IN5480 - Specialization in Research in Design of IT

Individual assignment fall 2021 - Stinelly

Table of contents

Module 1

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

The term AI has been around for decades. The term was first used f in 1956 by John McCarthy, a logician and mathematician, who sent out a call for participation in a workshop (Grudin, 2009, p. 49). McCarthy was not the only one to ponder over the possibility of computers performing tasks and behaving in ways that are characteristic of the human intellect. The logician and mathematician Alan Turing, wrote in 1949 an article published in the *London Times* where he address this exact topic (Grudin, 2009, p. 49).

If we pick out two random people from the street and ask them to define AI, it is likely that their answers will differ. Hence, it is useful to look at a couple of different definitions before we move forward. One definition of AI is that "artificial intelligence is the simulation of human intelligence processes by machines, especially computers " (Burns et al., 2021). This definition refers to the human intelligence processes, such as facial recognition and speech recognition. The Merriam-Webster dictionary defines AI in two ways. It states that artificial intelligence is "a branch of computer science dealing with the simulation of intelligent behavior in computers" and that AI is "the capability of a machine to imitate intelligent human behavior" (*Artificial Intelligence*, n.d.). To imitate intelligent human behavior means that the AI in itself does not display human behavior, instead the machine follows a set of rules on how to behave that is intended to appear as human behavior. A final definition of AI is presented by the Norwegian government in *The National Strategy for Artificial Intelligence*. Here they state that "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 give goal. Some AI systems adapt their behavior by analyzing how the environment is affected by their previous actions" (Ministry of Local Government and Modernisation, 2020). Taking all these definitions of AI into consideration AI can be describes as "machines that use informational input in order to simulate human intelligence and behavior".

In *Towards a Framework for Human-Robot Interaction*, author Sebastian Thrun writes about the different kinds of robots, robot autonomy and interfaces that the different individuals of

society might interact with. Thrun starts by introducing the three different kinds of robots relevant within the field of information technology. These three are industrial robots, commonly used in the automotive industry, the professional service robots, which "assist people in the pursuit of their professional goals" and finally the personal service robots, which "assist or entertain people in domestic settings or in recreational activities" (Thrun, 2004, pp. 11–12). Further, Thrun explains the different kinds of robots have different levels of autonomy. Industrial robots have the lowest level of autonomy, as their workspace is quite specific and easy to control, while professional service robots and personal service robots require a higher level of autonomy, as they work in environments where it is necessary for the robots to adapt to the unpredictability of human behavior (Thrun, 2004, pp. 14–15). Finally, Thrun looks at the different interfaces of robots, distinguishing between robots that require direct interaction, and those that require indirect interaction. This is an important distinction as the interface of a robot that can act on its own would not have the same requirements as a robot that needs to be operated by a person. All in all, I found this a well written and easy to understand article. I do wish that Thrun would have written another section on the consequences and effects of choosing one type of interface over another as I find this topic quite interesting, but that is more of a personal preference than a critique on the quality of the article.

One contemporary company that is well established within the field of AI is Microsoft, with their Microsoft Azure platform. Microsoft presents AI as a research area, focusing on "pursuing computing advances to create intelligent machines that complement human reasoning to augment and enrich our experience and competencies" (Microsoft, n.d.-a). Their focus is on AI as a product, and how their Azure platform is a tool for developers to create even better AI, which is reflected through the quote "Invent with purpose" (Microsoft, n.d.-b).

A documentary that goes into depth about AI and the consequences machine learning can have on society is *The Social Dilemma(2020)*. The Social Dilemma looks closer at how social media companies use the vast amount of user information available on their platforms to create detailed and deeply connected maps about their users and their interest. The

companies do this to show the user the content that is most likely to promote interest and interaction from the user (Orlowski et al., 2020). It explains how all aspects of machine learning and AI that users interact with on a daily basis is constantly gathering our data. As Tristan Harris puts it, “Social media isn’t a tool that’s just waiting to be used. It has its own goals and it has its own means of pursuing them by using your own psychology against you” (Orlowski et al., 2020).

1.2 Robots and AI systems

The word robot long predates the current day definition of what a robot is.
Robot means “forced labor” and derives from the Czech *robota*, which appeared in the play *R.U.R.: Rossum’s Universal Robots* by Karel Čapek from 1920 (Britannica, n.d.).

One definition of robot is from the Robot Institute of America, who in 1979 defined robot as “a programmable, 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). This definition looks at robots being able to move elements through physical space, like serving customers in restaurants or handing a surgeon a tool in the Operating Room. Another definition of the robot is “any automatically operated machine that replaces human effort” (Britannica, n.d.). This definition is a bit broader than the previous one. It does not only look at robots performing physical human effort such as movement, but may also include speech and solving math equations, which would require mental effort. Personally, I would define robot as “an object created by humans, that can replace human tasks which require physical movement”.

On one side, robots and AI are not that different. Both are man-made and can replace different human performed actions. In addition, both require written programs to know what to do and what rules to follow. On the other hand, robots and AI are as different as can be. Robots are physical and can only follow a predefined set of rules, while AI is designed to learn from the environment that interacts with it. If the robot walks into a wall, it will continue walking into the same wall each time unless someone updates the software. If an

AI is faced with a picture of a banana it will after a while recognize that these pictures have something in common.

An example of a contemporary robot is the robot vacuum cleaner. The iRobot comes with an app that lets you map the size of your rooms in your home (*IRobot Home App | IRobot*, n.d.). It also allows you to specify which areas you want to send your robot vacuum to, and which areas are so called “no-go-zones”. However, if your living room has a stairwell in one corner that you have not defined, and the robot happens to fall down the stairs, it will continue to make the same mistake until the inconsistency between the actual living room, and the mapped out living room is changed.

1.3 Universal Design and AI systems

The Center for Excellence in Universal Design was established by the National Disability Authority in 2007 (About Us | Centre for Excellence in Universal Design, n.d.). They define Universal Design as “the design and composition of an environment so that it can be accessed, understood and used to the greatest extent possible by all people regardless of their age, size, ability or disability” (What Is Universal Design | Centre for Excellence in Universal Design, n.d.). This definition means that Universal Design is design that from the beginning of the design process takes all possible user into consideration, and designs with the intention of the product being accessible to all. This stands in contrast to designing for one possible audience, and then later adapting the product to a wider audience. With respect to inclusion, I understand Universal Design as designing for the few before designing for the many. An example would be to prioritize writing descriptive code so that individuals using e-readers can access the sight, and then later making the sight visually. The aesthetic of a web page is not a criterion in order to understand its content, but clean code is necessary so that individual using an e-reader can understand and navigate the content of the web page.

AI is a promising field when it comes to inclusion. Our perception of the society around us is influenced by a bunch of different factors. For one, emotions and current mental state can fog our perception, or make us less observant. AI on the other hand will work

the same every time, so while we might not notice that someone is agitated, the AI can. This is connected to the potential of AI regarding human movement. Not only can the AI recognize out emotions, it can also be able to anticipate our movements. It is not inconceivable that this kind of AI could be used to give feedback on exercise form and technique. One example of AI that does this today is the app 'Smart Baduanjin'. An app that uses AI human pose detection to give feedback to elderly on their Baduanjin practice (Au Yeung, n.d.). Lastly, AI also holds potential when it comes to human cognition. In this case, human cognition means the mental processes our minds perform every day. Such as noticing patters around us or performing repetitive tasks. One example of AI that takes use of this is IBMs Crypto Anchor Verifier. The Crypto Anchor Verifier proves the authenticity of products by using AI and optical imaging (Dillenberger, 2018).

One example of how AI can help with inclusion is that it makes content easily available to the masses. There is AI that automatically text videos, or even university lectures as they are happening. This helps people who are hard of hearing, making it easier for them to participate in aspects of society that otherwise would not be that accessible. Examples of such products are Otter.ai, sonix.ai, and OTranscribe just to name a few. Looking at this AI seems perfect, like a gift to humanity making our lives easier. However, AI also has the potential to exclude everything from opinions to certain people from our online experience. One example of this is how the algorithm behind TikTok's "For You" page suppress videos of disabled users, among others (Biddle et al., 2020). Thus, excluding them from our user experience.

Understanding the environment around us and how things are connected is important in order to comprehend and perceive our environment with as little bias as possible. I believe that understanding the meaning of something, requires knowledge of the bigger picture. In example, it might be necessary to hold knowledge about Islam in order to understand the importance of Ramadan. With this in mind, I would further argue that machines do not understand. A machine in itself does not understand what emotions or anger means, nor does it understand culture or norms. instead, it has been trained to

detect different human emotions, and behave a certain way. It is all in the code and training data

1.4 Guidelines for Human-AI interaction

The fifth of Microsoft's eighteen guidelines for human-AI interaction is to "match relevant social norms". An example of this would be when taking an English web page and making it available in Arabic. In this case it is not enough to just translate the content, you would also have to change the direction it is intended to be read. The reason for this is that Arabic is read right to left in contrast to English left to right.

Microsoft is not the only company that has defined a set of design guidelines. Jakob Nielsen defined ten broad rules of thumb for design, called the *10 Usability Heuristics for User interaction Design* (Nielsen, 2020). These are a bit different than Microsoft's guidelines. Nielsen's rules of thumb are fewer and broader than Microsoft's guidelines. That being said, Nielsen and Microsoft address all the same concerns. Both highlight that it is important for the user to be told what is happening, that the system should match the real-world norms and expectations and that information about what the system does, and how it does what it does, should be easily available.

Module 2

2.1 Characteristics of AI-infused systems

AI-infused systems can be defined as systems that use AI to enrich or enhance the user experience. The system may respond differently to the same prompts from person to person and from day to day (Amershi et al., 2019, p. 2). This definition does not mention whether the AI has to be visible to the user or not. For that reason, the term AI-infused systems can be systems where the implemented AI is visible to the user, and systems where it is not.

A key characteristic of AI is that it is based on machine learning, thus constantly evolving. The dataset the algorithm learns from is big datasets that is gather whenever a user interacts with the AI. The fact that AI-infused systems often rely on user input and data to improve and evolve, this is not all fun and games. Amershi et al. describes AI-infused system as able to display "unpredictable behaviors that can be disruptive, confusing, offensive and even dangerous" (Amershi et al., 2019, p. 1), in addition to being prone to errors that can be difficult to prevent (Amershi et al., 2019, p. 2) . This description reflects a negative, yet important aspect of AI-infused systems. They are manmade, and their code and training data may be unintentionally biased. This can lead to offensive situations such as wrongfully categorizing something or someone, or when showing different search suggestions and recommendations based on the location and data of the person interacting with the AI. The systems are, as mentioned, error-prone, which is natural. AI works on training data and conditions specified in their code, but his training data and code has been written and often provided by humans. We, and AI, are imperfect, and sometimes the results of our action and code is not as well thought through or written as initially intended.

An example of an AI-infused system that exemplifies some of the above key characteristics is Siri. As users we are aware that Siri is an AI, based on machine learning algorithms. This makes Siri an AI-infused system where the AI is directly visible to the end-user. Siri does not try to hide that she is imperfect. When Siri can't decode exactly what the user says she will respond along the lines of "Sorry, I did not catch that". Siri is also error prone. It is not uncommon that she will show the search result of the wrong topic, play another song than what you ask for, or call the wrong contact. Siri will also adapt her suggestions based on the

datasets collected. For example, we can ask “Siri, what will the weather be like in Kingston today?”. In this case we have not specified whether we mean Kingston, Jamaica or Kingston, New York. In some cases, she might ask which one we meant, or give an answer based on our geolocation. If we ask, “what is the weather today?”, without any specified location, Siri will likely give the forecast for the city we are currently in.

With Siri, there will likely not be any serious problems if the answer we get is faulty. Calling up the wrong person is not the end of world, nor is not bringing an umbrella because we got the forecast of the wrong place. What can be troublesome is errors that occur when Siri misunderstand or completely misses what the user says because of their accent. It might feel discriminating to not be able to use such a handy tool because of an accent, but it won’t be your 13th reason.

2.2 Human-AI interaction

In the article “Guidelines for Human-AI Interaction the authors introduce 18 guidelines for how to design AI (Amershi et al., 2019). The authors believe that these guidelines “can serve as a resource to practitioners working on the design of applications an features that harness AI technologies, and to researchers interested in the further development of guidelines for human-AI interaction design”(Amershi et al., 2019, p. 1). The guidelines are divided into 4 categories: Initially, during interaction, when wrong and over time. They underwent several rounds of testing by both the authors and practicing designers, to ensure that they are easy to understand and to test against AI products. Amershi et al. conclude by stating that there is a lot to address when it comes to fairness and ethical considerations, and that the guidelines do not cover the entirety of all that needs to be considered (Amershi et al., 2019, p. 12). However, it is a start.

In the article “Will You Accept an Imperfect AI? Exploring Designs for Adjusting End-user Expectations of AI Systems” the authors Kocielnik, Amershi and Bennet study the impact of different kinds of expectations setting on AI systems. The AI system they use in the study is supposed to automatically detect whether or not there is a meeting request in free-text emails (Kocielnik et al., 2019, p. 1). They find that it is better that a system has high

precision, where False Positives are avoided, instead of the system having high recall, where false positives are not filtered out.

Previously I wrote about Siri as an AI-infused system and how it exemplifies different characteristics of AI. I will take two of the guidelines proposed in Amershi et al.'s article and discuss whether Siri adheres to or deviated from these.

The first guideline is nr.7, Support efficient invocation. With this guideline they mean to make it easy to invoke or request the AI systems services. To invoke Siri on an iPhone, all the user has to do is say "Hey Siri", another option is to press and release the side button. On MacBook's with a touch bar, it is possible to tap the Siri icon, as well as saying "Hey Siri", to invoke the interaction. That being said, all these options require the user to speak. If you want the option to type to Siri this has to be turned on inside the Accessibility tab in the settings app. I argue that Siri technically does adhere to this guideline, however improvements can still be made.

The second guideline is nr. 12, Remember recent interactions. With this the authors mean whether the system is able to maintain a short-term memory an allow the user to make efficient references to that memory. Siri does not completely satisfy this guideline. When interacting with Siri, it is only possible to see Siri's most recent response. Siri also keeps a log over all user interactions, but it is not quite clear for how long. Additionally, the data from some interactions can be saved by the user, such as "Siri, Call me your highness". All in all, while Siri does keep a log over user interactions, it is not possible for the user to properly navigate this data. Thus, Siri does not properly satisfy this guideline.

In regard to guideline 7, it is definitely possible to make some improvements to Siri as an AI-infused system. The option to type instead of speaking should not be hidden away in settings. An option could be to have a keyboard, or a button to show keyboard, appear when Siri is called in a way other than by speech.

Is it possible to make improvements in regards to guideline nr. 12. However, I believe it is more important to ask whether it is necessary to make these changes. When using Siri to set

an alarm or show the weather forecast, we do not need to see the reminder we asked her to set last time. Nor is it relevant to see how many cups is in a liter that we asked for when making dinner. Siri is meant to make things easier, and in some cases, seeing an interaction log can feel more like clutter in the interface than an improvement to the AI.

"On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" is an article written by Bender et al. in 2021. The article is a critical discussion of a specific type of AI-infused system, based on large language models. First of all, the authors point out the huge environmental and economic impact of such systems. In example that training a single base model like BERT requires the same amount of energy as a trans-American flight (Bender et al., 2021). The authors also point out how incredibly costly such systems can be. Not only is it expensive to train them, but to keep them up and running as well. Another problematic aspect the authors mention is the unfathomable amount of training data needed to create deep learning models with such a high level of accuracy. With such big datasets it is impossible to know exactly what they include, which increases the risk of encoding bias into the system. Or, as the authors say it, "Size doesn't guarantee diversity" (Bender et al., 2021). Bender et al. goes into a very detailed description of just how problematic biases like stereotypical associations or negative sentiments towards specific groups can be, and how nested into the system they can become. The authors conclude the articles by stating that "applications that aim to believably mimic humans bring risk of extreme harms", and that AI work should be focused on models where downstream effects are possible to be understood in order to "block foreseeable harm to society and different social groups" (Bender et al., 2021).

2.3 Chatbots/ conversational user interfaces

Designing chatbots and conversational interfaces can be challenging to say the least. The most important aspect to keep in mind is to remember to keep the design of AI human-centered. Google sums up the reason for this quite nicely on their site for the UX of AI. They write "if you aren't aligned with a human need, you're just going to build a very powerful system to address a very small, or perhaps nonexistent – problem" (*The UX of AI - Library*, n.d.). This means that the main focus should always be to design AI that people want or

need. If not, there will be placed a lot of recourses into something people are not going to use.

One key challenge when designing chatbots and conversational interfaces is that it can be hard to grasp what the AI is capable of. This lack of proper understanding can further lead to misleading design or design that does not address all that needs to be addressed. This is strengthened by the paper of Yang et al. where they write that “designers frequently report that it is difficult to grasp what AI can or cannot do”(Yang et al., 2020). When the design does not correspond with the capabilities of a given AI, the user experience will be influenced negatively. It won’t matter how great your AI is if the users do not enjoy using it.

Another key challenge when designing chatbots and conversational interfaces is to create good design despite the adaptive character of AI. Machine learning algorithms are always learning, which means that our interaction with the AI today, may be different from our interaction with the same AI a week before, and then again, a week before that. Yang et all writes that when system outputs that cannot be simulated, the system design is at its most complex (Yang et al., 2020). How do we design for a system when we don’t even know how it will evolve?

I believe that adherence to the first and second guidelines for human-AI interaction could help resolve these challenges. The first guideline, “make clear what the system can do” is intended to clarify to the user what the capabilities of the AI is (Amershi et al., 2019). This guideline directly addresses the first challenge above. Making the systems capabilities clear can make it easier for designers to grasp what the AI can and cannot do. This will make the design process easier and likely also lead to a more pleasant user experience.

The second guideline states “make clear how well the system can do what it can do” (Amershi et al., 2019). This guideline could help improve both of the challenges above. In regard to the second challenge, it does not predict how the system response will evolve an or how the machine learning will make it grow. However, it could give some guidelines as to what the possibilities are by narrowing down the landscape of possibilities to a handful of paths instead.

References

About Us | Centre for Excellence in Universal Design. (n.d.). Retrieved September 9, 2021, from http://universaldesign.ie/about-us/

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). *Guidelines for Human-AI Interaction*. 1–19. https://doi.org/10.1145/3290605.3300233

Artificial Intelligence. (n.d.). Merriam-Webster. Retrieved September 2, 2021, from https://www.merriam-webster.com/dictionary/artificial%20intelligence

Au Yeung, G. (n.d.). *Smart Baduanjin*. Retrieved September 9, 2021, from https://geneauyeung.com/project/0449b6a9/Smart-Baduanjin

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

Biddle, S., Ribeiro, P. V., & Dias, T. (2020, March 16). *TikTok Told Moderators: Suppress Posts by the "Ugly" and Poor*. https://theintercept.com/2020/03/16/tiktok-app-moderators-users-discrimination/

Britannica. (n.d.). *Robot | Definition, History, Uses, Types, & Facts*. Retrieved September 2, 2021, from https://www.britannica.com/technology/robot-technology

Burns, E., Laskowski, N., & Tucci, L. (2021, August). *What is Artificial Intelligence (AI)? - AI Definition and How it Works*. SearchEnterpriseAI. https://searchenterpriseai.techtarget.com/definition/AI-Artificial-Intelligence

Dillenberger, D. (2018, May 23). *Verifier Pairs AI with Optical Scanning for Real-World Product Authentication*. IBM Research Blog. https://www.ibm.com/blogs/research/2018/05/ai-authentication-verifier/

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

IRobot Home App | iRobot. (n.d.). Retrieved September 2, 2021, from https://www.irobot.com/irobot-home-app

Kocielnik, R., Amershi, S., & Bennett, P. (2019). *Will You Accept an Imperfect AI?: Exploring Designs for Adjusting End-user Expectations of AI Systems*. 1–14. https://doi.org/10.1145/3290605.3300641

Microsoft. (n.d.-a). Artificial Intelligence. *Microsoft Research*. Retrieved September 2, 2021, from https://www.microsoft.com/en-us/research/research-area/artificial-intelligence/

Microsoft. (n.d.-b). *Cloud Computing Services | Microsoft Azure*. Retrieved September 2, 2021, from https://azure.microsoft.com/en-us/

Ministry of Local Government and Modernisation. (2020, January 14). *The National Strategy for Artificial Intelligence* [Plan]. Government.No; regjeringen.no. https://www.regjeringen.no/en/dokumenter/nasjonal-strategi-for-kunstig-intelligens/id2685594/

Nielsen, J. (2020, November 15). *10 Usability Heuristics for User Interface Design*. https://www.nngroup.com/articles/ten-usability-heuristics/

Orlowski, J., Harris, T., Seibert, J., & Richardson, B. (2020, September 9). *The Social Dilemma* [Documentary, Drama]. Exposure Labs, Argent Pictures, The Space Program.

The UX of AI - Library. (n.d.). Google Design. Retrieved October 15, 2021, from https://design.google/library/ux-ai/

Thrun, S. (2004). Toward a Framework for Human-Robot Interaction. *Human-Computer Interactions*, *19*, 9–24. http://dx.doi.org/10.1207/s15327051hci1901&2_2

What is Universal Design | Centre for Excellence in Universal Design. (n.d.). Retrieved September 9, 2021, from http://universaldesign.ie/What-is-Universal-Design/

Yang, Q., Steinfeld, A., Rosé, C., & Zimmerman, J. (2020). Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3313831.3376301

Appendix

Feedback 1

In the feedback I received I was appraised for my structure, use of sources and reflections, in addition to my summary of Thrun's article. I kept this in mind while writing iteration two, asking myself how the structure was, whether the flow of felt natural and if the reflections and definitions were thorough enough.

I have also added a bit more to the paragraph regarding Nielsen's 10 Usability Heuristics. This is because the final feedback I got was that it was not entirely clear to the reviewer what these guidelines actually entailed. I hope this is a bit clearer in this updated version.

Feedback 2

These are the "stars" I received for module 2:

1. I think the changes you have done to module 1 are well, and your description of Nielsens Heuristic guidelines are short and well formulated.
2. Overall good structure, which makes it easy to follow along. I also think that you generally explain enough to make it easy to understand what you are describing.
3. I especially like all you write about Siri, good descriptions of both functionality and reflections that makes it easy to understand your thoughts and the overall issues with Siri.

These are the "wishes" I received for module 2:

1. I am not entirely sure I understand what you are saying I the section with main take-aways from Kocielnik et al. I am struggling to understand what type of result is the best. Many relevant or few irrelevant instead of many of both? Maybe use some of the terms regarding this that Kocielnik et al uses in their article.
2. You are writing in your section on Bender et al about biases in big datasets. I would like to see a deeper explanation on what type of biases it is talked about.

I have made changes to the section about Kocielnik and the section about Bender. I have tried to be clearer in my formulations and used more of the terms that was used by the authors in their articles.