Module 1

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

With myths spanning back to ancient times about automations and thinking machines, we can tell that the thought of AI is ancient(*Artificial Intelligence Definitions*, n.d.). But it was after the second world war, the potential of computation was established. In 1950, Alan Turing, a leading code breaker, published "Computing Machinery and Intelligence". The term Artificial Intelligence was coined by Stanford professor John McCarthy in 1955. He defined it as "the science and engineering of making intelligent machines". (*Artificial Intelligence Definitions*, n.d.)

Three different definitions:

- **From UIB:** "Artificial intelligence (AI) is a common description of systems that perform actions in the physical or digital dimension by perceiving their environment, and processing and interpreting huge amounts of information and data. AI systems have the ability to adapt their behavior by analyzing how the environment and conclusions are affected by previous actions." (*What Is Artificial Intelligence?*, n.d.)

- **From Wikipedia:** "Artificial intelligence (AI), in its broadest sense, is intelligence exhibited by machines, particularly computer systems." (*Artificial Intelligence*, n.d.)

- **From IBM:** "Artificial intelligence (AI) is technology that enables computers and machines to simulate human learning, comprehension, problem solving, decision making, creativity and autonomy." (*What Is Artificial Intelligence (AI)?*, 2024)

My definition: "AI is technology that enables computers to simulate algorithms and processes that are found in humans, such as learning, reasoning, and problem-solving."

Both humans and machines are able to make decisions independently based on available information. Conversely, human autonomy is deeply rooted in consciousness, free will, emotions and ethical considerations, which influence decisions in complex and often unexpected ways. With a mindset of computationalism consciousness however, free will and emotions might be caused by predefined algorithms and are not what we have made it out to be.

Both concepts (automatic and autonomous) refer to processes without direct human intervention or control. In both cases a system performs tasks after being initialized often following protocols and algorithms.

The article "Intelligence on Tap: Artificial Intelligence as a New Design Material" by Lars Erik Holmquist offers a detailed exploration of the rapid advancements in AI and its implication for interaction design. He emphasizes that AI will increasingly become accessible to non-experts, likening it to “intelligence on tap”. A key part of the article is about AI´s inability to truly comprehend the world or engage in natural language conversations effectively. The article was written in 2017 and ChatGPT was released in 2022. It highlights the pace of the AI revolution, as just a few years later, we see AI models like ChatGPT capable of remarkably sophisticated natural language processing and generation.

OpenAI presents AI primarily as a platform and a service. They are now also focusing on AGI (artificial general intelligence), with the goal of creating a model that can complete the same tasks a human can. AGI might be an allconsuming new idea that will change the world drastically.

In the film “I, Robot” based on Isaac Asimov´s works, we are introduced to a world where robots are integrated in everyday life. The film explores the complex dynamics between humans and AI. One of the main points of the movie is that the more technology, AI and robots advance, the line between machine and humans may blur, leading us to question the very nature of intelligence, autonomy and humanity itself.

1.2 Robots and AI systems

The word Robot came from a Czech play called “R.U.R” as a critique of mechanization and the way it can dehumanize people. Robot comes from the word “robota”, forced labor, with it´s slavic root, “rab”, slave. (Jordan, 2019)

Definition of robot:

- “A machine that resembles a living creature in being capable of moving independently (as by walking or rolling on wheels) and performing complex actions”. (Stamper, 2024)

- “A robot is an autonomous machine capable of sensing its environment, carrying out computations to make decisions, and performing actions in the real world.” (Guizzo, 2023)

My definition: “A tangible machine that performs tasks by following predefined scripts, rules or responding to sensory input”.

A fundamental difference between robots and AI is that robots handle physical tasks, while AI handles cognitive tasks. However, both operate autonomously and can perform tasks without human control. Integrating AI into robots makes them “smarter” and more adaptable, allowing them to handle more complex tasks and learn from their environment.

Boston Dynamics Spot walks on four legs, mimicking the movement of a four-legged animal. It can perform dynamic movements such as running, jumping and navigating on uneven terrain. Spot has several use cases such as searching disaster sites for survivors, helping law enforcement officers to resolve hostage situations, assisting with drug lab investigations and even conducting structural surveys after fires. (*Spot to the Rescue*, n.d.) Human operators control Spot remotely using a tablet or a computer interface. They can direct the robot's movements, configure its sensors, and monitor live video feeds from Spot´s camera (*Spot to the Rescue*, n.d.)

1.3 Universal Design and AI systems

Definition of Universal Design from Center for Excellence in Universal Design:

“Universal Design (UD) is 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. An environment (or any building, product, or service in that environment) should be designed to meet the needs of all people who wish to use it. This is not a special requirement, for the benefit of only a minority of the population. It is a fundamental condition of good design”. (*About Universal Design*, n.d.)

Universal Design focuses on creating environments and products that are available for everyone, including people with diverse abilities and needs. It is about ensuring accessibility and usability from the start.

AI has the potential to greatly enhance human capabilities in perception, movement, and cognition/emotions. For human perception, AI can improve experiences through technologies like augmented reality. For example, Microsoft HoloLense uses AI to overlay digital information onto the real world. (Roach, 2019). Another example is Apple's Vision Pro that uses AI to automatically enhance the clarity, color accuracy and overall visual appeal of images and videos. (Pogla, 2024)

In terms of movement, Companies such as German Bionic offer exoskeletons powered by AI that makes physically demanding jobs easier. Their products are worn like a backpack and provide up to 30 kg of extra force to your back, core, and legs. (Geschwindt, 2023)

For cognition and emotions, AI can aid medical health and personalize learning. Woebot is an AI chatbot designed to offer cognitive-behavioral therapy, helping users manage stress and anxiety. (*Woebot Health*, n.d.) Similarly, Duolingo employs AI to tailor language lessons to individual needs, optimizing the learning experience. (Salvador, 2023)

Potential for including: AI enables easier creation of programs that help people with disabilities communicate. Projects focusing on text to speech or speech to sign language can use AI to further develop and help their users.
Potential for exclusion: Bias in the dataset is one of the biggest challenges we are facing with AI at the moment. I will exemplify with a bizarre example: Imagine an AI software that can “see” who has arrived

in class. This particular model is only trained on white people and therefore struggles to identify black students. They might then not be registered and might face challenges in the school because of this.

In my understanding the concepts of understand and understanding refer to the ability of users to comprehend and make sense of information presented to them. In human context; understand means grasping the meaning of information, instructions, or interactions. While understanding involves a deeper cognitive process where a person can integrate new information with existing knowledge, enabling meaningful interaction and decision-making.

Conversely, machines can simulate understanding by processing and responding to information, but their capabilities are not equivalent to human comprehension or cognitive understanding as of now.

1.4 Guideline for Human-AI Interaction

04 Show contextually relevant information. This guideline emphasizes the importance of presenting information that is pertinent to the user's current task and environment. An example is a driving app. The app uses AI to display real-time traffic conditions and suggest an alternative route. The app also provides relevant information, such as nearby gas stations or restaurants, based on the driver´s preferences and the current traffic situation.

I found Nielsen´s Ten Heuristic Principles. (Nielsen, 1994) Both guidelines focus on user understanding where they emphasize making information and interactions clear and understandable. For example, Nielsen's heuristic on “Visibility of System Status” aligns with the Human-AI guideline on providing clear and understandable information.

1.5 Journal - everyday engagement, interaction and use

Pen and paper

The physical act of writing with a pen felt slow but steady. My handwriting is neat at the start but becomes messier as I go along. I had to focus on forming the letters and words, which made me think more carefully about what i wanted to say before writing. It felt good to feel the pen move across the paper.

QWERTY keyboard and display

Using a keyboard to type was quick and effective. Compared to writing things down by hand, I was able to express my ideas faster. It is very simple to update, allowing me to swiftly correct errors and reorder without leaving any trace. I found that when I type, I don't always think things through as well as I did when I wrote by hand.

Speech-to-text and microphone

It felt normal to speak my ideas out loud, but it was uncomfortable when certain words or phrases were misinterpreted by the speech recognition software. I discovered that I had to talk slowly and clearly, which broke up my flow. How quickly I could "write" without using my hands at all was interesting.

Module 2

2.1 Characteristics of AI-infused systems

AI-infused systems are transforming the way users interact with technology, incorporating machine learning and AI capabilities into everyday applications. Unlike traditional systems, which follow explicit rules and predictable behaviours, AI-infused systems are dynamic, adaptive, and often unpredictable based on its nature. These systems leverage vast amounts of data and advanced algorithms to offer features directly exposed to the end user, as defined by Amershi et al, (2019), but also come with unique challenges. In this article, we will explore the key characteristics of AI-infused systems.

One of the fundamental characteristics of AI-infused systems is their inherent uncertainty. AI-systems rely on probabilistic models to make decisions. This means that their predictions are prone to errors, such as false positives–where irrelevant information is identified as relevant– or false negatives, where important information is missed. These errors are a natural consequence of the way AI models generalize from data rather than following predetermined rules. The probabilistic nature of AI ensures that even well trained systems rarely operate at a perfect accuracy, which users must account for when interpreting results (Amershi et al, 2019).

In addition to the uncertainty associated with AI-infused systems, inconsistency presents another important characteristic of AI-infused systems. Traditional user interface design principles place great emphasis on consistency, with predictable behavior and a stable interface appearance (Yang et al, 2020).

In contrast, many AI components are inherently inconsistent. For example, AI-infused systems may respond differently depending on lighting or noise conditions that are not perceived as distinct differences by end-users. Systems may also react differently to the same text input over time, such as in cases where autocomplete systems suggest words after updates to the language model. This inconsistency can be attributed to the probabilistic nature of AI-systems, their complexity and reliance on contextual factors, as well as their ability to learn and adapt over time (Amershi et al, 2019). Inconsistent and unpredictable behavior can lead to user confusion, erode their trust in the system, and even cause them to stop the use of AI technology (Kocielnik et al, 2019).

Another important characteristic is their ability to personalize experiences for users, often behind the scenes (Amershi et al, 2019). This personalization may involve automatic filtering or prioritization of content based on the system's interpretation of user preferences or past behavior.

Spotify as an AI-infused system

Spotify, one of the most popular music streaming platforms globally, is a prime example of an AI-infused system. It uses AI to deliver personalized features and services directly to the end user. While Spotify's AI-driven recommendations and curated playlists have revolutionized music discovery, the system also embodies several of the characteristics previously discussed which bring with it their own set of challenges. The following examples are drawn from my personal experience using Spotify, though I do not know how their algorithm works.

Firstly, recommendation algorithms frequently make predictions with less than perfect accuracy. AI systems, being probabilistic, attempt to predict user preferences based on patterns in listening behavior but these predictions are subject of errors. Both false positives (recommending music the user dislikes) and false negatives (failing to recommend music the user would love). For example, while Spotify’s discover weekly playlist may occasionally surprise users with a perfect song, it is equally likely to include tracks that don’t fit their tastes. This unceranity stems from the complexity of human preferences, which are difficult to quantify entirely because human taste changes over time or mood.

Secondly, a small deviation in input, such as exploring a different genre for a day or listening to music in a different mood, can shift Spotify’s recommendations significantly. This sensitivity means the system is consistently adjusting to perceived changes in taste, even if those changes are temporary or context-specific. For example, a user might enjoy classical music while working but switch to pop during

leisure time. However, if they binge on classical music for a few days, Spotify might suddenly prioritize this genre across all playlists and suggestions, even when the user has no desire to listen to classical outside of a work context.

Finally, by tracking user preferences, the platform automatically filters and curates content that aligns with each user's tastes. While this feature can streamline the music discovery process, it comes with a downside: undesired information hiding. By constantly refining its recommendations to match known preferences, Spotify can create a filter bubble, where users are less exposed to music outside their usual listening habits. This personalization can limit the diversity of content that users see, potentially hiding tracks or artists that don't align with the system's understanding of their taste.

2.2 Human-AI interaction design

Both Amershi et al. (2019) and Kocielnik et al. (2019) address interaction design for AI-infused systems. They identify various challenges associated with designing such systems and propose strategies to address them.

Amershi et al. (2019) present a set of 18 design guidelines for AI systems based on a review of existing research and industry best practices. These guidelines cover several themes, including the importance of making the system's functionality clear to users, setting realistic performance expectations, providing users with control over the system's behavior, explaining the AI system's decisions, and fostering user trust in the system.

Kocielnik et al. (2019) focus on how to manage user expectations for AI systems, arguing that it is crucial to prepare users for the imperfections of these systems, such as their tendency to make errors. The authors conducted two studies to test various design techniques for managing expectations. They found that strategies like communicating the system's accuracy and granting users a degree of control can enhance user satisfaction and acceptance. Their research also indicates that users may be more tolerant of false positives than false negatives in certain contexts.

Evaluating if Spotify adhere to G1 and G11

G1: Make clear what the system can do, and G11 Make clear why the system did what it did, are two design principles that can be used to evaluate Spotify's use of AI in their platform.

The principle of clearity regarding the capabilities of the system emphasizes the importance of users understanding what the AI system can and cannot do (Amershi et al. 2019). Spotify partially adheres to this principle with its intuitive layout, where features such as "Discover Weekly" and "Release Radar" are clearly labeled and explained. However, there is room for improvement. The factors influencing recommendations remain unclear, as users do not know the extent to which the system considers genre, mood, time of day, location, or other user habits. The second principle G11; make clear why the system did what it did, encompasses to give the user insight to why the AI-system has taken a decision. Here Spotify also has potential for improvement. While the app provides a brief description of functions such as "discover weekly", it lacks a more in-depth explanation of how the recommendations are generated. Users are left without any understanding of why specific songs or artists appear in their recommendations.

Problematic aspects of textual content and solutions based on LLMs

Bender et al (2021) critically analyze large language models (LLM), highlighting several problematic aspects associated with textual content generated by these AI systems, along with potential solutions.

One major issue is the distorted worldviews reflected in the training data. LLMs are trained on extensive text data from the internet, which often contains biases mirroring a hegemenic perspective. As a result, these models can absorb and amplify these biases, leading to the discrimination and marginalization of various groups. Accountability is another significant issue. While LLM-generated text may appear coherent, it lacks genuine understanding and communicative intent. This "stochastic parrot" effect raises questions about responsibility for the content produced by these models. The potential for misuse is also alarming, as LLMs can generate fake news, propaganda, and hate speech, making it difficult for users to discern real from false information.

To midigate these challenges, Bender et al. propose investing in smaller, carefully curated datasets to reduce bias, considering the environmental and economic costs of developing LLMs, and focusing on creating AI systems that genuinely understand meaning rather than merely imitating human language.

2.3 Conversational user interfaces to larger models

Conversational interfaces for large language models (LLMs) like ChatGPT introduce unique design challenges. Their probabilistic nature can cause unpredictable behavior, leading to inconsistent responses, which disrupt conversational flow (Yang et al. 2020). Users typically expect continuity and coherence, but LLMs may provide responses that, while contextually relevant, lack a logical progression in the overall conversation. For instance, an LLM might suddenly shift topics or misunderstand references to past dialogue, which can leave the user confused and detracted from the user experience (Amershi et al. 2019).

Another significant challenge is the need to mitigate biases that emerge in LLM-generated content Trained on vast internet data, LLMs may inadvertently reflect discriminatory language or spread misinformation, as they cannot distinguish between facts and fiction. This can result in responses that reinforce stereotypes or provide factually incorrect answers, posing risks for both user trust and ethical use.

To address these issues, guidelines from Amershi et al. (2019) are helpful. G1 - Make system functionality clear: Users need to understand the system's limitations. LLMs should be clearly presented as machines, not humans. G2 - Set quality expectations: Users should be informed about the system's potential for errors, such as misunderstanding context or providing inaccurate responses. Adhering to these principles can reduce confusion and build trust. Clarity about LLM limitations helps users set realistic expectations, while transparency about errors enhances credibility.

References

Web Sources:

About Universal Design. (n.d.). Centre for Excellence in Universal Design. Retrieved September 10, 2024, from https://universaldesign.ie/about-universal-design

Artificial intelligence. (n.d.). Wikipedia. Retrieved September 3, 2024, from https://en.wikipedia.org/wiki/Artificial_intelligence

Artificial Intelligence Definitions. (n.d.). Stanford HAI. Retrieved September 3, 2024, from https://hai.stanford.edu/sites/default/files/2020-09/AI-Definitions-HAI.pdf

Geschwindt, S. (2023, May 3). *This AI-powered exoskeleton does the heavy lifting so you don't have to*. TheNextWeb. Retrieved September 10, 2024, from https://thenextweb.com/news/ai-exoskeleton-wearable-german-bionic-startup

Grudin, J. (2009, Winter). AI and HCI: two fields divided by a common focus. *AI Magazine*, *30*(4), 48+. https://link-gale-com.ezproxy.uio.no/apps/doc/A216267417/AONE?u=oslo&sid=bookmark-AONE&xid=d85837d1

Guizzo, E. (2023, August 9). *What is a Robot? - ROBOTS: Your Guide to the World of Robotics*. Robots Guide. Retrieved September 10, 2024, from https://robotsguide.com/learn/what-is-a-robot

Jordan, J. M. (2019, July 29). *The Czech Play That Gave Us the Word 'Robot'*. The MIT Press Reader. Retrieved September 3, 2024, from https://thereader.mitpress.mit.edu/origin-word-robot-rur/

Pogla, M. (2024, February 23). *AI Empowered: Vision Pro's Secret Sauce for Visual Excellence*. Auto-GPT. Retrieved September 10, 2024, from https://autogpt.net/ai-empowered-vision-pros-secret-sauce-for-visual-excellence/

Roach, J. (2019, November 7). *Making the HoloLens 2: Advanced AI built Microsoft's vision for ubiquitous computing*. Microsoft News. Retrieved September 10, 2024, from https://news.microsoft.com/source/features/innovation/hololens-2-shipping-to-customers/

Salvador, R. (2023, March 14). *Duolingo Max Uses OpenAI's GPT-4 For New Learning Features*. Duolingo Blog. Retrieved September 10, 2024, from https://blog.duolingo.com/duolingo-max/

Spot to the Rescue. (n.d.). Boston Dynamics. Retrieved September 10, 2024, from https://bostondynamics.com/blog/spot-to-the-rescue/

Stamper, J. (2024, August 26). *Robot Definition & Meaning*. Merriam-Webster. Retrieved September 3, 2024, from https://www.merriam-webster.com/dictionary/robot

What is artificial intelligence? (n.d.). What is artificial intelligence? | UiB AI. Retrieved September 3, 2024, from https://www.uib.no/en/ai/152317/what-artificial-intelligence

What Is Artificial Intelligence (AI)? (2024, August 16). IBM. Retrieved September 3, 2024, from https://www.ibm.com/topics/artificial-intelligence

Woebot Health. (n.d.). Woebot Health: Scalable Enterprise Solution for Mental Health. Retrieved September 10, 2024, from https://woebothealth.com/

Conference Sources:

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)

Appendix 1

The feedback i got on iteration 1 was the following:
I miss a bit more about the history of AI, for example, in what context the terminology was presented. There is a bit more about this in Grudin (2009) in the article "AI and HCI: two fields divided by a common focus," which is part of the syllabus (there is also a bit about Turing here if you find it relevant for your assignment), and in the conference proposal by McCarthy that I found here: http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf.

I used this feedback to rewrite the mentioned part and included more information about the history of AI.

Appendix 2

I got the following feedback on the second iteration:
"I'm not sure if the assignment actually asks for this; however, in task 2.3 on 'key challenges in the design of conversational interfaces,' it might be interesting to mention the challenges in the design process of conversational interfaces. In this way, one acknowledges the central role AI plays in the design process itself and how challenges here are the root of further challenges users experience in the interface. Yang et al. 2020 (pp. 2 and 3) mention this (as you already reference here)."

With this feedback i went back to 2.3 and expanded my exploration of challenges in the design of conversational interfaces.