in your daily routine – in your home – might lead to users connecting and personifying their robot and interaction with it like a companion.

1.3 Universal Design and AI systems

A definition on Universal Design

The term “Universal Utforming” is defined in ’Likestillings- og diskrimineringsloven’ (2017) as “Med universell utforming menes utforming eller tilrettelegging av hovedløsningen i de fysiske forholdene, inkludert informasjons- og kommunikasjonsteknologi (IKT), slik at virksomhetens alminnelige funksjoner kan benyttes av flest mulig, uavhengig av funksjonsnedsettelse.” (Likestillings- og diskrimineringsloven, 2017, Kap. 3, §17)
This definition of universal design is one that focuses on the accessibility of all public and open-to-public sites and websites. It specifies both ICT and non-ICT situations as sites for inclusion unless these precautions are themselves a hinder to the corporation. Disability is mentioned as something that should not hinder usage of the main functions, but it does not specify to what extend it is expected beyond the core functions. In return is disabilities not specified and this could be interpretated widely. As this definition is taken from a legal perspective, it probably includes broader limits, applicable to the boarder society, than a definition only concerned with interaction design.

The potential of AI with respect to diversity in users

AI has – and will probably gain more – the ability to transform or modify how we can perceive and take in the world around us. One example of how AI systems can include “more” users, is via face recognition and image recognition software. This can help users with little or no sight interact with their surroundings. Another example is the potential in making better AI systems, by providing more diverse datasets, in order for the systems to recognize and acknowledge neurodiversity. This could help users in i.e., job seeking, but also other “automatic” sorting in that ballpark.

The potential of AI for both including and excluding people.

Using AI technology to process the world around us can help include people with various degrees of disabilities. Examples of this includes automatic texting of videos and sounds and voice assistants. On the other hand, AI is also based on data and processing these in ways that is not transparent. Some degree of ‘transparency’ demanded in the WCAG (WCAG 2018) – but the processes behind AI and the feedback coming from decisions hereof are not transparent for most people, thus excluding people i.e., not working in tech.

My understanding of “understanding” + Do machines understand?

To me, understanding something relates to both how I perceive and receive information and how it is processed and connected with previous knowledge. This leaves us with a highly subjective process. Often, I would also argue that to understand and understanding a subject or a concept, mean to be able to use elements of it in practice or in relevance to other topics. This involves a complex procedure where multiple nuances need to be considered and related to eachother.

Though, above is my interpretation of human understanding. If only looking at a matter-of-fact processing process of taking in new information and taking it into a greater spider-web of knowledge, it would be possible to claim that machines can understand. We know AI-systems are capable of connection points and recognize patterns. Though, I do think part of the process of understanding is also that perception, experiences, and thus the spider-web of knowledge is deeply individual and subjective and to be able to use the knowledge on new unknown situations. In this regard, my impression is that machines can be more "uniform" at least when sharing the same coding and processing algorithms. Then the question for me is: are machines able to make different connections and subjective experiences while having the same wiring and coding as another robot – meaning, can we apply the discussion of nature vs. nurture to machines?

1.4 Guideline for Human-AI interaction

Microsoft's guidelines for human-AI interaction
One of the guidelines defined by Microsoft, is Guideline 6: Mitigate social biases (Amershi et al. 2019). This guideline is supposed to address how social biases should mitigated i.e., by using language that does not assume gender and other such prejudices we might hold in social settings.
One other example that this guideline could address, could be that of data bias in AI and how AI react and act differently to people of different skin tones. These biases are seen today in i.e., face recognition and could be mitigated by a more conscious effort to 'feed' the systems with more diverse data material.

Human-AI interaction guidelines and WCAG
The Human-AI interaction guidelines put a lot of emphasis on the feedback of the system to the user (and back) and how to guide and meet the expectations of the user in interaction with the system (esp. guidelines 1, 2, 5, 11, 16 & 18) (Amershi et al., 2019). These guidelines should help the user in the interaction with the AI system and ultimately make the system more transparent. The chosen examples are somewhat in spirit similar to the third principle of the WCAG 2.1 guidelines: making systems understandable (WCAG, 2018). They are similar in the way that the guidelines are created for designers to have something concrete to measure AI-infused systems on. Here the first guidelines are focused on how users should be able to understand what the system can do – sort of an equivalent how content should be understandable in WCAG 2.1. And then, they operate with their own field of understanding, and this is partly why these new guidelines were needed (Amershi et al., 2019). Where the two sets of guidelines are fundamentally different, is in the field of the how the AI-infused system should be learning from the user and the expectation for the user to 'teach' the system (i.e., guidelines 7, 8, 9, 12 & 13). This notion of interaction is not at all present in the guidelines for web content that is the WCAG 2.1.

References – iteration 1

Amershi, S. et al. (2019). Guidelines for Human-AI interaction. CHI 2019, May 4–9, 2019.

https://www.microsoft.com/en-us/research/uploads/prod/2019/01/Guidelines-for-Human-AI-Interaction-camera-ready.pdf

Black Mirror (2019). "*Rachel, Jack and Ashley Too*", Episode 5, Season 3. Released 5th of June 2019, Netflix.

Dennis, M. (2021). "Marvin Minsky". Britannica, Aug 5 2021.

https://www.britannica.com/biography/Marvin-Lee-Minsky

Det Etiske Råd (1/9 2007). "Robot – Maskinslaven". Accessed 10/9 2021.

https://www.etiskraad.dk/etiske-temaer/optimering-af-mennesket/homo-artefakt/leksikon/robot

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

https://aaai.org/ojs/index.php/aimagazine/article/view/2271

Likestillings- og diskrimineringsloven (2017). *Lov om likestilling og forbud mot diskriminering*. (LOV-2019-06-21-57), Lovdata:
https://lovdata.no/dokument/NL/lov/2017-06-16-51/KAPITTEL_3#%C2%A717

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)

https://www.jstor.org/stable/55330?seq=9#metadata_info_tab_contents

Tesla (n.d.a). "Autopilot". Accessed 10/9 2021.
https://www.tesla.com/autopilot

Tesla (n.d.b). "Artificial Intelligence & Autopilot". Accessed 10/9 2021.
https://www.tesla.com/AI

The Norwegian Ministry of Local Government and Modernisation (printed 2020): "National Strategy for Artificial Intelligence".
https://www.regjeringen.no/contentassets/1febbbb2c4fd4b7d92c67ddd353b6ae8/en-gb/pdfs/ki-strategi_en.pdf

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

https://www.tandfonline.com/doi/pdf/10.1080/07370024.2004.9667338

Verne, G. & Bratteteig, T. (2018). Does AI make PD obsolete?; exploring challenges from Artificial Intelligence to Participatory design. In *Proceedings of PDC 2018, Aug. 2018, 5 pages.*

https://dl.acm.org/citation.cfm?id=3210646

WCAG 2.1 (2018). Web Content Accessibility Guidelines (WCAG) 2.1.
W3C Recommendation 05 June 2018.
https://www.w3.org/TR/WCAG21/#understandable

Iteration 2

2. 1. Characteristics of AI-infused systems

2.1.1.

AI-infused systems are defined and used by Amershi et al. (2019) as "…systems that have features harnessing AI capabilities that are directly exposed to the end user" (2019, p. 1). This definition is also presented in Følstad (2021) as a way into categorizing different types of AI; as artificial *super / general / narrow* intelligence (2021, slide 12). So, while AI-infused systems come in different forms, some characteristics of the limits of these, are presented in the articles of Amershi et al. (2019), Kocielnik et al. (2019), and Yang et al., (2020), all combined with their suggestion for countermeasures to address these limits and characteristics.

Amershi et al. (2019) define AI-infused systems as being filled with uncertainty, inconsistency in their behaviour, and being able to do behind the scenes personalization without much transparency of the process. These characteristics are by the team addressed in a four-phase process of developing guidelines for how interaction between end users and AI-infused systems can be designed (Amershi et al., 2019).

Yang et al. (2020) define the key characteristic features of AI-infused systems through five main "challenges" that arise when trying to design these systems. The reported design challenges were "1) uncertaincy surrounding AI's capabilities and 2) AI's output complexity, spanning from simple to adaptive complex" (2020, p. 1). With these challenges, AI is characterized as some form of black box where possibilities and limitations and how the outputs from the algorithms behind, are neither clear nor transparent for the designers.

The AI-infused systems are in Kocielnik et al. (2019) more generally described as probalistic (and thus often not accurate), impacted by user actions, and having transparency issues. Kocielnik et al. (2019) makes define these characteristics in context of how interaction with AI-infused systems can become difficult for end users, when they do not know what to expect when they don't understand the system and might have previous firsthand experiences tainting their interaction (2019, p. 2).

2.1.2.

To illustrate an AI-infused system I am familiar with, I have chosen the voice assistant Google Home. The Google Home is an assistant with which one interact with through a voice interface. In my experience this assistant exemplifies several the above-mentioned characteristics.
Kocielnik et al. (2019) describe the AI-infused systems as probalistic and thus the inaccuracy in their responses is affected (2019, p. 3). For me, the voice assistant will hardly ever either 1) hear me or 2) understand me correctly in my requests (which did funnily enough once result in a Rick-Roll). Some studies have shown that voice assistants have a bias against women's voices and non-native English speakers, which can contribute to show that the data behind is flawed and will lead to inaccurate responses (Bajorek 2019; Lloreda, 2020). These issues in the system also contributes to a lack of transparency and inconsistency for the user, as the voice assistant will sometimes respond very well and sometimes hardly at all. The transparency and inconsistency here is especially that the users are not informed on why they are experiencing such variance in responses.
The consequences of these examples lead to an unreliable service (for some). Most people annoyed by this might quit and disconnect the service. But people who are relying on a voice interface, might be excluded as a result.

2. 2. Human-AI interaction design

2.2.1.

Both Amershi et al. (2019) and Kocielnik et al. (2019) address the challenges of designing interaction for AI-infused systems. The focus of Amershi et al. (2019) is mainly concerned with developing design guidelines that could help interaction designers designing these interactions. The need for new design guidelines comes from the changed field for designers, as AI have its challenges with transparency in regard to when errors occur and inconsistent and unpredictable behaviour (2019, p. 2). The guidelines can be used to analyse and evaluate already existing products and help designers work with future ideas (2019, p. 11).

In the article of Kocielnik et al. (2019), the actual interaction between end user and system is in focus. Here, they address how we as users meet systems and interacts with then, based on a set of expectations and anticipation. If these two key factors are not met or only met sporadically, the faith in the interaction will disappear and lead to less interaction between users and the system in question (2019, pp. 1-3).

2.2.2

In my experience of interacting with the voice assistant from Google, the guidelines G11 and G13 could be fitting to discuss.

The guideline G11 is "Make clear why the system did what it did". In relation to the voice assistant, this guideline illustrates a deviation of the voice assistant. It is never clear to the user how it will find the answers it will give, where misunderstandings, and misinterpretation origin from the user's input. This can be seen in contrast to i.e., Apple's voice assistant, Siri, where it does have a visual representation in addition. Here, it will write what it thought you said. For a purely audial interaction with the Google Home, none of that feedback – and thus processing, are revealed. These potential misunderstandings (i.e. if the users is not an native English speaker) could be made transparent by the speaker asking if it has understood the user correctly.

The other guideline, G13, says "Learn from user behaviour". Here, there is potential, but also a point from where it deviates. Could the voice assistant become costumed to the voice of the woman at the house – or could it slowly learn to recognize accented English speakers or special voiced, thus learning from the user, no matter the language chosen in the settings.

2.2.3.

The article of Bender et al. (2021) present the argument that large language models are not neutral and therefore should not be treated as such. The argument of the article it that data sourced online for these large language models are not without risks and potential harm. Sourcing language material from online fora could bring several biases regarding language, tone of voice and uneven representation. By sourcing primarily English material online, a biased picture is created; are native English speakers overrepresented and quoted more? The thought behind sourcing language-material online, is that the internet is a global and diverse place and thus the data would be likewise. But bias exist in who are writing the texts that the language models are gathering. Some viewpoints and arguments are pushed forward and thus can obtain hegemonic status, that will then be perpetrated in the language models. This language used by the ones holding hegemonic status, can contain biases towards curtain ethnic, religious or otherwise marginalized groups. By using this data, the risk is that these biases could become encoded in the language model – without regard to social contexts and sensitivities. Another point is the bias in who have access – and especially stabile access – to the internet. Here, the representation would become uneven, in favour of richer communities.

One solution proposed is to address the documentation debt. This means to document, and to budget for documentation, the creation of datasets (Kocielnik et al., 2019, p. 615). This solution would make room for more transparency of the datasets that language models are based on and learning from. By documenting these training data, one would be able to adjust the gathering to mitigate some of the biases.

2. 3. Chatbots / conversational user interfaces

2.3.1.

The development of and transition towards more chatbot-human-interaction builds on the challenges presented earlier. Especially one challenge is repeatedly mentioned: the shifting focus of design of the visual experience to the design of conversations (Følstad & Brandtzaeg, 2017, p. 40) and how this poses new challenges for designers designing AI systems (Yang et al., 2020). This shift in the design focus and need is presented as a challenge because design usually is dealing with guiding the user in the interaction. When the meaning of interaction with the system becomes the conversation between the user and the chatbot, the role of the designer changes. These changing roles and work with AI-infused systems again poses a range of other challenges related to the design of AI (as described earlier).

Another challenge caused of the design of chatbots, is the need for large language models (Bender et al., 2021). In order for the chatbot to be able to respond in natural language in the conversation, it needs a large dataset. The potential risks and harm of these datasets were presented in section 2.2.3. above. Meanwhile, all chatbots are not fully equipped in the world of language as humans yet, and that causes challenges. Designers usually work with the needs and expectations of the users, with designing systems. This, however, is difficult to meet with the state of chatbots now. Some aspects of this is, one, designers not having enough insight into the possibilities and limitations of AI (Yang et al., 2020, pp. 2-3) and second, chatbots not yet having enough data to be accurate enough in their understanding, causing usability and expectation problems for the users (Kocielnik et al., 2019).

2.3.2.

Guidelines G1 and G2 respectively suggests that the chatbot should make it clear for the user what it can do, and how well it can do it. The work of making the chatbot clear on its abilities and limitations, seems like a core task and path for how interaction between users and chatbots can be designed. Working to make the possibilities and limitations of the chatbots more transparent, might be relevant for both what Følstad & Brandtzaeg (2017) and Kocielnik et al. (2019) are addressing. The changing focus of design, conversation, might be clearer when the relationship between the chatbot-interface and its language model-database behind is made clear, i.e., will it become clearer how the chatbot could actually interact with the user. Here, Følstad & Brandtzaeg (2017) see a range of opportunities for a new phase of the field of HCI (2017, pp. 41-42). At the same time, being clear on the limitations could mitigate the challenges presented by Kocielnik et al. (2019) regarding users' expectations when meeting chatbots.

Appendix 1 – Management of feedback on Individual Assignment, iteration 1

The first wish presented for the revision of Iteration 1, is to write together the paragraphs "my understanding" and "do machines understand?" to make it more coherent. This section has now been revised.

The second wish was for part 1.4 to be elaborated, in order for the differences between WCAG and Amershi et al. (2019)s guidelines to be clearer. This section has now been revised.

References – iteration 2

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

Bajorek, J. P. (10/5 2019). *Voice Recognition Still Has Significant Race and Gender Biases.* Harvard Business Review. Accessed on 22/10 2021. https://hbr.org/2019/05/voice-recognition-still-has-significant-race-and-gender-biases

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

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

Følstad, A. (29/10 2021). *Session 1; Interaction with AI – Module 2.* https://www.uio.no/studier/emner/matnat/ifi/IN5480/h21/lecture-notes/interacting-with-ai-2021---module-2---session-1---handout.pdf

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

Lloreda, C. L. (5/7 2020). *Speech Recognition Tech Is Yet Another Example of Bias*. Scientific American. Accessed on 22/10 2021. https://www.scientificamerican.com/article/speech-recognition-tech-is-yet-another-example-of-bias/

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