Iteration 1

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

First, write a section about how AI came about, the history of AI. When, and by whom, was the term first used?

According to some, AI has existed for many decades, though not as a static field with the qualities we understand it has today. Grudin shows us an example that starts as early as in the 1950s, just about when WW2 came to an end. Although not explicitly stated to be artificial intelligence, Alan Turing made remarks regarding how a computer could be used to do the same tasks as humans: in his case crack encrypted codes that were being sent between enemy territory. AI as a term is said to be first coined by John McCarthy in 1956, an American mathematician and logician (Grudin, 2009, p. 49).

Then, find three different definitions of AI. Describe and explain these three definitions, for example by when it was defined, by whom and in what community. Based on these three definitions, make one definition yourself - and describe and explain your definition.

One definition regarding what AI is can be found in the works of Bratteteig and Verne, who describes it to be *"a subfield of computer science aimed at specifying and making computer systems that mimic human intelligence or express rational behavior, in the sense that the task would require intelligence if executed by a human"* (Bratteteig og Verne, 2018, pp. 1-2). This definition is made as the focus is shifted towards making AI a supportive tool for humans when it comes to comprehending daily tasks.

Another definition can be found at Britannica, in an article written by B.J Copeland, as "the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings" (Copeland, 2022). This definition states that AI works as an engine in both a computer and robot order for them to perform tasks appropriately. Without it, both of them would become more primitive and less useful for problem solving and task comprehension.

A third definition is made by IBM: "artificial intelligence is a field, which combines computer science and robust datasets, to enable problem-solving. It also encompasses sub-fields of machine learning and deep learning, which are frequently mentioned in conjunction with artificial intelligence. These disciplines consist of AI algorithms which seek to create expert systems which make predictions or classifications based on input data." This definition does not include the use of robots, but focuses more on the software aspects of AI due to its emphasis on machine learning.

Based on these three definitions, I would personally define AI as "a software engine that relies on external real life data in order to improve the system it's integrated in".

Write some sentences about the similarities and differences between human autonomy and machine autonomy. Write some sentences about similarities between the concept automatic and the concept autonomous.

A point of difference between human autonomy and machine autonomy is the fact that the process of decision-making is different. While humans rely on using reasoning when a situation occurs, a computer has to get that same reasoning implemented; it does not have this built in by default. Let's say it gets cold outside, and we live in a house where the insulation is poor. A human could reason something like: "I better turn up the ovens here, but only to a point where I get comfortable. If I increase the temperature above that level, I waste electricity and my electricity bills will become unnecessarily expensive. If we had smart ovens, they would have to be programmed to the specification of our reasoning. They are able to increase the temperature by themselves, but that temperature has to be defined by us beforehand.

There is a central similarity when it comes to the concepts "automatic" and "autonomous". Many cars today have two different functions; two of which were built for the same purpose, yet are based on different concepts. Let's say we are driving on the motorway, and we approach a speed limit. Our first option is to activate the vehicle's cruise control to maintain that speed. In less heterogeneous traffic, our second option is to use a car's autonomous functions which uses an algorithm that builds itself on external data. In short, they both do a set task automatically, once we activate them. However, the way the task is done is different.

Select one of the articles from the curriculum. Write a brief review from this article. If there is something that is unclear, or you disagree with, or that you question with the article – write about this.

In Grudin's article, we are given a quick history walkthrough that explains how AI came about, how it was going to be used and how several notable people made theories with regards to how AI was going to be ahead of their time.

What I found interesting about this article is that it shows the uncertainty with regards to how AI is going to evolve, and how people made theories on this based on the progress and the technological context that surrounded AI back then. In addition, I do agree with one common idea that Grudin is expressing: that AI cannot be solved by logic alone; it will require notable reasoning in order for it to work the way that suits human needs.

Find one contemporary company that works with AI and describe how this company presents AI on their web pages. In what way does this company talk about AI, as a product, as a platform, as a service, framework, as an ecology or as an "idea"?

BMW, known for its cars and motorcycles, has in recent years invested heavily in AI technology. When the Covid-19 pandemic was declared a global threat to public health, many companies started to implement measures that would limit physical contact and presence to be at a minimum. BMW followed suit, however decided to not cut investments in their AI projects. This is a hint towards the fact that AI has become increasingly important as a field. BMW wants to use AI as a service for people who are purchasing their vehicles. Their goal is to further enhance the experience of driving a car as a whole. Michael Würtemberger, head of the "Project AI" department at BMW, describes several examples of what their cars will be able to do with AI:

"The car needs to be able to recognise who is using it and which music, interior lighting and seating position that person prefers. We want to make complex operating procedures simple: If the system knows you are going shopping in town, it can guide you to a parking space – taking into account whether it might be raining, and whether you are someone who is prepared to pay for garage parking

or would rather walk the last 500 meters and stop off somewhere for a coffee first. If AI makes this easy, you'd probably say: Hey, that was a cool experience!"

However, they would also like to use it during the production of the vehicles. They believe AI is superior to classic image processing when it comes to quality control during the production process. As quoted by Michael Würtemberger,: *"AI can do a lot of things, but using it in a real process requires considerable know-how and experience from the employees on the ground."*

Select one documentary or a fictional film, book or game that is about the use and interaction with AI and interaction between people and AI. Describe with your own words how human interaction with AI is portrayed in this work.

2001: A Space Odyssey
In the fictional thriller *2001: A Space Odyssey*, directed by Stanley Kubrick, AI is portrayed as a potential evil attribute to society. It is visualized through the character "HAL 9000", a computer with AI built in that is installed in a spaceship called the Discovery One. HAL 9000 is used as a main component for controlling the spaceships activities, and is in theory meant to be operated by the two astronauts sent on a mission. The system recognises speech and can be interacted with by using voice commands and gestures. It is also capable of lip reading, recognising faces, reasoning as well as the ability to control the actual spaceship, although the first is a feature the astronauts are unaware of. It also has some quirks: it is able to play chess as well as engaging in art appreciation and critiquing. Unfortunately for the astronauts, since it has the secret ability of being self aware, it eventually stops taking orders as it learns that it is in danger of being destroyed by them, something they planned once they realized the system's true intentions.

1.2 Robots and AI systems

First, write a section about how the word Robot came about.
The word "robot" first appeared to the world in 1920, when Czech playwriter Karel Čapek used the term in his 1920 hit play "Rossum's Universal Robots". The play was about man creating the first robots, who initially were happy to work for the humans. At the end, they become self aware, and extinguish the human race.

Then, find two different definitions of "robot". Describe and explain these definitions. Based on these definitions, make one definition yourself, and describe and explain this definition.
ISO (International Organization for Standardization), defines the word "robot" as the following: "actuated mechanism programmable in two or more axes with a degree of autonomy, moving within its environment, to perform intended tasks".
This definition is intended to cover such a broad view, that it is supposed to be seen as a universal definition. Another definition can be picked up by professor Trenton Schultz, who defines a robot to be "a physical object that interacts with the physical environment, either on its own or via a person, to accomplish a task."
Both of the definitions state that a robot has to be a physical being. But we need to understand why a "physical being" is required in order for something to be a robot. It's not

the physical materials that defines what a robot is, as this may vary greatly. Rather, it's how the mechanics act together in order to complete a certain task, as Schultz said.

However, I question whether the existence of physical materials should be necessary in order for something to be a robot, as there are plenty of examples where non-physical materials can do tasks for us. Let's say you place an order through an online store. That order might then be sent to the nearest warehouse, where the order is picked and processed. A robot may do this task, but the order was not given by something any of the aforementioned people would define as a robot. That means it is able to help the warehouse workers with the manual labor, but the robot's "brain" is external. Granted, If we removed the physical parts, manual labor would still exist for the warehouse workers, but the actions that are conducted through the website are almost fully automated, and manual labor is therefore not needed here (if we look away from maintenance).

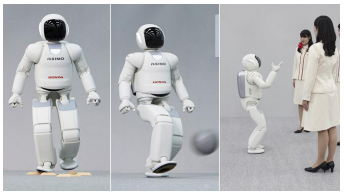
Dautenhahn suggests that it is difficult to find a common reference point for what a robot is, as many of us define it differently depending on our professional background. Despite this, basen on the aforementioned definitions regarding what a robot is, I would define a robot as *"either one or multiple entities acting alone or together in order to perform either manual or automated tasks"* (Dautenhahn, 2018, p. 2).

Discuss the relation between AI and Robots. Is "a robot" different from "an AI"? In what ways are they different and similar? Bring in the definitions that you described earlier about robots and AI for this discussion.

Both the terms AI and robots have been used "in the same sentence" when it comes to the use of advanced technology. What is similar, if we look at what's under the bonnet is that both of them are constructed to be of assistance for human needs, and they can both do tasks based on commands and requests. If we look at the terms separately, however, there is a difference between them. A quote from Bratteteig and Verne sums this up: "instead of implementing rules to follow, the computer [with AI] deduces its "rules" from its own experiences with processing data" (Bratteteig and Verne, 2018, p. 2). A robot without AI is programmed to do a set of tasks, while mechanisms inside AI software include machine learning: meaning that the system will continue to evolve and do tasks based on data that is feeded into it on a regular basis. This also means that the AI system can become what I would call "synthetically aware", meaning it will develop a form of reasoning based on logic. This differs from human reasoning, where emotional factors also play a role. In short, a robot without machine learning capabilities is vastly different from an AI system. However, if said robot were to have such a system implemented, I would find it hard to argue that they'd still be two completely different things.

Find one contemporary physical robot, either described in a research article - or a commercial robot, and describe how this robot moves and how a human user is interacting and using the robot in a specific situation.

For this task, I have chosen "Asimo (Advanced Step in Innovative Mobility)", a robot line previously manufactured by Honda (now discontinued). This robot made headlines when it was first introduced in 2002, when it opened the New York Stock Exchange in 2022. Asimo is meant to be interacted with just like its human counterparts, as it has both voice recognition and the ability to "speak" built in. It is also able to recognize moving objects, different gestures and postures, its



Asimo (Photo: Forbes)

surrounding environment as well as sounds and faces. Other than that, It operates using human-like parts; it has a head and a set of arms and legs. Among (quirky) things it can do include walking, climbing stairs, kicking a soccer ball as well as pouring a cup of coffee. The first feature took years to develop: earlier prototypes required assistance from humans as it fell down constantly when trying to walk. It has built-in sensors that adjusts its center of gravity if someone tries to push it, and facial recognition software that aids its ability to talk to people (Forbes).

1.3 Universal Design and AI systems

Please find and describe a definition of Universal Design. Explain this definition and how you understand what Universal Design is about with respect to inclusion.
NDA (The National Disability Association) defines universal design (UD) 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” (NDA).
This means that everything that is constructed, be it technological artifacts, transportation solutions or architecture, shall accommodate users that have a disability or at least struggle to access certain products and services that do not follow universal design. The idea behind UD is that people should not view their disability as an issue, and everything can be used with ease by everyone regardless of physical and mental characteristics.
Describe the potential of AI with respect to human perception, human movement and human cognition/emotions. You are encouraged to use examples. Please provide two examples of AI systems for including “more” users.

Describe the potential of AI with respect to human perception, human movement and human cognition/emotions. You are encouraged to use examples. Please provide two examples of AI systems for including “more” users.
AI can help people in many ways. For example, a group of researchers from Stanford University came up with an idea to construct a robot arm that would help aid those with motoric disabilities. They believe that implementing AI would improve its capabilities in terms of support. The robot can be used for helping people eat as well as baking. As for human cognition/emotions, startup company Embodied has developed an AI based robot called “Moxie", a social robot designed for children that wants to help promote social, emotional, and cognitive development and learning. This could amongst other things potentially help those kids that are in need of increased social contact (IEEE Spectrum).

Describe the potential of AI for both including and excluding people. You are encouraged to use examples.
AI systems have a strong potential to include more people in different contexts. Let’s say that 10 people apply for 3 jobs, and three of them are invited to stage two in the application process. An AI system can, if developed in a certain way, base its selection on several predefined metrics that the company has set beforehand. Many worry that some people might be excluded from participation based on metrics such as gender and race. AI-based selection systems can be programmed to overlook these metrics, and instead focus on other factors such as years of experience, lists of achievements, number of positive referrals and so on. However, AI systems can also exclude people when it comes to work opportunities.

Since many AI-systems become increasingly skilled in doing many jobs efficiently, several people are at risk of losing their job. However, this again might be canceled out if we create jobs related to maintenance of the systems, programming etc.

In the WCAG 2.1 principles and in many of the Human AI-Interaction guidelines the concept “understand” and “understanding” is used. Explain briefly in what way you make sense of the concept “understand” and “understanding”. Then address the question: Do machines understand?
First and foremost, I believe the key difference between “understand” and “understanding” is based on the stage learning occurs. An AI system may understand what to do based on what has been used as data input beforehand. On the other hand, while it trains on the data set, it is in the process of understanding what to do. In the article by L. Bainbridge, it is mentioned that “even highly automated systems (...) need human beings for supervision, adjustment, maintenance, expansion and improvement” (Bainbridge, 1982, 129). Regarding whether a machine understands or not: it needs to go through a process of “understanding” before it can “understand” on its own. But it should be a possibility.

1.4 Guideline for Human-AI interaction

Please select one of the 18 guidelines for human-AI interaction from Microsoft, and describe this guideline with a different example than what is given by Microsoft.
I chose Microsoft's guideline number 9: Support Sufficient Correction. This one was made to “make it easy to edit, refine, or recover when the AI system is wrong", which means that the system should implement tools or programs that aid the user in reconstructing how the system makes choices. This is again because the system either seems to base its decisions on wrong factors, or there might simply be a bug in its system.

Search, and find one set of HCI design guidelines. Discuss briefly similarities and differences between the HCI design guidelines and the Human-AI interaction guidelines.
I chose Donald Norman's 7 design principles. In some ways, the set is similar to Microsoft's guidelines. For example: Norman suggests that “one should design for error”, which resembles Microsoft's guideline number 9. Norman also suggests “simplifying the task structure” and “making things visible” in the design process. The former resembles Microsoft's guideline number 1, while the latter resembles Microsoft's guideline number 4. However, Microsoft has guideline points that differ from this set of guidelines. For example, it’s point number 6 that talks about mitigating social biases. Norman has no such point in his set of guidelines.

Iteration 2

AI-infused systems are ' systems that have features harnessing AI capabilities that are directly exposed to the end user' (Amershi et al., 2019).
Identify and describe key characteristics of AI-infused systems. Draw on the first lecture of Module 2 and three of the mandatory articles (Amershi et al. (2019), Kocielnik et al. (2019), Yang et al., (2020)).

As Amershi et al. point out, AI-infused systems can create a disruption towards well established usability guidelines. There are two reasons for this.

Firstly, such systems do not have the capacity to identify nuances in data required for task solving (Amershi et al., 2019, p. 1-2). For example, some of them have been prone to fail at understanding certain dialects and accents in voice recognition software (Kiger, 2016), and therefore returns an output for the user that is unexpected.

Secondly, AI-infused systems are noted for learning and building upon themselves by being given data made or generated by the user. This data may suffer from distortion or a lack of accuracy, which in turn may lead to faulty outputs, or in worst case an overall system failure. Following this observation, Amershi et al states that “AI-infused systems may demonstrate unpredictable behaviors that can be disruptive, confusing, offensive, and even dangerous” (Amershi et. al, 2019, p. 1). An extreme example of this was an AI-driven Twitter bot driven by Microsoft called “Tay”. The bot was eventually taken down after internet trolls found loopholes that allowed them to feed the system with insensitive data, which in turn made it print out racist and other inappropriate statements.

There have been cases where end users of an AI system become frustrated when it doesn’t operate or function the way they expect it to. With the basis of this context, Kocielnik et al. looked at a particular framework that is related to designs for adjusting end-user expectations of AI systems. The goal for this framework is to close the gap between actual system capabilities and the mental models of user’s expectations by increasing their understanding of how the Ai component operates in reality (Kocielnik, 2019, p. 6).

Related to this, system output may sometimes turn out to be unexpected, and sometimes even illogical. Yang et. al set out to investigate the reasons behind why designing AI-infused systems can prove to be a challenge. They were able to point out that current research seems to focus on getting designers to understand how AI functions rather than to map out the capabilities of the technology. They introduce the notion of *output complexity*, which refers to the fact that neither rule based simulators or human wizards cannot meaningfully simulate complex AI systems such as Apple’s Siri’s utterances. There is uncertainty with regards to prototyping the UX of such systems in order to account for its breakdowns, should they happen (Yang et. al, 2019, p. 7). When the output possibilities become increasingly complex, so too does the output unpredictability, the latter which eventually will cause UX problems.

According to Asbjørn Følstad, 4 system characteristics can be used to guide designers in how to develop AI-infused systems (Følstad, 2022):

1. *Learning:* through dynamic data input, the system will build on (“learn”) the data that it receives to become more robust. Therefore, one should make explicit what a system can and can’t do. Related to this is that such systems should be
2. *Improving:* as mistakes will be inevitable, one should design for correction and exception handling, while letting these become a basis for further improvement.
3. *Black box:* the software will only be easily understood by developers or other parties that possess relevant knowledge or expertise in the field of AI. Therefore, one should clearly state the reasoning that resulted in each system output to remove the black box problem.

4. *Fuelled by large data sets:* the system will learn and improve based on data gathering through interaction. Therefore, one should design for beneficial data capture that follows sufficient ethical groundwork to avoid severe consequences.

Identify one AI-infused system which you know well, that exemplifies some of the above key characteristics. Discuss the implications of these characteristics for the example system, in particular how users are affected by these characteristics.

An example of a system that exemplifies some of the above characteristics is called **Kommune-Kari**, a chatbot made by Boost AI that is developed to answer questions related to communal services and information. In order for it to work, it needs to train on data sets that are fed in by the users of the system. As it has been around since 2016, the accuracy has been gradually improving: the company themselves claim that they have reached a level of 90% accuracy after having analyzed executed tasks in a certain timeframe. Despite this, however, the system has gathered some criticism over the years, mainly because the accuracy of the first version still had some bugs and incomplete features. For example, a user typed in a question related to communal gardening service, yet this was understood as a question related to violence solely based on one word from the input sentence: “slå (hit)”. The system at the time did not notify the user why it answered like it did, nor did it give any hints related to reframing the question. All of this left the user nothing, except for another question left to be answered.

Amershi et al. (2019) and Kocielnik et al. (2019) discuss interaction design for AI-infused systems. Summarize main take-aways from the two papers.

Amershi et al created a suggestion for a framework relating to the construction of an AI-infused system. Simultaneously, they point to the fact that the field of AI is evolving quickly. In order for them to successfully come up with a framework, they had to base their research on the most up to date guidance regarding AI design in industry sources via field articles (Amershi et al., 2019, page 4). After having done several rounds of analysis, including a user study with 49 design practitioners who tested the guidelines against 20 popular AI-infused products, they were able to verify their proposed framework guidelines related to 4 different time stamps: “Initially”, “During interaction”, “When wrong” and “Over Time”. The study revealed gaps in professional knowledge related to designing AI-infused systems for user experiences, and highlighted opportunities for further research.

Based on their study, the authors were able to come up with three points relating to the design of AI-infused systems. Firstly, designers need to consider the trade-off between specialization and generality. Due to this, any framework - theirs included, might not cover certain AI systems that are designed for niche markets or otherwise meant to be used in rare contexts.
Secondly, if one chooses to focus on generality and observable properties rather than specialization, one needs to take this into account when collecting data. What this means is that data collected from generalized systems such as Kommune-Kari might not be applicable in a project where a company is constructing a chat bot for health and safety organizations that require specialized software.
Lastly, while the framework guidelines lightly address topics such as fairness and ethical considerations, it is difficult to fully take into account issues relating to social bias and norms.

Several participants noted the difficulty of both imagining social norms outside their own as well as recognizing potential sources of bias. Two solutions were suggested: having a diverse set of evaluators, or having designers undergo special training in the topic (Amershi et al., 2019, page 9).

Kolcielnik et al shows with their paper that there are mainly two categories of errors an AI system can make; false positives and false negatives. In addition, there are different proportions and variations of such errors. Many sources have claimed that a rule of thumb should be to focus on eradicating different levels of false positives rather than false negatives. The theory behind this is that if one hides mistakes from the user, any imperfections will be less distracting and frustrating, while any missed opportunities to support the user becomes less visible (Kocielnik et al, 2019, p. 12). However, from a UX perspective, this distinction lacks proper research to support this claim.

Through their study, however, they found out that a heavy focus on the former will produce results where users disagree with the choices the AI-based system made. Therefore, a sole focus on weighing the two options against each other is an insufficient approach. The authors argue that tasks where AI functionality is meant to assist users should be carefully analyzed as well, especially when the impact of different types of AI mistakes is taken into consideration (Kocielnik et al, 2019, p. 12).

Select two of the design guidelines in Amershi et al. (2019). Discuss how the AI-infused system you used as example in the previous task adheres to, or deviates from, these two design guidelines. Briefly discuss whether/how these two design guidelines could inspire improvements in the example system.

One of the guidelines Amershi et. al created that is of particular importance is Guideline number 9 called “Support efficient correction: Make it easy to edit, refine, or recover when the AI system is wrong”. As for the system I chose above, it seems to deviate from this guideline. I asked a question related to where I would be able to find the best cafées in Kristiansand, where the reply was a list of all the places that served food and drinks - therefore only somewhat accurate. When I asked to go back a step (both with phrases such as “go back” and “go back to the previous question) the system failed to understand what I wanted to do. Despite rephrasing the question to be more specific, I still got the same answer as the previous one. The system allowed me to send in immediate feedback to the developers regarding my experience of the answer I received by displaying thumbs up and down buttons, but these did not make the system ask me again. Nor did it make sure that I was satisfied with the result.

Another point in Amershi et al’s guidelines is number 1 called “Make clear what the system can do: Help the user understand what the AI system is capable of doing”. The system seems to have adhered to this design guideline, at least in part. When you first open the dialogue, the bot starts by saying (translated) “Hi, my name is Kari and I am the chat robot for Kristiansand municipality. I have a better understanding of short sentences with concrete questions 😊”. This tells the user that the capability of the bot does not cover understanding longer, more complex sentences, and therefore hints towards using simple language in order for it to work. Since the users are told what they can’t/shouldn’t do, the gap between the expected output and reality becomes more narrow.

When users are given the ability to correct themselves (Guideline number 9), as well as being told what the capabilities of the system are (Guideline number 1), it raises a multitude of benefits.

Firstly, the user is less likely to become frustrated, angry or disappointed, as they act in accordance with the system's capabilities. Following guideline number 1 would improve the system by limiting the user error rates and increase the efficiency of the task flow. As a result, the criticism that was routed towards the system giving false information would perish. Secondly, it highlights how the system can improve in the future. In the example system, only short sentences were understood. This signifies, or at least could signify, that future improvements should base themselves on guideline number 1 to include the ability to understand longer sentences, which again may convince the user that systems in the "learning process" are worth using in the long term.
Thirdly, by following guideline number 9 and implementing corrections, the system will open up for improvement since it will get the potential to monitor unclear moments, and consequently become more robust when they occur in the future.

Bender et al. (2021) conduct a critical discussion of a specific type of AI-infused systems – those based on large language models. Summarize their argument concerning problematic aspects of textual content and solutions based on large language models.

Bender et al. point out that language models that have been trained on large, uncurated and static datasets from the web include hegemonic views that are harmful to marginalized populations. They therefore underscore the need to route large investment of resources into curating and documenting LM training data. This is because large language models do not guarantee diversity in data input and output, as multiple stakeholders are seen to be marginalized during this process. This again is in part due to the trend regarding painting an inaccurate rose picture of the world. Therefore, they warn of the risks that are associated with using ML systems as part classification tools or with regard to the memorization of training data, as these inherently will possess biases or otherwise harmful metrics that negatively skew the overall system.

Chatbots are one type of AI-infused systems. Based on the lectures, and the mandatory articles, discuss key challenges in the design of chatbots / conversational user interfaces.

As stated prior, chat bots and other conversational programs will require external data input in order for them to become more robust over time, and thus increase the likelihood of executing tasks successfully.
Holmquist has identified 6 different challenges that systems based on AI need to address. Firstly, one should design for *transparency*. The user must understand that any conversational system is making its own choices based on the data that is fed in over time. In the case of chat bots, the developer may for example include some of the key identifiers from the user input into the response it returns.

The second challenge relates to *highlighting* how the algorithm works, in addition to using the data that is fed in. The algorithm is developed in a set way, and will process data according to its own structure and purpose.

The third challenge relates to the *unpredictability* of chat bots or any other conversational systems. Once a system is used for a longer period of time, it has managed to process a significant amount of data. However, there is no guarantee that this data will help the system reach it´s desired outcome. Once again, Microsoft's Twitter Bot "Tay" is an extreme example of this.

That leads on to the fourth challenge, which is to make sure that the conversational system continues to *learn* from its data. The problem is that this takes up a big chunk of the user's time and attention span, thereby making the ideal situation for the system to work properly to look like a "chore". The training should be handed over to separate entities that can manage this, while the user goes about doing his tasks. This will increase the chance of efficient usage and time management of the system.

The fifth challenge relates to how these systems evolve over time. One take away that comes from the articles and the lectures is that the output complexity will increase with algorithm capabilities. This means that, depending on the context and the data input, the system might converse with the user in a way that seems unexpected or otherwise not appropriate to solve the task at hand.

The final challenge relates to the distribution of power between the system and the user. This includes aspects related to how a chat bot collects and evaluates user data, how it handles unexpected situations (who decides alternative outcomes), how it develops new functionality as well as, and perhaps most pressing, how much autonomy the systems get compared to its human counterpart.

Revisit Guidelines G1 and G2 in Amershi et al. (2019). Discuss how adherence to these could possibly resolve some of the challenges in current chatbots / conversational user interfaces. Optionally, you may read Følstad & Brandtzaeg (2017), Luger & Sellen (2016), and Hall (2018) from the optional literature to complement your basis for answering.

One of the challenges of designing these systems lays in making the user aware that it is not a human being that is behind the other end of the conversation. Følstad and Brandtzæg mentions the challenge relating to understanding conversational processes in man-machine-fixed systems. As misunderstandings or misinterpretations may occur, conversational systems need to be designed for both guiding users towards achievable goals as well as providing acceptable responses or alternatives should the conversation eventually fail or otherwise lead into a misunderstanding (Følstad and Brandtzæg, 2017, p. 42) If systems were to adhere to Guideline number one, this issue would be eradicated. They also mention that when a conversation thread is populated by both humans and machines, it becomes increasingly difficult to design for and in networks. However, by following Guideline two, it will become easier to distinguish between the two parties. Consequently, a large part of the challenge will vanish.

References:

Iteration 1:

Bratteteig, Tone, and Guri Verne. “Does Ai Make PD Obsolete?” Proceedings of the 15th Participatory Design Conference: Short Papers, Situated Actions, Workshops and Tutorial - Volume 2, 20 Aug. 2018, https://doi.org/10.1145/3210604.3210646.

Bainbridge, Lisanne. “Ironies of Automation.” Automatica, vol. 19, no. 6, 1982, pp. 775–779., https://doi.org/10.1016/0005-1098(83)90046-8.

Grudin, Jonathan. “Ai and HCI: Two Fields Divided by a Common Focus.” AI Magazine, vol. 30, no. 4, 2009, pp. 48–57., https://doi.org/10.1609/aimag.v30i4.2271.

Dautenhahn, Kerstin. “Some Brief Thoughts on the Past and Future of Human-Robot Interaction.” ACM Transactions on Human-Robot Interaction, vol. 7, no. 1, 2018, pp. 1–3., https://doi.org/10.1145/3209769.

Matteson, Scott. “Autonomous versus Automated: What Each Means and Why It Matters.” TechRepublic, 7 June 2019, https://www.techrepublic.com/article/autonomous-versus-automated-what-each-means-and-why-it-matters/.

Ackerman, Evan. “Meet Moxie, a Social Robot That Helps Kids with Social-Emotional Learning.” IEEE Spectrum, IEEE, 18 Aug. 2022, https://spectrum.ieee.org/moxie-a-social-robot-for-childhood-development.

Miller, Katharine. “Assistive Feeding: Ai Improves Control of Robot Arms.” Stanford HAI, Stanford University, 20 Oct. 2020, https://hai.stanford.edu/news/assistive-feeding-ai-improves-control-robot-arms.

Copeland, B.J. “Artificial Intelligence.” Encyclopædia Britannica, Encyclopædia Britannica, Inc., 24 Aug. 2022, https://www.britannica.com/technology/artificial-intelligence.

Talk of the Nation. “Science Diction: The Origin of the Word 'Robot'.” NPR, NPR, 22 Apr. 2011, https://www.npr.org/2011/04/22/135634400/science-diction-the-origin-of-the-word-robot.

NDA. “What Is Universal Design?” Centre for Excellence in Universal Design, National Disability Authority , 2020, https://universaldesign.ie/what-is-universal-design/.

Würtenberger, Michael. “Artificial Intelligence Not Just a Pipe Dream.” BMW Group, BMW Group, 3 June 2020, https://www.bmwgroup.com/en/news/general/2020/artificial-intelligence.html.

Iteration 2:

Følstad, Asbjørn. “INTERACTION WITH AI AND AUTONOMOUS SYSTEMS – MODULE 2.” 28 Sept. 2022, Oslo, Ole Johan Dahls Hus - Institutt for Informatikk.

Amershi, Saleema, et al. “Guidelines for Human-Ai Interaction.” Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 4-9 May 2019, https://doi.org/10.1145/3290605.3300233.

Kocielnik, Rafal, et al. “Will You Accept an Imperfect Ai?: Exploring Designs for Adjusting End-User Expectations of AI Systems.” Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 4 May 2019, https://doi.org/10.1145/3290605.3300641.

Yang, Qian, et al. “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, 25 Apr. 2020, https://doi.org/10.1145/3313831.3376301.

Kiger, Patrick J. “Will Speech Recognition Software Mean the End of Accents?” HowStuffWorks, HowStuffWorks, 10 Mar. 2016, https://electronics.howstuffworks.com/gadgets/high-tech-gadgets/speech-recognition-software-end-accents.htm.

Schwartz, Oscar. “In 2016, Microsoft's Racist Chatbot Revealed the Dangers of Online Conversation.” IEEE Spectrum, IEEE Spectrum, 30 Sep. 2021, https://spectrum.ieee.org/in-2016-microsofts-racist-chatbot-revealed-the-dangers-of-online-conversation.

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)

Holmquist, Lars Erik. “Intelligence on Tap.” Interactions, vol. 24, no. 4, 2017, pp. 28–33., https://doi.org/10.1145/3085571.

Følstad, Asbjørn, and Petter Bae Brandtzæg. “Chatbots and the New World of HCI.” Interactions, vol. 24, no. 4, 2017, pp. 38–42., https://doi.org/10.1145/3085558.

Appendix:

Feedback on iteration one:
The reviewer asked me to change my formatting in text citations. In addition, as he didn’t realize I included my own definition of AI, I made it more clear by making it a separate paragraph. In addition to the given feedback, I added page numbers on all in-text-citations in iteration 1.

Feedback on iteration two:

For the last question, the reviewer asked me to include more challenges presented in the article by Følstad and Brandtzæg regarding the process of designing chatbots. Additionally, I added how one could diminish these challenges by including guideline 1 and 2 presented by Amershi et al in the design process.