Specialization in Research in Design of IT

IN5480 Individual assignment fall 2020

Module 1

Concepts, definition and history of AI and interaction with AI

History of AI

The term *artificial intelligence (AI)* was coined by American mathematician and logician John McCarthy in 1956. He first used the term when he called for participation in a workshop regarding the topic (Grudin, 2009).

While John McCarthy was the first person to use the term, the topic of AI had been discussed in earlier years. The British mathematician and logician Alan Turing stated his opinion on the topic when writing in the *London Times* in 1949, where this statement would later become sensational. He shared that he did not see any reason why computers should not enter any of the fields normally covered by human intellect (Grudin, 2009).

Definitions of AI

AI has been proven to be a vast topic, and people do not always agree on the definitions of the term. The broader definition is often described as an approach to simulate intelligent behavior using technology, however it gets a bit trickier when going into detail. Despite this, there have still been attempts to pinpoint what exactly an AI is. Some of the many definitions for AI is has been described as:

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

John McCarthy had previously coined the term, and redefined it in 1998. He puts emphasis on it being especially applicable for intelligent computer programs, simultaneously stating that its intelligence is not necessarily confined to what's biologically observable.

2. *Artificial intelligence (AI) is a wide-ranging branch of computer science concerned with building smart machines capable of performing tasks that typically require human intelligence. (Built In, n.a.)*

Built In is an online tech newspaper, and much different from McCarthy, describes the term as machines that are capable of performing tasks that typically require human intelligence. This definition restricts artificial intelligence to human-like abilities or intelligence.

3. *Artificial intelligence enables computers and machines to mimic the perception, learning, problem-solving, and decision-making capabilities of the human mind. (IBM, 2020)*

Similarly to Built In, IBM defines AI as a machine with human-like abilities, and expands on what these abilities include.

I believe that in order for me to define AI, I have to break down what the two words in the term mean by themselves. In the Cambridge Dictionary, artificial is described as “made by people, often as a copy of something natural” (Cambridge Dictionary, 2020a), while intelligence is described as “the ability to learn, understand, and make judgments or have opinions that are based on reason” (Cambridge Dictionary, 2020b). In my opinion, humans are not the only beings on earth to have these abilities, albeit being the most advanced at them. In my own attempt to define AI, I resonate the strongest with McCarthy’s statement as his statement does not limit AI to what’s biologically observable, but I want to expand even further on the limits. I believe that a machine that mimics other animals whose intelligence equals that of Cambridge Dictionary’s description of intelligence can also be regarded as AI. In summary, my definition of AI is a machine that can mimic intelligent behaviour of different creatures by learning, understanding and making judgements like them.

AI in Facebook Inc

Facebook Inc is a multinational Internet corporation, and owns Facebook, a social media platform with 2.7 billion monthly active users (Statista, 2020). While being most known for this platform, Facebook also has a research group within the field of AI called Facebook AI Research (FAIR). One of FAIRs work includes AI Habitat, which is a simulation platform for research in embodied AI[1]. On their website, AI Habitat is presented as a service used to aid research within embodied AI. They aim to use this research train and test their embodied AI agents in simulation before applying this in the real world (AI Habitat, 2019).

AI in Portal and Portal 2

Portal is a franchise of puzzle-platform video games consisting of two games: Portal and Portal 2. The game revolves around the protagonist Chell, who is the only human you get to meet through the entire game. Chell is held hostage for the entertainment of a malicious AI called GLaDOS at Aperture, a former science facility center. Chell's goal throughout the game is to abide by GLaDOS' orders to go through different test chambers in hopes of eventually being released. As you progress through the story, you learn that she was created by scientists working at Aperture with the intention of making an AI capable of human emotions.

GLaDOS is portrayed as an entity lacking conscience almost throughout the entire series, showing no mercy towards her victims. As many other fictional stories, Portal is a tale of the dangers of an AI exceeding the intelligence and lacking the morals of a human. The game raises a lot of ethical questions, such as "will we reach a point where AI becomes a danger to humans?" and "should AI be able to truly mimic human emotions?"

Robots and AI systems

Origin of robots

The word *robot* originates from the Czech word "robota", which translates to "forced labor". It first appeared in Karel Čapek's play R.U.R. (Rossum's Universal Robots) in 1920. It was his brother, Joseph Čapek, who had come up with the term. In the play, "robot" is used to describe human-like machines (Lexico, 2020).

Definitions of robots

The Cambridge Dictionary defines a robot as "a machine controlled by a computer that is used to perform jobs automatically" (Cambridge Dictionary, 2020c). This definition removes humans out of the equation, and is the part that differentiates robots from other types of

[1] Embodied AI is the study of intelligent systems with a virtual or physical embodiment, such as robots and personal assistants.

machines. The controller of a robot is said to be a computer instead of the typical human operator. It also says that the jobs should be performed automatically.

NASA describes robots as “a machine that is built to do a certain job again and again, or to do work that might be dangerous for humans” (May, 2020). Contrary to Cambridge Dictionary’s definition, this definition does not state whether or not a computer has to be fully independent in order to qualify as a robot. Instead, it mentions repetitiveness in tasks and its purpose.

Personally, I find NASA’s definition too specific and limiting to agree with. I do not believe that a robot is always made with the intention to do work that could be dangerous for humans. While it is indeed sometimes the case, I also see robots being used for simplifying and reducing tasks for humans and other times simply for entertainment. My own definition of a robot is a physical machine that can perform tasks independently without the assistance of humans. This description is much broader and covers what I envision as robots.

Relation between AI and robots

While AI and robots might seem very similar at first, they both have qualities that differentiate the two. I do however not think that one should exclude the other. In my earlier definition of AI, I described it as a machine with the ability to learn, understand, and make judgments or have opinions based on reason. Here, I did not specify if the AI had to take shape in a physical form. For robots, I defined them as physical machines that can perform tasks independently without the assistance of humans. This means that while not necessary, a robot can contain an AI. Science fiction often portrays robots as intelligent. Both also require programming in order to operate and are artificial, which is why they can often be mistaken as the same thing.

Robotic vacuum cleaners

Robotic vacuum cleaners are commercial automatic vacuum cleaners, and have found home in increasingly many households since the mid 2000s. They are designed to have wheels to move around the room easily. With today’s technology, most of these robots have sensors that allow them to change direction when approaching an obstacle.

Universal Design and AI systems

Definition of Universal Design

Centre for Excellence in Universal Design defines 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.” (Centre for Excellence in Universal Design, n.a.) I chose this definition since it reflects my opinion on what universal design is. Moreover, I want to add ethnic background into the list. Universal design is all about inclusion and is a measure to prevent discrimination of people with different qualities. Especially people who are part of a minority will often meet different challenges throughout their day. By adding universal design into our services and everyday products, we contribute to removing an obstacle in their life.

The Potential of AI

Implementing AI into robots can start a new way of interaction between humans and robots. A robot that mimics or recognises human perception, movement, cognition and emotions could possibly bond with humans in need of therapy. The health sector in Norway is struggling to find enough nurses for retirement homes and are not equipped for the age wave that is rapidly approaching. Using AI-implemented robots could benefit both the workers in the health sector and the elderly people in need for care. Many elderly people can often feel lonely in retirement homes. In the case of people with dementia, a lot of them feel angry, confused and lost. Having specialised robots with the intelligence to understand their emotions and heal their pain could help reduce the work tasks of a nurse, which in turn would allow the nurses to shift their attention to other tasks.

AI also has the potential to include people. For instance, using an image recognition tool could help users understand images on web services. On the other end, there is also potential to exclude people with AI. Some facial recognition softwares have a hard time recognising people with darker skin tones. This is usually due to the AI being trained with limited sample data. This is a problem that researchers and creators of AI should address to make sure that their AIs are inclusive and non-offensive to users.

Guideline for Human-AI interaction

Match relevant social norms

Microsoft proposes 18 design guidelines for human-AI interaction. Guideline number 5 is to “match relevant social norms” (Microsoft, 2019). This guideline is relevant during interaction with the AI and means that the AI should act in a way that is expected according to the user’s social and cultural context. This guideline has similarities to one of Jakob Nielsen’s 10 heuristics:

> *Match between system and the real world: the system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order. (Nielsen, 1994)*

I chose this heuristic as it mentions that the system should behave according to the real world and create familiarity to the user. This could be achieved by matching social norms. Additionally, I have interpreted the word “language” to not be restricted to sound, but to include hand movements and body language - both of which can differ in various social and cultural contexts.

HCI design guidelines and the Human-AI interaction guidelines

Nielsen’s 10 heuristics and Microsoft’s guidelines for human-AI interaction both take into account the users of the systems. They facilitate the user’s behaviour and needs, inclusion and matching the systems with the real world. Both also put importance on providing clear and efficient information when something has gone wrong. What mainly differentiates the two is that Nielsen’s heuristics focus on what should be provided for the users, while Microsoft’s guidelines is more about how the system should behave.

Module 2

Characteristics of AI-infused systems

In Asbjørn Følstad's lecture, he identifies four key characteristics in AI-infused systems: *learning, improving, black box* and *fuelled by large data sets.*

Learning

Learning entails that AI-based systems dynamically learn and adapt through information based on a user's behaviour. These systems usually come with a lot of flaws upon launch, and they can use this learning characteristic to improve themselves.

Improving

AI can improve by processing and learning new behaviour from continuously receiving new information. This characteristic goes hand in hand with learning and is important for the AI in order to perform tasks better over time as they will learn from mistakes and expand their understanding of the user.

Black box

The term *black box* is used to describe the sides of AI that users do not typically get to see or understand: how the AI processes and outputs the data. This can create a confusing experience for some people, especially if they are not familiar with interacting with AI-infused systems. Kocielnik suggests a solution to this, stating that providing explanations will lead to a better understanding of how the AI system works.

Fueled by large data sets

An AI-infused system is fueled by large data sets. It collects data from its interaction with users and uses this information to learn and improve itself. It is however also possible for the system to overlearn itself on this information to the extent where it starts doing or saying things that the creators did not originally intend for it to do. Yang et al. (2020) describe that since there is uncertainty associated with what the AI can do, it can be difficult designing user experiences.

Google Maps

Google Maps is a web mapping service provided by Google, and is an example of an AI-infused system. It offers detailed information about geographical regions and sites around the world, and uses AI to predict upcoming traffic, estimate travel time, and find the most optimal routes from a destination to another (Lau, 2020).

Google Maps uses three key characteristics in AI-infused systems to help its accuracy in predicting traffic: learning, improving and fueled by large data sets. The system analyses historical traffic patterns for roads, and learns and improves from the data sets over time. The more data the system collects, the more accurate the predictions become. The system can also be described as a black box as the user does not get to know how the predicted traffic was calculated or what data was used for this purpose.

Human-AI interaction design

Amershi et al. state that the existing guidelines are not tailored for design of human-AI interaction, and discuss how AI can create new challenges when designing user interfaces. To solve these challenges, the authors propose 18 design guidelines for human-AI interaction, categorised into four different categories based on when the interaction takes place: initially, during interaction, when wrong and over time. Researchers, designers and usability practitioners collaborated through four iterations to test and define the guidelines. These are intended to be resources for working with AI-infused services and systems, and for further research within the field of AI (Amershi et al., 2019).

The paper by Kocielnik et al. explores “techniques for shaping end-user expectations of AI-powered technologies prior to use and study how that shaping impacts user acceptance of those technologies.” They use a scheduling assistant which uses AI to automatically suggest meeting requests in emails after it detects meeting spots, dates and times. The study explores two different versions of the same system with different focuses on which errors to avoid. The findings show that our expectations affect how good we think a system is, and that clarifying in advance where the system might go wrong can make users more open to accept an imperfect AI (Kocielnik et al., 2019).

Google Maps and the human-AI interaction guidelines

Based on the design guidelines proposed by Amershi et al., I will describe how Google Maps adheres to the guideline 2 and 12.

Guideline 2: Make clear how well the system can do what it can do.

The second guideline revolves around letting the user know what the system can do, and helping the user understand how often the AI may make mistakes (Amershi et al., 2019). An example of where Google Maps uses this guideline can be seen in figure 1, where the user is presented with a message informing the user that the conditions may differ from their predictions. The text is however only available after the user chooses to see more details about the route by clicking on a link. The font here is also smaller than all of the rest of the page, as well as being gray in colour. This makes the message less accessible to users.

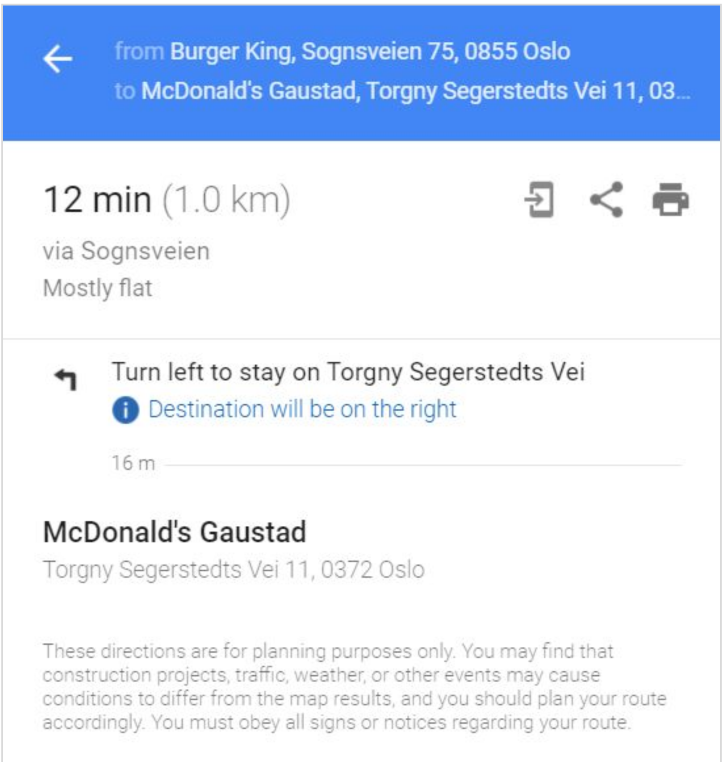

Figure 1: Google Maps lets users know that the conditions may differ from their prediction.

Guideline 12: Remember recent interactions.

The twelfth guideline entails maintaining a short-term memory and allowing the user to make efficient references to that memory (Amershi et al., 2019). While typing in the address bar in Google Maps, the user will be shown previous addresses that they have searched for.

Chatbots/controversial user interfaces

Challenges in the design of chatbots/conversational user interfaces

Følstad and Brandztæg (2017) list several challenges in the design of chatbots or conversational user interfaces. One of the key challenges is designing conversations. Graphic and visual design is less important in this area, while conversations and words are crucial for a chatbot. This makes the conversation a design object, where the designers have to focus on what the chatbot should say, what it should respond to, and the flow and the limit of their conversations. The user also has more responsibility on their end, as they have to know what to say to get the right responses from the chatbot.

Another challenge is the need to change the focus from UI design to service design. Følstad and Brandztæg state that “we need to move from seeing design as an explanatory task to an interpretational task". While there used to be focus on making everything comprehensible for the user, the attention has shifted to understanding the users and their needs.

Yang et al. (2020) also discuss about the challenges around AIs. One of the topics they present in the article tackle challenges in understanding AI capabilities. AIs are capable of changing and adapting as it learns, and as a result of this, it is difficult to anticipate what the system can do, when it will fail. This makes it hard to design suitable interactions for these scenarios.

Guideline 1: Make clear what the system can do.
Guideline 2: Make clear how well the system can do what it can do.
Guidelines 1 and 2 in Amershi et al. (2019) propose solutions for some challenges. Clarifying in advance what the system can and can not do minimises the amount of conversations a designer has to design. It also makes the experience less frustrating for users, as they are more open to accept an imperfect AI if they know what to expect from it (Kocielnik et al., 2019). These guidelines do however not solve the challenge to change focus from UI design to service design. With these guidelines, the designer still has to focus on the users and their needs.

Module 3

Human AI collaboration

Mental models help humans understand how systems function and make predictions about future system behavior, and are crucial to how a robot will be perceived (Rouse and Morris, 1986). Phillips et al. (2016) discuss the need to utilise robots as teammates instead of tools in situations where it is difficult to operate a robot *and* maintain situation awareness, and how our knowledge of human-animal teams can be used to help create better collaboration between humans and robots. Robots that are inspired by these relationships may improve interactions between humans and robots as they take inspiration from well-established mental models and human tendencies. In order to mimic animals, the robot needs to have a certain degree of autonomy. The article *Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy* by Schneiderman (2020) mentions Sheridan and Verplank's approach to evaluate the autonomy of computers through ten levels of autonomy.

AI has become an important field in human-robot interaction, however, users want greater transparency through explainable AI (XAI) systems. XAI is an AI whose actions can be easily understood and analysed by humans (Hagras, 2018).

PARO Therapeutic Robot

PARO is a robot modeled after a seal, and was designed to alleviate depression by providing companionship to elderly people. The appearance of the seal has been important to the purpose of the seal, as it has been found to impact trust within patients (Phillips et al., 2016). It is used as animal therapy in environments such as hospitals and extended care facilities, and has been shown to reduce stress, improve relaxation, socialisation and motivation. PARO uses five different sensors to perceive people and its environment: tactile, light, audition, temperature, and posture sensors. These sensors make it possible for it to recognise speech (although very limited), move its body and make sounds.

Using the ten levels of autonomy by Sherian and Verplank, I interpret PARO as level 8 on the scale: the computer informs the human only if asked. I chose this level as PARO is able to learn to behave in a way the users prefer, but is reliant on humans as an input. The robot will learn from and respond to our commands, and is otherwise not its own being outside of that.

By increasing the level of autonomy, PARO would behave more independently like an animal. The 10th level is described as “the computer decides everything and acts autonomously, ignoring the human.” PARO would then have its own wants and needs, and this level of automation could both be an advantage and a disadvantage. The advantage is that PARO would feel more life-like, and earning their trust would feel more rewarding and reciprocating. This could be a form of therapy in itself, knowing that a sentient being chooses to bond and stay with you. The disadvantage, however, is that the robot’s behaviour becomes unpredictable. They may leave their assigned tasks and become less available, or they may choose to abandon their patients completely. They can also become malicious, by physically hurting the patients or being difficult

I would regard PARO as an XAI as it displays predictable behaviour and adaptability towards humans. I would also argue that this is the needed explainability for this kind of robot, as the patients might struggle with memory loss, fatigue, or other mental or physical disabilities.

BigDog

BigDog was developed by Boston Dynamics, and was a legged robot that walked, ran, and carried heavy loads for military purposes (Pillips et al. 2016). BigDog navigated rough terrain, otherwise not accessible by wheeled or tracked vehicle, using sensors and its control system to stay upright and move its legs. The robot was controlled by a human operator who provided steering and speed input. This project was eventually shelved as BigDog was considered too loud for combat.

In the ten levels of autonomy, I would choose to define the levels twice for BigDog: one for its ability to keep its balance, and the other for its navigation. For the first point, I would define BigDog as level 10: “the computer decides everything and acts autonomously, ignoring the human.” This is because BigDog does not rely on human input to keep its balance and does this automatically. For the navigation of BigDog, I regard it as level 1: “the computer offers no assistance; the human must take all decisions and actions.” This is because of how BigDog cannot navigate without a human operator telling it where and how fast to go.

By decreasing the level of autonomy on BigDog's balancing abilities, it would be extremely difficult for humans to operate the robot. The operators would have to learn when the robot would be close to tipping over, which would require taking many factors into consideration and training.

By increasing the navigation of BigDog to level 5 (the computer executes that suggestion if the human approves), BigDog could help operators by providing suggestions to routes and have the human approve it. This would leave room for the operators to evaluate the routes but also help operators find alternative routes that might not otherwise have been suggested, which could save time for the operators. However, by increasing the level to 10, BigDog could ensue chaos. Executions of military operations are crucial to the outcome, and BigDog would leave too much uncertainty, instability and unpredictability to be part of these operations.

I would categorise BigDog as an XAI as its shape suggests what it can be used for. It is relatively small compared to other vehicles and independently moving legs, suggesting it can take routes that would otherwise not be accessible by traditional vehicles. At the same time, it has the shape of a dog, which could suggest aid.

References

AI Habitat (2019). *About AI Habitat*. Available at: https://aihabitat.org/ (Accessed: 09. September 2020)

Amershi, S., et al. *In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (paper no. 3)*. ACM.

Built In (n.a.) *What is Artificial Intelligence?* Available at: https://builtin.com/artificial-intelligence (Accessed: 09. September 2020)

Cambridge Dictionary (2020a) *Artificial.* Available at: https://dictionary.cambridge.org/dictionary/english/artificial (Accessed: 09. September 2020)

Cambridge Dictionary (2020b) *Intelligence.* Available at: https://dictionary.cambridge.org/dictionary/english/intelligence (Accessed: 09. September 2020)

Cambridge Dictionary (2020c) *Robot.* Available at: https://dictionary.cambridge.org/dictionary/english/robot?q=robots (Accessed: 09. September 2020)

Centre for Excellence in Universal Design (n.a.) *What is Universal Design.* Available at: http://universaldesign.ie/What-is-Universal-Design/#:~:text=Universal%20Design%20is%20the%20design,%2C%20size%2C%20ability%20or%20disability. (Accessed 09. September 2020)

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

Lau, J. (2020) Google Maps 101: How AI helps predict traffic and determine routes. Available at: https://blog.google/products/maps/google-maps-101-how-ai-helps-predict-traffic-and-determine-routes/ (Accessed: 21. October 2020)

Grudin, J. (2009). AI and HCI: Two Fields Divided by a Common Focus. *AI magazine, 30* (4), p. 48-57.

Hagras, H. (2018). Toward Human-Understandable, Explainable AI. Computer, 51(9), 28-36. https://doi.org/10.1109/MC.2018.3620965

IBM Cloud Education (2020) *Artificial Intelligence (AI).* Available at: https://www.ibm.com/cloud/learn/what-is-artificial-intelligence (Accessed: 09. September 2020)

Lexico (2020) *Robot.* Available at: https://www.lexico.com/definition/robot (Accessed: 09. September 2020)

May, S. (2020) *Robot.* Available at: https://www.nasa.gov/audience/forstudents/k-4/dictionary/Robot.html (Accessed: 09. September 2020)

McCarthy, J. (1998). What is Artificial Intelligence? Available at: http://cogprints.org/412/index.html (Accessed: 09. September 2020)

Microsoft (2019) *Introduction to guidelines for human-AI interaction*. Available at: https://docs.microsoft.com/en-us/ai/guidelines-human-ai-interaction/ (Accessed: 09. September 2020)

Nielsen, J. (1994) *10 Usability Heuristics for User Interface Design.* Available at: https://www.nngroup.com/articles/ten-usability-heuristics/ (Accessed: 09. September 2020)

Phillips, E. K., et al. (2016). Human-Animal Teams as an Analog for Future Human-Robot Teams: Influencing Design and Fostering Trust. *Journal of Human-Robot Interaction*, 5(1), 100-125. https://doi.org/10.5898/JHRI.5.1.Phillips

Rouse, W. B., & Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin*, 100(3), 349-363. Doi: 10.1037/0033-2909.100.3.349

Ben Shneiderman (2020) Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy, *International Journal of Human–Computer Interaction*, 36:6, 495-504. Doi: 10.1080/10447318.2020.1741118

Statista (2020) *Number of monthly active Facebook users worldwide as of 2nd quarter 2020*. Available at: https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/ (Accessed: 09. September 2020)

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

Appendix A

Feedback: first iteration

For my first iteration, I received feedback that both addressed the strengths of the paper and the areas in which it could be improved. My peer suggested I make my definitions of AI and robots clearer, as I had only discussed the topics and not made a clear conclusion. For the second iteration, I forgot to take this feedback into account. By the third iteration, I have added a clear definition to them both by summing up my thoughts into one sentence.

Feedback: second iteration

After my second iteration, I received feedback that I had forgotten to add a section about the feedback from the first iteration. I had also forgotten to make changes to my paper to improve it based on the feedback. Lastly, my peer suggested that I also mention some examples from Yang under the chapter “Challenges in the design of chatbots/conversational user interfaces”, as Yang also had good points about the topic. For the third iteration, I added one of Yang’s challenges under this section and describe why this is a challenge.