# Modul 1

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

### 1.1.1 History of AI

The first work in the field of artificial intelligence was done by Alan Mathison Turing. All modern computers are in essence Turing machines. His published work "Computer Machinery and Intelligence" led to what is known as The Turing Test, which experts used to measure computer intelligence. At this point the word "artificial intelligence" was not used, but the idea of it was introduced. Later, John McCarthy used the word "artificial intelligence" in a workshop he held. AI became mainstream quickly and today; AI is a part of our daily lives. The collaboration between humans and computers continue to evolve (Tableau).

### 1.1.2 Definition

John McCarthy defined Artificial Intelligence in 1955 as "the science and engineering of making intelligent machines" (Stanford University, 2020).

Britannica defines AI as the ability of computer or computer-controlled robot to perform tasks that are commonly associated with the intellectual processes characteristic of humans, such as the ability to reason.

SAS on Artificial intelligence, "AI makes it possible for machines to learn from experience, adjust to new inputs and perform human-like tasks."

Artificial intelligence is to be close cognitively to a human. Since intelligence is defined as being able to learn and solve problems, this is artificial in a way that it is able to replicate a human since it's based on data from humans. Even humans are not able to fully understand the human brain and how it works completely, so AI is also far

from achieving this feature. AIs for example lack emotional intelligence, experience, morals, ethics and forming deep human connections.

### 1.1.3 Differences and similarities between human autonomy and machine autonomy

Autonomous systems are capable of doing tasks automatically without outside control. While the human autonomy is about being able to make decisions based on not only logic, morals, social norms, but also emotions. Humans also won't produce the same "results" every time and it can fluctuate depending on the situation, meaning the human is more independent than the machine. It is similar in the way that they both can execute tasks, a human can drive a car, while an AI can drive a car based on algorithms.

Automation or automatic systems are powerful enough to take over control that used to be done by people (Norman, 1990). However, an autonomous system goes further than pre-set instructions and can adapt to make a decision.

### 1.1.4 Intelligence on tap

This article is about how far artificial intelligence has come. Back in the days it was believed that machines would soon surpass humans, but it hasn't been until recent years that AI really has surpassed humans in intelligence with deep learning (Holmquist, 2017). Having access to huge amounts of data has led to significant results in different areas of AI. Designers need to understand the limitations and possibilities with AI. Being able to replace human-to-human interactions is a limitation that AI has.

The challenges of integrating AI into products and services revolve around both user understanding and the inherent complexities of AI. It's crucial for users to grasp how AI impacts their interactions, which means clear communication is needed so they know what to expect. However, AI's unpredictable nature and the fact that it often functions in ways that even developers can't fully explain add layers of complexity. AI systems continuously learn and improve, which can lead to outcomes that surpass initial expectations, further complicating user comprehension and trust. Managing

these dynamics requires a careful balance between innovation and transparency.

If AI is constantly improving and evolving, how can designers design it better? Is our role to just inform the end-user of the capabilities of the AI? Design the machine so that it gives feedback to the user? How can we give control to the end-user so that it's not too much feedback?

### 1.1.5 Samsung

Samsung incorporates AI directly into their products, especially smartphones. For instance, Android phones have a voice assistant called Bixby and AI-powered cameras. Samsung AI's can optimize battery life on phones and provide several other services.

SmartThings is platform where AI is used to facilitate the automation and integration of smart home devices. Through this AI, it is easier for different devices to work together. Bixby is also an AI platform that can integrate with other applications and build new services on.

Samsung's chatbots are an example of a service that Samsung provides to enhance customer service and provide efficient support.

Samsung provides developers with SDKs and APIs to build, integrate and deploy AI functionalities within their applications. This framework supports the creation of AI solutions across different platforms and devices. These SDKs and APIs form the building blocks of Samsung AI's framework.

Samsung's AI is designed to work within a broad ecosystem of devices and services. Whether it's smartphones, home appliances, wearables and IoT devices, Samsung's AI acts as a central component that links different devices. It creates a connected environment that enhances user experience throughout the entire Samsung ecosystem.

Samsung's idea for the future of AI goes beyond electronic devices and they believe that it can unlock new possibilities in healthcare, automotive technologies and more.

### 1.1.6 Movie, book or game

In Iron Man there is an AI called J.A.R.V.I.S that takes care of the Iron Man suits and internal systems in Stark's buildings. It can understand human speech and respond to it accordingly. It evolves throughout the movies, where it goes from being serious to more understanding as the movies progress. This reflects the idea of AI continuously improving. As of today J.A.R.V.I.S is more advanced than most AI machinery.

## 1.2 Robots and AI systems

In 1979, the Robot Institute of America defined a robot as "a reprogrammable, multifunctional manipulator designed to move materials, parts, tools, or specialized devices through various programmed motions for the performance of a variety of tasks" (Russel & Norvig, 1995).

Meriam Webster's (1993) definition is "An automatic device that performs functions normally ascribed to humans or a machine in the form of a human."

It can be hard to define what a robot is, because robot designs change rapidly (Dautenhahn, 2018), but based on these definitions a robot is machine that's able to perform human tasks in the real world, while an AI doesn't necessarily exist in a physical form and is usually smarter. An AI can be installed inside a robot to perform more cognitively like a human, meaning it's able to learn and adapt. A robot doesn't necessarily need to have an AI installed if it's a robot that has predefined instructions, like some industrial robots. A chatbot would therefore be more of an AI in my definition. The HelpMate robot is an example of a robot that transports food and medications in hospitals that has more direct interaction with people (Thrun, 2004).

## 1.3 Universal Design and AI systems

### 1.1.1 Definition

From Centre for Excellence in Universal Design, “Universal design” is described 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.” From this definition I understand universal design as to be including of all humans, so that their possible limitations will not be a limitation.

### 1.1.2 Potential of AI

There are a lot of technologies that explore the human senses and emotions, like virtual reality or eye-scanning. AI may or may not be as inclusive, people with disabilities might have a hard time using virtual reality depending on what kind of movement it requires.

### 1.1.3 Understand and understanding

“Understand” and “understanding” is being able to make sense of information. In my opinion machines do not “understand” on the same level of a human. Understanding goes beyond logic and can require emotions, which goes beyond of what a machine is able to comprehend. You can argue that a machine, like a human, is able to understand, because it’s fed with previous “experiences” or “information”. The machine is however not fool-proof and I would say it’s easier to feed it with wrong information. If a human has gone through the experience, it’s much harder to fool with wrong information.

## 1.4 Guideline for Human-AI interaction

05 Match relevant social norms.
Ensure the experience is delivered in a way that users would expect, given their social and cultural context.

I chose guideline 5. In this case it’s important to understand which social and cultural context the AI is going to be used. People will interpret information differently depending on where they come from and what background they have. This could for

example be simple things like symbols can be interpreted differently from country to country. Aligning with cultural context and social gives better user experience.

One set of HCI design guidelines are Nielsen's 10 usability heuristics. HCI design guidelines and Human-AI Interaction guidelines are similar in that they both focus on user-centered design. They both serve the users' needs and make it more user-friendly. However, Human-AI Interaction address problems that one might meet with AI's like the need for transparency and communication is important, because of how the AI might be able to adapt and learn quickly.

1.5 Journal – everyday engagement, interaction and use

Writing with a pen and paper was slower, it was harder to rewrite your thoughts, but there was something about writing things by hand that made my mood feel calmer. Having to type with the keyboard was more efficient and it was easier to edit. However, I felt some type of detachment, I could just write without thinking too much about what I was writing. The recording of my voice was like telling a story to someone I didn't know or having some type of presentation. It was faster, but also unstructured.

## Resources


Dautenhahn, K. (2018). Some Brief Thoughts on the Past and Future of Human-Robot Interaction. *ACM Transactions on Human-Robot Interaction*, *7*(1), 1–3. https://doi.org/10.1145/3209769

Holmquist, L. E. (2017). Intelligence on tap: Artificial intelligence as a new design material. *Interactions*, *24*(4), 28–33. https://doi.org/10.1145/3085571

Norman, D. A. (1990). The «Problem» with Automation: Inappropriate Feedback and Interaction, not «Over-Automation». *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *327*(1241), 585–593.

Tableau. (n.d.). The history of artificial intelligence. Tableau. https://www.tableau.com/data-insights/ai/history

Thrun, S. (2004). Toward a Framework for Human-Robot Interaction. *Human–Computer Interaction*, *19*(1–2), 9–24. https://doi.org/10.1080/07370024.2004.9667338

Stanford University. (2020). AI definitions. *Stanford Institute for Human-Centered Artificial Intelligence (HAI).* https://hai.stanford.edu/sites/default/files/2020-09/AI-Definitions HAI.pdf

# Modul 2

## 2.1 Characteristics of AI-infused systems

### 2.1.1 Identify and describe key characteristics of AI-infused systems.

Amershi et. al. (2019) defines AI-infused systems as ' systems that have features harnessing AI capabilities that are directly exposed to the end user'. One characteristic to AI-infused systems is the characteristic of evolving by learning. Systems may respond differently to the same text input over time or behave differently from user to user. Inconsistency can lead to abandonment of AI technology. However, errors are common in AI-infused systems and it's difficult to obtain reliability (Amershi et. al., 2019).

### 2.1.2 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 AI-infused system that I know well is Spotify's music recommendation system. It tailors' playlists to users based on their listening habits and it's continually learning by refining its recommendations. It automatically generates playlists for users, and it's directly exposed to the end user. However, these specially curated playlists can lead to frustration, because it offers less diversity in recommendations and users don't know why those specific songs are recommended. Users lose a sense of control, because their musical experience is dictated by an algorithm.

## 2.2 Human AI-interaction design

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

Guidelines for Human-AI Interaction

This paper highlights 18 guidelines for design aimed at improving the interaction between humans and AI-infused systems. For over 20 years, human-computer interaction (HCI) has proposed guidelines, principles and strategies for designing user interfaces and interaction for AI-systems, but still designers and developers struggle with creating intuitive and effective AI-infused systems. These guidelines were developed to help people understand 20 years of learning in AI-design. The 18 guidelines are used to provide AI-systems that are intuitive, effective and user-friendly, while also taking ethics into consideration and the importance of transparency (Amershi et. al., 2019).

Will you accept the imperfect AI? Exploring Designs for Adjusting End-user Expectations of AI Systems

The article *"Will You Accept an Imperfect AI? Exploring Designs for Adjusting End-user Expectations of AI Systems"* by Kocielnik et. al (2019) investigates how to set appropriate user expectations for AI-powered systems, specifically to improve user satisfaction and acceptance of imperfect AI systems. The study uses an AI-powered Scheduling Assistant (which automatically detects meeting requests in free-text emails) to explore the effects of different techniques for shaping user expectations.

2.2.2 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.

Guideline 13: Learn from user behavior

Spotify's recommendation system constantly tracks and learns from users' listening habits, such as the songs they play, the songs they skip, favorite genres, repeated tracks and different interactions with playlists. Over time the system understands the user's preferences to offer

personalized recommendations. However, the learning process can lead to limitations in diversity of suggestions. Therefore, Spotify could improve their system by giving users the opportunity to adjust their preferences so that users could explore other genres if they'd like to.

Guideline 18: Notify users about changes

Spotify does not explicitly notify users when changes are made in the recommendation algorithm. Every week there's a new playlist generated like "Discover Weekly" without knowing why such tracks are recommended, which can lead to confusion or frustration. Users are left unaware of how the system evolves over time and how their behaviour impacts the recommendations they receive. Spotify could improve their system with more transparency and explain why specific songs are being recommended.

2.2.3 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 langue models.

Bender et al. (2021) argues that large language models (LMs) are trained on vast datasets that often reflect societal biases, such as racism, misogyny, ableism and other forms of discrimination. These biases can manifest subtly. LMs also tend to reproduce and even amplify stereotypes present in their training data. This occurs in various ways such as referring to certain professions as male by default or using language that demeans certain social groups. This can cause harm to certain groups that encounter this biased text.

Another issue discussed is the overconfidence users place in machine translation systems that output fluent but inaccurate translations. The authors cite a case where an Arabic phrase was mistranslated, leading to an arrest, as an example of the real-world consequences of such errors (Bender et. al., 2021).

To mitigate these risks the author, suggest bias detection, better techniques to detect and filter subtle biases in LMs' outputs. The output should also be clearly labelled as machine generated. The data collection should be carefully done, and the researchers should value

sensitive design. This helps identifying whose values are expressed and supported through technology, and how the lack of support might cause harm (Bender et. al., 2021).

## 2.3 Conversational user interfaces to large language models

2.3.1 Conversational interfaces to large language models such as ChatGPT are a recent addition to AI-infused systems. Based on the lectures, and the mandatory articles, discuss key challenges in the design of such conversational user interfaces.

Designers often struggle to fully grasp the capabilities of AI, leading to challenges such as uncertainty around what the AI can do and the complexity of its outputs. This makes it difficult to follow a traditional user-centered design process when crafting human-AI interactions. The double diamond framework, for example, is typically used to define a design goal based on what current technologies can achieve. However, with AI, predicting outcomes is much more difficult. AI's performance evolves as it learns, meaning its strengths and errors are not consistent over time. As a result, introducing AI systems to users becomes more challenging, as both designers and users may find it hard to anticipate how the AI will behave in different scenarios (Yang et. al., 2020).

2.3.2 Revisit Guidelines G1 and G2 in Amershi et al. (2019). Discuss how adherence to these could possibly resolve some of the challenges in current conversational user interfaces to large language models.

Adhering to Guidelines G1 and G2 from Amershi et. al. (2019) can address some of the challenges in conversational user interfaces to large language models.

Guideline 1: Make clear what the system can do

One of the main difficulties in designing conversational user interfaces with large language models is the uncertainty of what the AI is capable of. If designers followed G1 by clearly communicating the system's capabilities to users, this uncertainty could be reduced. Transparency about the system's abilities could also improve trust, making the AI more predictable.

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

LLMs often deliver responses with inconsistent quality, making user interactions more challenging. By following G2, systems could inform users about their confidence or limitations in certain situations, helping to set more realistic expectations and prevent user frustration. For example, displaying confidence levels or signaling uncertainty in responses (e.g., "I'm not entirely sure, but this could be the answer") would allow users to make better judgments when interpreting the AI's output. This approach would help address the problem of fluctuating performance, especially since LLMs can sometimes generate incorrect or misleading information. For example, ChatGPT is prone to generating fictional or erroneous content, but presented to be factual (Skjuve et. al., 2024). Clear communication of the system's confidence would create a more effective collaboration between the user and the AI, allowing users to adapt their actions based on the AI's strengths and potential weaknesses.

References

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., ... & Teevan, J. (2019). Guidelines for human-AI interaction. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (paper no. 3). ACM. (https://www.microsoft.com/en-us/research/uploads/prod/2019/01/Guidelines-for-Human-AI-Interaction-camera-ready.pdf)

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

Kocielnik, R., Amershi, S., & Bennett, P. N. (2019). Will You Accept an Imperfect AI?: Exploring Designs for Adjusting End-user Expectations of AI Systems. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (paper no. 411). ACM. (https://www.microsoft.com/en-us/research/uploads/prod/2019/01/chi19_kocielnik_et_al.pdf)

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)

Skjuve, M., Brandtzaeg, P. B., Følstad, A. (2024) Why do people use ChatGPT. First Monday. (https://firstmonday.org/ojs/index.php/fm/article/view/13541)