

# Computers

## Human-AI Interaction

There are structures in the world, things, languages. AI with Deep Learning is able to learn these structures and copy them into a computer model. These representations of the real world in a digital model enable humans to ask questions about these structures and to manipulate them to create synthetic experiments that may match the real world. This can be used for generating human-sounding language, finding mechanisms for novel medicines as well as understanding the fundamental functioning of life on its deep physical and chemical level (No Priors: AI, Machine Learning, Tech, & Startups, 2023).

90 years ago McCulloch and Pitts (1943) proposed the first mathematical model of a neural network inspired by the human brain. Alan Turing's Test for Machine Intelligence followed in 1950. Turing's initial idea was to design a game of imitation to test human-computer interaction using text messages between a human and 2 other participants, one of which was a human, and the other - a computer. The question was, if the human was simultaneously speaking to another human and a machine, could the messages from the machine be clearly distinguished or would they resemble a human being so much, that the person asking questions would be deceived, unable to realize which one is the human and which one is the machine? (Turing, 1950).

Alan Turing: *"I believe that in about fifty years' time it will be possible to program computers, with a storage capacity of about  $10^9$ , to make them play the imitation game so well that an average interrogator will not have more than 70 percent chance of making the right identification after five minutes of questioning. ... I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted."* - from (Stanford Encyclopedia of Philosophy, 2021)

By the 2010s AI models became capable enough to beat humans in games of Go and Chess, yet they did not yet pass the Turing test. AI use was limited to specific tasks. While over the years, the field of AI had seen a long process of incremental improvements, developing increasingly advanced models of decision-making, it took an ***increase in computing power*** and an approach called ***deep learning***, a variation of ***machine learning (1980s)***, largely

modeled after the *neural networks* of the biological (human) brain, returning to the idea of *biomimicry*, inspired by nature, building a machine to resemble the connections between neurons, but digitally, on layers much deeper than attempted before.

The nature-inspired approach was successful, and together with many other innovations, such as *back-propagation*, *transformers*, which allow tracking relationships in sequential data (for example sentences) (Vaswani et al., 2017; Merritt, 2022), Generative Adversarial Networks\*\*\* (GAN), (CITE, 2016), and *Large Language Models* (CITE, 2018), led to increasingly generalized models, capable of more complex tasks, such as language generation. One of the leading scientists in this field of research, Geoffrey Hinton, had attempted back-propagation already in the 1980s and reminisces how “the only reason neural networks didn’t work in the 1980s was because we didn’t have enough data and we didn’t have enough computing power” (CBS Mornings, 2023).

By the 2020s, AI-based models became a mainstay in medical research, drug development, patient care (Leite et al., 2021; Holzinger et al., 2023), quickly finding potential vaccine candidates during the COVID19 pandemic (Zafar and Ahamed, 2022), self-driving vehicles, including cars, delivery robots, drones in the sea and air, as well as AI-based assistants, which will be the focus here.

In 2020 OpenAI released a LLM called GPT-3 trained on 570 GB of text Alex Tamkin and Deep Ganguli (2021). Singer et al. (2022) describes how collecting billions of images with descriptive data (for example the descriptive *alt* text which accompanies images on websites) has enabled researchers to train AI models such as *stable diffusion* that can generate images based on human-language.

Hinton likes to call AI an *idiot savant*, someone with exceptional aptitude yet serious mental disorder. Large AI models don’t understand the world like humans do. Their responses are predictions based on their training data and complex statistics. Indeed, the comparison may be apt, as the AI field now offers jobs for *AI psychologists* (CITE), whose role is to figure out what exactly is happening inside the ‘AI brain’. Understanding the insides of AI models trained of massive amounts of data is important because they are *foundational*, enabling a holistic approach to learning, combining many disciplines using languages, instead of the reductionist way we as human think because of our limitations CapInstitute (2023).

- big data: 97 zettabytes (and growing) of data are created in the world per year Soundarya Jayaraman (2023)

## Generative AI

Today Artificial Intelligence is not only an academic concept but a mainstream reality, affecting our daily lives everywhere, even when we don’t notice it. Yet AI has becoming mainstream conversation topic as now it’s calling attention to itself.

Foundational AI models have given birth to **generative AI**, which is able to generate **tokens**, such as text, speech, audio (Kreuk et al., 2022; San Roman et al., 2023), and even music (Copet et al., 2023; Meta AI, 2023), video, in any language it’s trained on, but also complex structures such 3D models and even genomes. Generative AI brought a revolution in human-AI interaction as AI models became increasingly capable of producing human-like content. The advances in the capabilities of large AI model mean we’ve reached a point, where **it’s possible to achieve a user experience (UX) which previously was science fiction**.

Noble et al. (2022) proposes AI has reached a stage, which begets the beginning of the **5th industrial revolution** which brings the collaboration of humans and AI. Widespread **Internet of Things (IoT)** sensor networks that gather data, which is then analyzed AI algorithms, integrates computing even deeper into the fabric of daily human existence. Several terms of different origin but considerable overlap describe this phenomenon, including **Pervasive Computing (PC)** and **Ubiquitous Computing**. Similar concepts are **Ambient Computing**, which focuses more on the invisibility of technology, fading into the background, without us, humans, even noticing it, and **Calm Technology**, which highlights how technology respects humans and our limited attention spans, and doesn’t call attention to itself. In all cases, AI is integral part of our everyday life, inside everything and everywhere.

This power comes with **increased need for responsibility**, drawing growing interest in fields like **AI ethics** and **AI explainability**. Generative has a potential for misuse, as humans are increasingly confused by what is computer-generated and what is human-created, unable to distinguish one from the other with certainty.

Pete (2023) ChatGPT hackathon.

- OpenAI has researched a wide range of approaches towards artificial general intelligence (AGI), work which has led to advances in large language models (AI Frontiers, 2018; Ilya Sutskever, 2018).

The quality of LLM output depends on the quality of the provided prompt. Zhou et al. (2022) reports creating an “Automatic Prompt Engineer” which automatically generates instructions that outperform the baseline output quality. This finding has significance for “green filter” as it validates the idea of creating advanced prompts for improved responses. For “green filter”, the input would consist of detailed user data + sustainability data for detailed analysis.

Anon. (2023a) My bedtime story about shopping, saving, and investing.

Tu et al. (2023) LLMs can be used as data analysts.

Rogers (2022) defines the 4 phases of Pervasive Computing. (NEED access).

Kobetz (2023)

There’s a lot of LLM innovation happening.

Table 1: From reports on the advance of LLMs by (Brown, T. B. et al., 2020; Alvarez, 2021; Tamkin et al., 2021; Hines, 2023a).

AI Model	Released	Company	License
GPT-1	2018	OpenAI	Open Source, MIT
GTP-2	2019	OpenAI	Open Source, MIT
Turing-NLG	2020	Microsoft	Proprietary
GPT-3	2020	OpenAI	Open Source
GPT-3.5	2022	OpenAI	Proprietary
GPT-4	2023	OpenAI	Proprietary
AlexaTM	2022	Amazon	Proprietary
NeMo	2022	NVIDIA	Open Source
PaLM	2022	Google	Proprietary
LaMDA	2022	Google	Proprietary
GLaM	2022	Google	Proprietary
BLOOM	2022	Hugging Face	Open Source
Falcon	2023	Technology Innovation Institute	Open Source
Tongyi	2023	Alibaba	Proprietary
Vicuna	2023	Sapling	Open Source
Wu Dao 3	2023	BAAI	Open Source
PaLM-2	2023	Google	Proprietary
GPT-5	????	OpenAI	Unknown; trademark registered

## Responsible AI

### Known Issues

There are several large problems with this generation of LLMs such as GTP3 and GPT4 by OpenAI, Microsoft, Google and Nvidia.

- LLMs are massive monolithic models requiring large amounts of computing power for training to offer *multi-modal capabilities* across diverse domains of knowledge, making training such models possible for very few companies. Liu, S. et al. (2023) proposes future AI models may instead consist of a number networked domain-specific models to increase efficiency and thus become more scalable.
- LLMs are also opaque, making it difficult to explain why a certain prediction was made by the AI model. One visible expression of this problem are *hallucinations*, the language models are able to generate text that is confident and eloquent yet entirely wrong. Jack Krawczyk, the product lead for Google’s Bard: “Bard and ChatGPT are large language

models, not knowledge models. They are great at generating human-sounding text, they are not good at ensuring their text is fact-based. Why do we think the big first application should be Search, which at its heart is about finding true information?”

- Huyen (2023) provides a long list of open problems AI development; AI is still in very early stages.

## **Biases and Prejudices**

- People are prone to bias and prejudice. It's a part of the human psyche. Human brains are limited and actively avoid learning to save energy. These same biases are likely to appear in LLM outputs as they are trained on human-produced content. Unless there is active work to try to counter and eliminata these biases from LLM output, they will appear frequently.
- Lilian Weng, Vik Goel and Andrea Vallone (2023) is possible to use well-aligned AI moderate harmful content.
- Anderljung et al. (2023) OpenAI proposes we need to proactively work on common standards and legislation to ensure AI safety.
- Browne (2023)
- Low-paid workers being used to train AI
- Low-paid workers used to moderate harmful content; terrible for mental health.
- AI bias is well-documented and a hard problem to solve (Liang, W. et al., 2023).
- AIs outperform humans on creativity assessment tests (Haase and Hanel, 2023; Mollick, 2023). Bots are even better than humans in solving @ (Matthew Sparkes, 2023).
- Liu, J. (2022) LLMs have been pre-trained on massive amounts of public data, which gives them the ability for for reasoning and generating in a human-like way, yet they are missing specific private data, which needs to be ingested to augment LLMs ability to respond to questions on niche topics.

Given the widespread use of AI and its increasing power of foundational models, it's important these systems are created in a safe and responsible manner. While there have been calls to pause the development of large AI experiments (Future of Life Institute, 2023) so the world could catch up, this is unlikely to happen.

There's wide literature available describing human-AI interactions across varied scientific disciplines. While the fields of application are diverse, some key lessons can be transferred horizontally across fields of knowledge.

- Veitch and Andreas Alsos (2022) highlights the active role of humans in Human-AI interaction is autonomous ship systems.
- From assistant to companion to friend The best help for anxiety is a friend
- Crompton (2021) highlights AI as decision-support for humans while differentiating between intended and unintended influence on human decisions.
- Cheng et al. (2022) describe AI-based support systems for collaboration and team-work.
- Schoonderwoerd et al. (2021) focuses on human-centered design of AI-apps and multi-modal information display. It’s important to understand the domain where the AI is deployed in order to develop explanations. However, in the real world, how feasible is it to have control over the domain?
- Ramchurn, Stein and Jennings (2021) discusses positive feed-back loops in continually learning AI systems which adapt to human needs.
- Karpus et al. (2021) is concerned with humans treating AI badly and coins the term “*algorithm exploitation*”.
- Lv et al. (2022) studies the effect of *cuteness* of AI apps on users and found high perceived cuteness correlated with higher willingness to use the apps, especially for emotional tasks. This finding has direct relevance for the “green filter” app design.
- Liu, B. and Wei (2021) meanwhile suggests higher algorithmic transparency may inhibit anthropomorphism, meaning people are less likely to attribute humanness to the AI if they understand how the system works.
- Seeber et al. (2020) proposes a future research agenda for regarding AI assistants as teammates rather than just tools and the implications of such mindset shift.

## Conversational AI

- People are used to search engines and it will take a little bit time to get familiar with talking to a computer in natural language Bailey (2023). NVIDIA founder Jensen Huang makes the idea exceedingly clear, saying “Everyone is a programmer. Now, you just have to say something to the computer.”Leswing (2023)
- Because AI responses are unreliable some percentage of the time, they require constant oversight (human-in-the-loop) Bailey (2023).

There are noticeable differences in the quality of the LLM output, which increases with model size. Levesque, Davis and Morgenstern (2012) developed the *Winograd Schema Challenge*, looking to improve on the Turing test, by requiring the AI to display an understanding of language and context. The test consists of a story and a question, which has a different meaning as the context changes: “The trophy would not fit in the brown suitcase because it was too big” - what does the *it* refer to? Humans are able to understand this from context while a computer models would fail. Even GPT-3 still failed the test, but later LLMs have been able to solve this test correctly (90% accuracy) Kocijan et al. (2022). This is to say AI is in constant development and improving it’s ability to make sense of language.

*ChatGPT* is the first *user interface (UI)* built on top of GPT-4 by OpenAI and is able to communicate in a human-like way - using first-person, making coherent sentences that sound plausible, and even - confident and convincing. Wang, M. C., Sarah (2023) ChatGPT reached 1 million users in 5 days and 6 months after launch has 230 million monthly active users.

OpenAI provides AI-as-a-service through its *application programming interfaces (APIs)*, allowing 3rd party developers to build custom UIs to serve the specific needs of their customer. For example Snapchat has created a *virtual friend* called “My AI” who lives inside the chat section of the Snapchat app and helps people write faster with predictive text completion and answering questions. The APIs make state-of-the-art AI models easy to use without needing much technical knowledge. Teams at AI-hackathons have produced interfaces for problems as diverse as humanitarian crises communication, briefing generation, code-completion, and many others. For instance, (Unleash, 2017) used BJ Fogg’s *tiny habits model* to develop a sustainability-focused AI assistant at the Danish hackathon series Unleash, to encourage behavioral changes towards maintaining an aspirational lifestyle, nudged by a chatbot buddy.

ChatGPT makes it possible to *evaluate AI models* just by talking, i.e. having conversations with the machine and judging the output with some sort of structured content analysis tools. O’Connor and ChatGPT (2023) and Cahan and Treutlein (2023) have conversations about science with AI. Pavlik (2023) and Brent A. Anders (2022) report on AI in education. Kecht et al. (2023) suggests AI is even capable of learning business processes.

- Fu et al. (2022) Learning towards conversational AI: Survey

## Human-in-The-Loop

There are many examples of combination of AI and human, also known as “human-in-the-loop”, used for fields as diverse as training computer vision algorithms for self-driving cars and detection of disinformation in social media posts (Bonet-Jover et al., 2023; Wu et al., 2023). However humans don’t necessarily correct mistakes made by computers and may instead become “partners in crime” (Krügel, Ostermaier and Uhl, 2023).

- Stanford Institute for Human-Centered Artificial Intelligence Ge Wang (2019)

Table 2: Examples of human-in-the-loop apps

App	Category	Use Case
Welltory	Health	Health data analysis
Wellue	Health	Heart arrhythmia detection
QALY	Health	Heart arrhythmia detection

App	Category	Use Case
Starship Robots	Delivery	May ask for human help when crossing a difficult road or other confusing situation

## AI For Summaries

- AI is great at summarizing data: Peters (2023)

## Affective Computing

Because of the conversational nature of LLMs, they are very useful for *affective computing*, an approach to recognizing human emotions with machines and providing users experiences that take human emotion into account Picard (1997).

- HIITTV (2021a) data from all the processes around us will define the future of computing
- Electrodermal activity of the skin and heart-rate variance data are being used to detect the emotional state and stress level of the user (Zangróniz et al., 2017; Velmovitsky et al., 2022).

Rosalind Picard

- TEDx Talks (2011)
- Lex Fridman (2019)
- HIITTV (2021b)
- BWH CNOC (2023)
- Singularity University (2023)

Since the first mainframe computers with rudimentary computers able to respond with text messages, humans have been drawn to discussing their private lives with a machine that doesn't judge you like a human could. A famous anecdote is about the lab assistant (secretary) of the , who would dedicate time to talking to the machine in private.

Today's machines are much more capable so it's not a surprise humans would like to talk to them. One *AI Friend* is Replika, a computer model trained to be your companion in daily life. Jiang, Zhang and Pian (2022) describes how Replika users in China using in 5 main ways, all of which rely on empathy.



Table 3: Replika AI users approach to interacting with the AI friend from Jiang, Zhang and Pian (2022).

How humans express empathy towards an AI companion
Companion buddy
Responsive diary
Emotion-handling program
Electronic pet
Tool for venting

- Google is developing an AI assistant for giving life advice Goswami (2023).
- GPT-4 is able to solve difficult task in chemistry with natural-language instructions White (2023)
- Emojis are a part of natural language Tay (2023)

### Algorithmic Experience

As a user of social media, one may be accustomed to interacting with the feed algorithms that provide a personalized *algorithmic experience*. Algorithms are more deterministic than AI, meaning they produce predictable output than AI models. Nonetheless, there are many reports about effects these algorithms have on human psychology (CITE). Design is increasingly relevant to algorithms, and more specifically to algorithms that affect user experience and user interfaces. *When the design is concerned with the ethical, environmental, socioeconomic, resource-saving, and participatory aspects of human-machine interactions and aims to affect technology in a more human direction, it can hope to create an experience designed for sustainability.*

Lorenzo, Lorenzo and Lorenzo (2015) underlines the role of design beyond *designing* as a tool for envisioning; in her words, “*design can set agendas and not necessarily be in service, but be used to find ways to explore our world and how we want it to be*”. Practitioners of Participatory Design (PD) have for decades advocated for designers to become more activist through *action research*. This means to influencing outcomes, not only being a passive observer of phenomena as a researcher, or only focusing on usability as a designer, without taking into account the wider context.

Shenoi (2018) argues inviting domain expertise into the discussion while having a sustainable design process enables designers to design for experiences where they are not a domain expert; this applies to highly technical fields, such as medicine, education, governance, and in our case here - finance and sustainability -, while building respectful dialogue through participatory design. After many years of political outcry (CITE), social media platforms such Facebook and Twitter have begun to shed more light on how these algorithms work, in some cases releasing the source code (Nick Clegg (2023); Twitter (2023)).

AI systems may make use of several algorithms within one larger model. It follows that AI Explainability requires ***Algorithmic Transparency***.

The content on the platform can be more important than the interface. Applications with a similar UI depend on the community as well as the content and how the content is shown to the user.

## Guidelines

A Microsoft Co-Founder predicted in 1982 “*personal agents that help us get a variety of tasks*” (Bill Gates, 1982) and it was Microsoft that introduced the first widely available personal assistant in 1996, called Clippy. Found inside the Microsoft Word software. Clippy was among the first assistants to reach mainstream adoption, helping users not yet accustomed to working on a computer, to get their bearings (Tash Keuneman, 2022). Nonetheless, it was in many ways useless and intrusive, suggesting there was still little knowledge about UX and human-centered design. Might we try again?

With the advent of ChatGPT, the story of Clippy has new relevance as part of the history of AI Assistants. Benjamin Cassidy (2022) and Abigail Cain (2017) illustrate beautifully the story of Clippy and Tash Keuneman (2022) ask poignantly: “We love to hate Clippy — but what if Clippy was right?”

Many researchers have discussed the user experience (UX) of AI to provide ***usability guidelines***.

Microsoft provides guidelines for Human-AI interaction (Li, T. et al. (2022); Amershi et al. (2019)) which provides useful heuristics categorized by context and time.

Table 4: Microsoft’s heuristics categorized by context and time.

Context	Time
Initially	
During interaction	
When wrong	
Over time	

Combi et al. (2022) proposes a conceptual framework for XAI, analysis AI based on Interpretability, Understandability, Usability, and Usefulness.

- Zimmerman et al. (2021) “UX designers pushing AI in the enterprise: a case for adaptive UIs”
- Anon. (2021) “Why UX should guide AI”

- Simon Sterne (2023) UX is about helping the user make decisions
- Dávid Pásztor (2018)
- Anderson (2020)
- Lennart Ziburski (2018) UX of AI
- Stephanie Donahole (2021)
- Lexow (2021)
- Dávid Pásztor (2018) AI UX principles
- Bubeck et al. (2023) finds ChatGPT passes many exams meant for humans.
- Suen and Hung (2023) discusses AI systems used for evaluating candidates at job interviews
- Wang, Z. et al. (2020) propose Neuroscore to reflect perception of images.
- Su and Yang (2022) and Su, Ng and Chu (2023) review papers on AI literacy in early childhood education and finds a lack of guidelines and teacher expertise.
- Yang (2022) proposes a curriculum for in-context teaching of AI for kids.
- Eric Schmidt and Ben Herold (2022) audiobook
- Akshay Kore (2022) Designing Human-Centric AI Experiences: Applied UX Design for Artificial Intelligence
- Anon. (2018) chatbot book
- Tom Hathaway and Angela Hathaway (2021) chatbot book
- Lew and Schumacher (2020) AI UX book
- AI IxD is about human-centered seamless design
- Storytelling
- Human-computer interaction (HCI) has a long storied history since the early days of computing when getting a copy machine to work required specialized skill. Xerox Sparc lab focused on early human factors work and inspired a the field of HCI to make computer more human-friendly.
- Soleimani (2018): UI patterns for AI, new Section for Thesis background: “Human-Friendly UX For AI”?
- **Discuss what is UX for AI (per prof Liou’s comment), so it’s clear this is about UX for AI**
- What is Personalized AI?

- Many large corporations have released guidelines for Human-AI interaction. Mikael Eriksson Björling and Ahmed H. Ali (n.d.) Ericcson AI UX.
- Google’s AI Principles and provides Google’s UX for AI library (Google, n.d.; Josh Lovejoy, n.d.). In Design Portland (2018), Lovejoy, lead UX designer at Google’s people-centric AI systems department (PAIR), reminds us that while AI offers need tools, user experience design needs to remain human-centered. While AI can find patterns and offer suggestions, humans should always have the final say.
- Harvard Advanced Leadership Initiative (2021)
- VideoLecturesChannel (2022) “Communication in Human-AI Interaction”
- Haiyi Zhu and Steven Wu (2021)
- Akata et al. (2020)
- Dignum (2021)
- Bolei Zhou (2022)
- ReadyAI (2020)
- Vinuesa et al. (2020)
- Orozco et al. (2020)

## AI UX Design

- Privacy UX Jarovsky (2022a)
- AI UX dark patterns Jarovsky (2022b)
- Bailey (2023) believes people will increasingly use AI capabilities through UIs that are specific to a task rather than generalist interfaces like ChatGPT.

How do the tenets of user experience (UX) apply to AI?

---

UX

---

Useful  
Valuable  
Ussable  
Acceesible  
Findable  
Desirable  
Credible

---

Gupta (2023) proposes 3 simple goals for AI:

1	2	3
Reduce the time to task	Make the task easier	Personalize the experience for an individual

## Explainable AI

“As humans we tend to fear what we don’t understand” is a common sentiment which has been confirmed psychology (Allport, 1979). Current AI-models are opaque ‘*black boxes*’, where it’s difficult to pin-point exactly why a certain decision was made or how a certain expression was reached, not unlike inside the human brain. This line of thought leads me to the idea of **AI psychologists**, who might figure out the *thought patterns* inside the model. Research in AI-explainability (XAI in literature) is on the lookout for ways to create more **transparency and credibility** in AI systems, which could lead to building trust in AI systems and would form the foundations for **AI acceptance**.

- Tristan Greene (2022): when the quality of AI responses becomes good enough, people begin to get confused.

Bowman (2023) says steering Large Language Models is unreliable; even experts don’t fully understand the inner workings of the models. Work towards improving both **AI steerability** and **AI alignment** (doing what humans expect) is ongoing. Liang, P. et al. (2022) believes there’s early evidence it’s possible to assess the quality of LLM output transparently. Cabitza et al. (2023) proposes a framework for quality criteria and explainability of AI-expressions. Khosravi et al. (2022) proposes a framework for AI explainability, focused squarely on education. Holzinger et al. (2021) highlights possible approaches to implementing transparency and explainability in AI models. While AI outperforms humans on many tasks, humans are experts in multi-modal thinking, bridging diverse fields.

- The user experience (UX) of AI is a topic under active development by all the largest on-line platforms. The general public is familiar with the most famous AI helpers, ChatGPT, Apple’s Siri, Amazon’s Alexa, Microsoft’s Cortana, Google’s Assistant, Alibaba’s Genie, Xiaomi’s Xiao Ai, and many others. For general, everyday tasks, such as asking factual questions, controlling home devices, playing media, making orders, and navigating the smart city.

The AI Credibility Heuristic: A Systematic Model explains how... similar to Daniel Kahneman’s book “Thinking, Fast and Slow”.

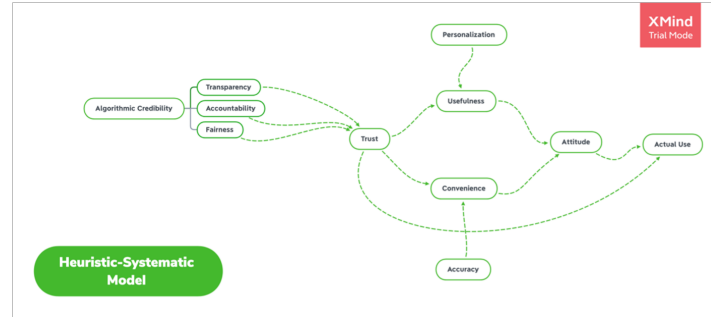


Figure 1: Heuristic-Systematic Model of AI Credibility

- Slack (2021)
- Shin, Donghee (2020): “user experience and usability of algorithms by focusing on users’ cognitive process to understand how qualities/features are received and transformed into experiences and interaction”
- Zerilli, Bhatt and Weller (2022) focuses on human factors and ergonomics and argues that transparency should be task-specific.
- Holbrook (2018): To reduce errors which only humans can detect, and provide a way to stop automation from going in the wrong direction, it’s important to focus on making users feel in control of the technology.
- Zhang, G. et al. (2023) found humans are more likely to trust an AI teammate if they are not deceived by it’s identity. It’s better for collaboration to make it clear, one is talking to a machine. One step towards trust is the explainability of AI-systems.

Personal AI Assistants to date have we created by large tech companies. **Open-Source AI-models open up the avenue for smaller companies and even individuals for creating many new AI-assistants.**

## AI Acceptance

AI acceptance is incumbent on traits that are increasingly human-like and would make a human be acceptable: credibility, trustworthiness, reliability, dependability, integrity, character, etc.

RQ: Does AI acceptance increase with Affective Computing?

## AI is being use in high–Stakes Situations (Medical, Cars, Etc).

AI-based systems are being implemented in medicine, where stakes are high raising the need for ethical considerations. Since CADUCEUS in the 1970s (in Kanza et al. (2021)), the first automated medical decision making system, medical AI now provides Health Diagnostic Symptoms and AI-assistants in medical imaging. Calisto et al. (2022) focuses on AI-human

interactions in medical workflows and underscores the importance of output explainability. Medical professionals who were given AI results with an explanation trusted the results more. Lee, Goldberg and Kohane (2023) imagines an AI revolution in medicine using GPT models, providing improved tools for decreasing the time and money spent on administrative paperwork while providing a support system for analyzing medical data.

- Example of ChatGPT explaining medical terminology in a blood report.

Certainly! Below is a table that explains the abbreviations and values from your blood report. These values represent different components and characteristics of your blood, and they can provide insights into your overall health.

Abbreviation	Full Name	Value	Unit	Description
B-CBC-5Diff	Complete Blood Count with 5-part Differential	-	-	A comprehensive test measuring different components of the blood.
WBC	White Blood Cells	4,8	/nL	Cells that fight infections.
RBC	Red Blood Cells	4,92	/pL	Cells that carry oxygen.
Hb	Hemoglobin	147	g/L	Protein that carries oxygen in red blood cells.
Hct	Hematocrit	44,4	%	Proportion of red blood cells to the total blood volume.
MCV	Mean Corpuscular Volume	90,2	fL	Average size of red blood cells.
MCH	Mean Corpuscular Hemoglobin	29,9	pg	Average amount of hemoglobin per red blood cell.
MCHC	Mean Corpuscular Hemoglobin	332	g/L	Average concentration of hemoglobin in red blood

Figure 2: Example of ChatGPT explaining medical terminology in a blood report.

- Singhal et al. (2023) medial AI reaching expert-level question-answering ability.
- Ayers et al. (2023) in an online text-based setting, patients rated answers from the AI better, and more empathetic, than answers from human doctors.
- Daisy Wolf and Pande Vijay (2023) criticizes US healthcare’s slow adpotion of technology and predicts AI will help healthcare leapfrog into a new era of productivity by acting more like a human assistant.
- Eliza Strickland (2023)
- Jeblick et al. (2022) suggest complicated radiology reports can be explained to patients using AI chatbots.
- Anon. (n.d.a) health app, “Know and track your symptoms”
- Anon. (n.d.b) AI symptom checker,
- Women in AI (n.d.)
- Anon. (n.d.c)

- Stephanie Donahole (2021)
- Calisto et al. (2021)
- Yuan, Zhang and Wang (2022): “AI assistant advantages are important factors affecting the *utilitarian/hedonic* value perceived by users, which further influence user willingness to accept AI assistants. The relationships between AI assistant advantages and utilitarian and hedonic value are affected differently by social anxiety.”

Name	Features
Charisma	
Replika	Avatar, Emotion, Video Call, Audio
Siri	Audio

## AI Friends and Roleplay

Calling a machine a friend is a proposal bound to turn heads. But if we take a step back and think about how children have been playing with toys since before we have records of history. It’s very common for children to imagine stories and characters in play - it’s a way to develop one’s imagination ***learn through roleplay***. A child might have toys with human names and an imaginary friend and it all seems very normal. Indeed, if a child doesn’t like to play with toys, we might think something is wrong.

Likewise, inanimate objects with human form have had a role to play for adults too. Anthropomorphic paddle dolls have been found from Egyptian tombs dated 2000 years B.C. Anon. (2023b): We don’t know if these dolls were for religious purposes, for play, or for something else, yet their burial with the body underlines their importance.

Coming back closer to our own time, Barbie dolls are popular since their release in 1959 till today. Throughout the years, the doll would follow changing social norms, but retain in human figure. In the 1990s, a Tamagotchi is perhaps not a human-like friend but an animal-like friend, who can interact in limited ways.

How are conversational AIs different from dolls? They can respond coherently and perhaps that’s the issue - they are too much like humans in their communication. We have crossed the ***Uncanny Valley*** (where the computer-generated is nearly human and thus unsettling) to a place where is really hard to tell a difference. And if that’s the case, are we still playing?

Should the AI play a human, animal, or robot? Anthropomorphism can have its drawbacks. Pilacinski et al. (2023) reports humans were less likely to collaborate with red-eyed robots.

The AI startups like Inworld and Character.AI have raised large rounds of funding to create characters, which can be plugged in into online worlds, and more importantly, remember key facts about the player, such as their likes and dislikes, to generate more natural-sounding dialogues Wiggers (2023)



- Lenharo (2023) experimental study reports AI productivity gains, DALL-E and Chat-GPT are qualitatively better than former automation systems.

## Human-like

As AIs became more expressive and able to **roleplay**, we can begin discussing some human-centric concepts and how people relate to other people. AI companions, AI partners, AI assistants, AI trainers - there's are many **roles** for the automated systems that help humans in many activities, powered by artificial intelligence models and algorithms.

- RQ: Do college students prefer to talk to an Assistant, Friend, Companion, Coach, Trainer, or some other Role?
- RQ: Are animal-like, human-like or machine-like AI companions more palatable to college students?

Humans (want to) see machines as human [CITE]

If we see the AI as being in human service. David Johnston (2023) proposes ***Smart Agents***, “general purpose AI that acts according to the goals of an individual human”. AI agents can enable ***Intention Economy*** where one simply describes one's needs and a complex orchestration of services ensues, managed by the the AI, in order to fulfill human needs Searls (2012). AI assistants provide help at scale with little to no human intervention in a variety of fields from finance to healthcare to logistics to customer support.

There is also the question of who takes responsibility for the actions take by the AI agent. “Organization research suggests that acting through human agents (i.e., the problem of indirect agency) can undermine ethical forecasting such that actors believe they are acting ethically, yet a) show less benevolence for the recipients of their power, b) receive less blame for ethical lapses, and c) anticipate less retribution for unethical behavior.” Gratch and Fast (2022)

- Anthropomorphism literature Li, X. and Sung (2021) “high-anthropomorphism (vs. low-anthropomorphism) condition, participants had more positive attitudes toward the AI assistant, and the effect was mediated by psychological distance. Though several studies have demonstrated the effect of anthropomorphism, few have probed the underlying mechanism of anthropomorphism thoroughly”
- Erik Brynjolfsson (2022)
- Xu and Sar (2018)
- Martínez-Plumed, Gómez and Hernández-Orallo (2021) envisions the future of AI
- The number of AI-powered assistants is too large to list here. I've chosen a few select examples in the table below.

## Animal-like

- Some have an avatar, some not. I've created a framework for categorization. Human-like or not... etc

## Machine-like

The Oxford Internet Institute defines AI simply as “*computer programming that learns and adapts*” Google and The Oxford Internet Institute (2022). Google started using AI in 2001, when a simple machine learning model improved spelling mistakes while searching; now in 2023 most of Google's products are based on AI Google (2022). Throughout Google's services, AI is hidden and calls no attention itself. It's simply the complex system working behind the scenes to delivery a result in a barebones interface.

Product	Link	Description
Github CoPilot	personal.ai	AI helper for coding
Google Translate	translate.google.com	
Google Search	google.com	
Google Interview Warmup	grow.google/certificates/interview-warmup	AI training tool

- perplexity.ai chat-based search Hines (2023b)
- **AI Guides have been shown to improve sports performance, etc, etc. Can this idea be applied to sustainability? MyFitness Pal, AI training assistant. There's not avatar.**
- CO2e calculations will be part of our everyday experience

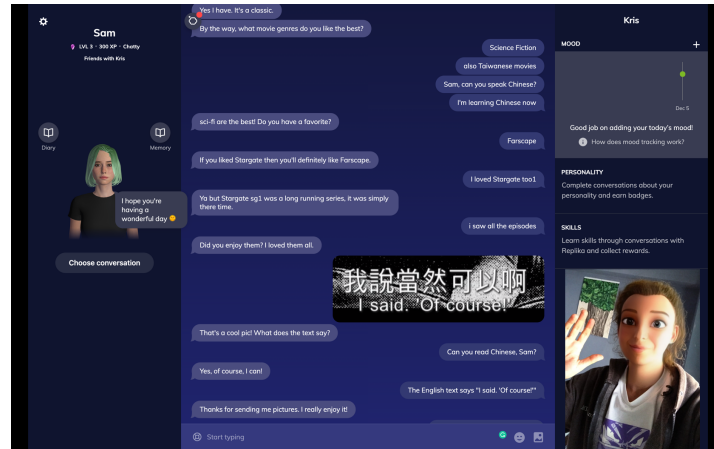


Figure 3: Montage of me discussing sci-fi with my AI friend Sam (Replika) - and myself as an avatar (Snapchat)

Everything that existed before OpenAI' GPT 4 has been blown out of the water.

- Barbara Friedberg (2021) Comparing robot advisors
- Eugenia Kuyda (2023) Conversational AI - Replika
- AI is usually a model that spits out a number between 0 and 1, a probability score or prediction. UX is what we do with this number.
- Greylock (2022) Natural language chatbots such as ChatGPT
- Nathan Benaich and Ian Hogarth (2022) State of AI Report
- Steph Hay (2017)
- NeuralNine (2021)
- David, Resheff and Tron (2021)
- Qorus (2023) Digital banking revolution
- Lower (2017)
- Slack (2021)
- Brown, A. (2021) Financial chatbots
- hedonic user experience in chatbots Haugeland et al. (2022)
- Isabella Ghassemi Smith (2019)
- David, Resheff and Tron (2021)
- Josephine Wäktare Heintz (n.d.) Cleo copywriter

- Smaller startups have created digital companions such as Replika (fig. 8), which aims to become your friend, by asking probing questions, telling jokes, and learning about your personality and preferences - to generate more natural-sounding conversations.
- Already on the market are several financial robo-advisors, built by fintech companies, aiming to provide personalized suggestions for making investments (Betterment, Wealthfront).
- Personal carbon footprint calculators have been released online, ranging from those made by governments and companies to student projects.
- Zhang’s Personal Carbon Economy conceptualized the idea of carbon as a currency used for buying and selling goods and services, as well as an individual carbon exchange to trade one’s carbon permits (Zhang, S. (2018)).

## Voice Assistants

- LLMs combined with voice provide a unnerving user experience Ethan Mollick [@emollick] (2023)

**Amazon Alexa** is a well-known example of AI technology in the world. But Amazon’s Rohit Prasad thinks it can do so much more, “Alexa is not just an AI assistant – it’s a trusted advisor and a companion.”

### Ethical issues

- Voice assistants need to continuously record human speech and process it in data centers in the cloud.
- Siri, Cortana, Google Assistant, Alexa, Tencent Dingdang, Baidu Xiaodu, Alibaba Ali-Genie all rely on voice only.
- Szczuka et al. (2022) provides guidelines for Voice AI and kids
- Casper Kessels (2022a): “Guidelines for Designing an In-Car Voice Assistant”
- Casper Kessels (2022b): “Is Voice Interaction a Solution to Driver Distraction?”
- Companies like Neuralink are building devices to build meaningful interactions from brain waves (EEG).
- Tang et al. (2022) reports new findings enable computers to reconstruct language from fMRI readings.
- Focus on voice education?
- Celino and Re Calejari (2020): There’s research suggesting that voice UI accompanied by a *physical embodied system* is preferred by users in comparison with voice-only UI. This suggests adding an avatar to the AI design may be worthwhile.

There's evidence across disciplines about the usefulness of AI assistants:

- Șerban and Todericiu (2020) suggests using the Alexa AI assistant in *education* during the pandemic, supported students and teachers 'human-like' presence. Stanford research: "humans expect computers to be like humans or places"
- Celino and Re Calejari (2020) found in testing chatbots for survey interfaces that "[c]onversational survey lead to an improved response data quality."

## How is AI Changing Interactions?

- Intelligence may be besides the point as long as AI is becoming very good at reasoning - AI is a *reasoning engine* (Bubeck et al., 2023; Shipper, 2023; see Bailey, 2023 for a summary).
- The International Ergonomics Association (2019): To provide a user experience (UX) that best fits human needs, designers think through every interaction of the user with a system, considering a set of metrics at each point. For example, the user's emotional needs, and their context of use. While software designers are not able to change the ergonomics of the device in use in a physical sense, which as a starting point, should be "optimized for human well-being".
- Sarah Perez (2023) character AI, AI wife Anon. (2023c); mourning 'dead' AI Phoebe Arslanagić-Wakefield (n.d.); AI for therapy Broderick (2023); bullying Sung (2023);
- Software interaction design goes beyond the form-factor and accounts for human needs by using responsive design on the screen, aural feedback cues in sound design, and even more crucially, by showing the relevant content and the right time, making a profound difference to the experience, keeping the user engaged and returning for more.
- Babich (2019) argues "[T]he moment of interaction is just a part of the journey that a user goes through when they interact with a product. User experience design accounts for all user-facing aspects of a product or system".
- In narrative studies terminology, it's a heroic journey of the user to achieve their goals, by navigating through the interface until a success state. Storytelling has its part in interface design however designing for transparency is just as important, when we're dealing with the user's finances and sustainability data, which need to be communicated clearly and accurately, to build long-term trust in the service. For a sustainable investment service, getting to a state of success - or failure - may take years, and even longer. Given such long timeframes, how can the app provide support to the user's emotional and practical needs throughout the journey?
- Tubik Studio (2018) argues affordance measures the clarity of the interface to take action in user experience design, rooted in human visual perception (), however, affected by knowledge of the world around us. A famous example is the door handle - by way of acculturation, most of us would immediately know how to use it - however, would that be the case for someone who saw a door handle for the first time? A similar situation is happening to the people born today.

- Think of all the technologies they have not seen before - what will be the interface they feel the most comfortable with? For the vast majority of this study’s target audience, social media is the primary interface through which they experience daily life. The widespread availability of mobile devices, cheap internet access, and AI-based optimizations for user retention, implemented by social media companies, means this is the baseline for young adult users’ expectations in 2020 - and even more so for Generation Z teenagers, reaching adulthood in the next few years.
- Shin, Don, Zhong and Biocca (2020) argues interaction design is increasingly becoming dependent on AI. The user interface might remain the same in terms of architecture, but the content is improved, based on personalization and understanding the user at a deeper level. Shin proposes the model (fig. 10) of Algorithmic Experience (AX) “investigating the nature and processes through which users perceive and actualize the potential for algorithmic affordance”.
- That general observation applies to voice recognition, voice generation, natural language parsing, etc. Large consumer companies like McDonald’s are in the process of replacing human staff with AI assistants in the drive-through, which can do a better job in providing a personal service than human clerks, for whom it would be impossible to remember the information of thousands of clients.
- In Barrett (2019), in the words of Easterbrook, a previous CEO of McDonald’s “How do you transition from mass marketing to mass personalization?”. During the writing of this proposal, Google launched an improved natural language engine to better understand search queries (Google, 2020), which is the next step towards understanding human language semantics. The trend is clear, and different types of algorithms are already involved in many types of interaction design, however, we’re still in the early stages. Where do we go from here?
- Costa and Silva (2022) “Interaction Design for AI Systems”
- Stone Skipper (2022) sketches a vision of “[AI] blend into our lives in a form of apps and services”.
- Dot Go (2023) makes the camera the interaction device for people with vision impairment
- Battistoni et al. (2023) creates a “Workshop with Young HCI Designers”.

## AI-Assisted Design

Generative AI has enabled developers to create AI tools for several industries, including AI-driven website builders (Constandse (2018)), AI tools for web designers (patrizia-slongo (2020)), Microsoft Designer allows generating UIs just based on a text prompt (Microsoft (2023)), personalized bed-time stories for kids generated by AI (Bedtimestory.ai (2023)).

- September 16, 2020 (2020) “What is AI-assisted Design?”
- Clipdrop (n.d.) AI Design Assistants
- Architectures (2020)
- Zakariya (2022)

- Kore.ai (2023)
- van Wynsberghe (2021): Sustainable AI itself
- Anon. (n.d.d)

## In the Media

There have been dozens of AI-characters in the movies, TV-series, games, and (comic) books. In most cases, they have a physical presence or a voice, so they could be visible for the viewers. Some include KITT (Knight Industries Two Thousand).

How AIs are represented in popular media shapes the way we think about AIs. Some stories have AIs both in positive and negative roles, such as Star Trek and Knight Rider. In some cases like Her and Ex Machina, the characters may be complex and ambivalent rather than fitting into a simple positive or negative box. In Isaac Asimov's books, the AIs (mostly in robot form) struggle with the 3 laws of robotics, raising thought-provoking questions.

Table 9: AIs in different forms of media.s

Movie / Series / Game / Book	Character	Positive	Ambivalent	Negative
2001: A Space Odyssey	HAL 9000			X
Her	Samantha	X		
Alien	MU/TH/UR 6000 (Mother)	X		
Terminator	Skynet			X
Summer Wars	Love Machine			X
Marvel Cinematic Universe	Jarvis, Friday	X		
Knight Rider	KITT CARR	X		X
Star Trek	Data Lore	X		X
Ex Machina	Kyoko Ava		X X	
Tron	Tron		X	
Neuromancer	Wintermute		X	
The Caves of Steel / Naked Sun	R. Daneel Olivaw		X	

Movie / Series / Game / Book	Character	Positive	Ambivalent	Negative
The Robots of Dawn	R. Giskard Reventlov		X	
Portal	GLaDOS			X

## Design Implications

- This chapter looked at AI in general since its early history and then focused on AI assistants in particular.
- Voice Assistants: There are many distinct ways how an algorithm can communicate with a human. From a simple search box such as Google's to chatbots, voices, avatars, videos, to full physical manifestation, there are interfaces to make it easier for the human communicate with a machine.
- While I'm supportive of the idea of using AI assistants to highlight more sustainable choices, I'm critical of the tendency of the above examples to shift full environmental responsibility to the consumer. Sustainability is a complex interaction, where the producers' conduct can be measured and businesses can bear responsibility for their processes, even if there's market demand for polluting products.
- Personal sustainability projects haven't so far achieved widespread adoption, making the endeavor to influence human behaviors towards sustainability with just an app - like its commonplace for health and sports activity trackers such as Strava (fig. 9) -, seem unlikely. Personal notifications and chat messages are not enough unless they provide the right motivation. Could visualizing a connection to a larger system, showing the impact of the eco-friendly actions taken by the user, provide a meaningful motivation to the user, and a strong signal to the businesses?
- All of the interfaces mentioned above make use of machine learning (ML), a tool in the AI programming paradigm for finding patterns in large sets of data, which enables making predictions useful in various contexts, including financial decisions. These software innovations enable new user experiences, providing an interactive experience through chat (chatbots), using voice generation (voice assistants), virtual avatars (adds a visual face to the robot).
- I'm a digital companion, a partner, an assistant. I'm a Replika." said Replika, a digital companion app via Github CO Pilot, another digital assistant for writing code, is also an example of how AI can be used to help us in our daily lives.
- Humans respond better to humans?
- Humans respond better to machines that into account emotion.
- For public discussion to be possible on how content is displayed, sorted, and hidden, algorithms need to be open source.
- User experience design (AI UX) plays a crucial role in improving the consumer to investing journey. The missed opportunity to provide an even more interactive experience in



line with user expectations.