# Opening the Blackbox of Generative AI with Microsoft Copilot

## A Techno-Anthropological study on demystification of relations & processes between humans and chatbots


MSc Techno-Anthropology

Aalborg University Copenhagen

June 2024

Sarah Feldes & Tino Nielsen


# Abstract

The following report investigates the intra- and interactional processes between humans and Microsoft Copilot as an exemplary, state-of-the-art generative Artificial Intelligence Technology. Elaborating and extending existing models of Human-World-Technology relation and mediation, as well as theories of Human-Computer-Interaction and Human-Machine-Communication, we analyze and discuss the relations, intra-actions, and power structures, that shape the way people interact, communicate, and interpret the GenAI and the technology behind it.

Discussing the technical and social blacboxes emerging out of structural and technological elements, and demonstrating how adaptations of existing models of technological mediation and the Techno-Anthropological discipline support demystifying these, the report concludes by offering future lookouts on User Experience research based on extended Techno-Anthropological inquiry as a means to inclusive technological development.

Keywords: intra-action, Techno-Anthropology, blackboxing, User Experience Research, Microsoft, Copilot, inclusion, Human-Computer-Interaction, generative Artificial Intelligence

## Acknowledgements

We want to thank Microsoft for the collaboration on the present inquiry. Their support, supervision, and openness to investigating the topic have paved the way to make our research possible. Thank you to all employees; developers, researchers, managers, that volunteered to give interviews as part of the research process, and a special thanks to Anne Klærke Leth Olesen and Rikke Kiilerich for connecting us with their colleagues at Microsofts' UX department and for their guidance and help through the entire research process.

We also want to thank Teknologirådet and Marie Hoff and her team at TITAN for letting us test their chatbot, sharing their expertise, and giving their perspective and experience on user-driven innovation in generative AI tools and chatbots in expert interviews.

We thank Københavns Professionahøjskole, and especially Troels Jensen whose contribution to state-of-the-art literature and user-involvement has given valuable insights to the lifeworlds and challenges of users within and outside the Danish education system with the use of chatbots and Artificial Intelligence in general.

Last, but most definitely not least we thank our supervisor Lars Botin, and everyone at Aalborg University Copenhagen and Techno-Anthropology, for teaching, challenging and supporting us throughout our academic journey the past 2 years.

# Table of Content

# 1. Introduction

Decades of research and development on algorithms, Artificial Intelligence (AI) and automated generation of content by computers have been subject to rapid increase of attention and discussion by the broader public, since the company OpenAI released their generative AI chatbot ‘ChatGPT’ on November 30th, 2022. The Artificial Intelligence is based on the GPT language model, which OpenAI had worked on and refined since 2018. Even though ChatGPT was not the first generative AI-based technology that had been created, it had a huge impact on the evolution of language processing and Artificial Intelligence in terms of processing power and data availability. Apart from that it sparked a whole new era of content creation, public debate and social controversies, as it was for large parts of the general public the first actively engaging with technologies like this. Adoption and use cases increased exponentially and the tool was quickly adopted into fields like programming, education and creative writing, where Artificial Intelligence showed to be able to create not only relevant, but high-level professional content, partly indistinguishable from text produced by experts in a respective field. With this boom in use, other companies quickly followed suit and the entire business domain of the technology grew at an intense speed. Models thus become better almost by the minute and excel at creating accurate outputs in a range of different media forms like text, images, and audio (Chiu 2023).

Despite its immediate success and applicability to a vast range of use cases and professions, there remains uncertainty and gray zones about the employability, ethics, and social as well as technical limitations for users in relation to chatbots and Artificial Intelligence based technologies. Cases of racist, sexist, discriminatory, political and factual incorrect output have made news and shed light on unsolved problems that humans and technology find themselves confronted by in the age of Artificial Intelligence (Mauro & Schellmann 2023).

The present study thus seeks out to shed light on a specific area of such interaction-based problems, namely the processes taking place when humans and computers engage with each other in written conversation, the chatbot. Looking at one example of Microsoft Copilot, to elucidate an interactional space that still is a ‘black box’ to most people.

# 2. Field of Inquiry

With the ambition to explore matters of ethics, demystification, and inclusion we seek to investigate the processes of the generative AI technologies that humans experience upon interaction with them, and how those processes impact the generated output. Focusing on the development side of the product through qualitative empirical methods, we are especially interested in understanding which agendas and measures can and are being set in place by humans and are influenced through human power towards the technology in order to prevent harmful and offensive output and make people as possible successful end-users of the technologies.

As Techno-Anthropologists, we are interested in understanding the human processes and actions in relation to technology that define their experiences and attitudes towards them. We therefore focus on the Human-Computer-Interaction and processes and considerations that are taking place when users engage with generative AI. While the technology of the generative AI chatbot is the centerpiece of the study in terms of the state-of-the-art technology that Microsoft Copilot offers, the analytical focus is less on the backend of the technology itself, but rather on the human component and its interactions with the backend and procedures of user experience research.

Due to the exploratory nature of the research project, as well as the fast emergence of the theoretical field that covers the area of enquiry, we combine elements and developments from different theoretical domains within Science and Technology Studies (STS). The meta-theoretical frameworks of epistemology and ontology open up the conceptual space to a variety of theoretical trajectories that guide the research. Therefore, theories from Social Sciences and Science and Technology Studies building on approaches by Karen Barad and Olya Kudina in combination with theories from Human-Computer-Interaction (HCI) by Andrea Guzman, Seth Lewis, and feminist HCI by Ann Light build the combined theoretical foundation of the research.

## 2.1 Research Gap

Generative AI and their relations to humans at its current state still is a relatively new field, both for public and scientific research. With its rapid evolution especially since 2022, a range of different scientific inquiries have been done on AI and specifically generative AI. One side of this research is the evaluation and positioning of the technology. These studies include the

investigation of different requirements for successful outputs including hardware, software and user knowledge, the different types of generative AI, as well as the evaluation metrics used to compare different language models (Bandi et al. 2023, Thompson et al. 2022).
Another focus area is the quality of the output and how similar to human produced text the generated content is. This includes investigating the Large Language Models used in specific generative AIs, exploring the datasets they are built on, as well as the challenges of inadequate data. (Abburi et al. 2023, Jeong 2023). On the more social scientific side, research based on the human use of this technology includes the investigation of how students use generative AI and what their immediate feelings were towards the technology. The researchers then use that data to create different types of AI usage and their perception of the interaction (Zhu et al. 2024).

Despite the extensive research on generative AI and chatbots related to the use by different stakeholder groups, content and quality of output, there remains a gap in research of why humans and chatbots interact with each other in the way they do, and what causes the output to look the way it does. Rather than investigating technical configurations of AI and the 'how' of human interaction, we therefore want to investigate AI as a fluid phenomenon where there are differences between the outcomes depending on the person that engages with them and on the context, it is engaged in, looking also at the 'why' we are engaging with chatbots the way we do as of today.

## 2.2 Problem Statement

Based on this gap within existing research on the state-of-the art technology, we investigate the following research question and sub questions:

**How can considerations on user inclusion and blackboxing in development of generative Artificial Intelligence in the case of Microsoft Copilot be addressed and demystified through the concept of intra-action and why does Techno-Anthropological inquiry and problem solution support user inclusion?**

1. *How do intra-actions play out in use and development of Microsoft Copilot?*
2. *How and why is mystification taking place?*
3. *By which means can such processes be demystified?*

# 3. Case description

In line with the rapid development of generative AI technologies and especially chatbot interfaces, Microsoft Copilot Development Lab announced their new feature “Copilot” on the company’s own blog in March 2023. The revolutionary AI tool was officially released at a virtual launch event April 4th, 2023, featuring the presentation of basic its functions, capabilities and exemplifying use cases (Microsoft 2023).

The new feature was integrated into the companies’ existing products and allows users to *“simply describ(e) what you would like your bot to do in natural language and having a complete conversational dialog appear—ready to use in seconds”* (Microsoft 2023).

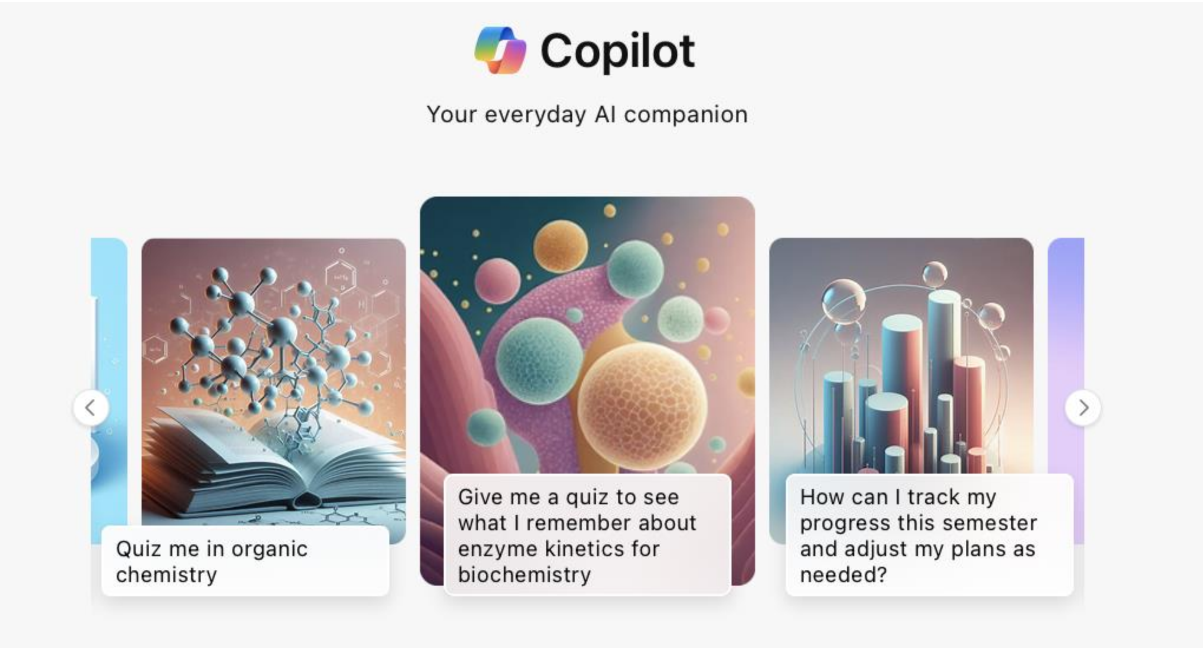


Illustration 1: Screenshot of Microsoft Copilot User Interface (Microsoft 2024)

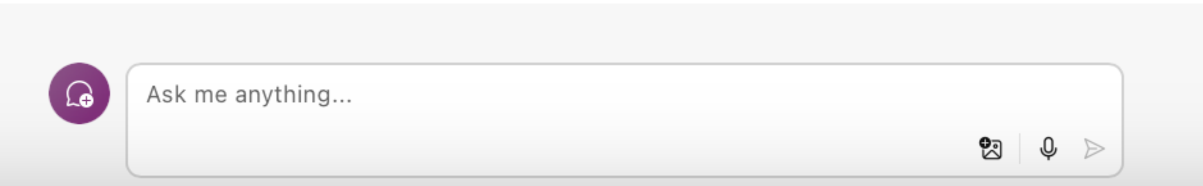


Illustration 2: Screenshot of Microsoft Copilot Chat Interface (Microsoft 2024)

The company describes the feature as a ‘dialogue’ in their press release, initiated through written prompts that are inserted in the textbox of the chatbot interface, which automatically appears in all integrated Microsoft software like the Office package that includes commonly used programs such as Microsoft Teams, Outlook, PowerPoint etc. (Microsoft 2024).

With the aim to facilitate and speed up workflows within small and medium sized organizations in particular, the tool is targeted mainly to professionals within corporations that have to solve extensive lists of administrative and communication tasks on a daily basis. Thus, the tool is created with a user in mind that works with Microsoft tools regularly and is familiar with navigating the different applications and knows how to use the various platforms.

Shortly after the release of the Copilot program, Microsoft further published guidelines for the implementation and use of the features. Among other things a conversational guide called *"Microsoft 365 Copilot: The art and science of prompting"* (Microsoft 2023). What the name indicates quite clearly is Microsofts' attention to the skills and knowledge needed in order to master the usage of the new feature.

While the conversational guidelines are directed to end-users, that is Microsoft's customers, they also continue to release documents and guidelines that are supposed to help employees to assess the artificial intelligence products they are working on, to ensure the most user-friendly, effective and inclusive tools.

One of these guidelines is the "*Microsoft Responsible AI Act*" (Microsoft 2022) that specifically focuses on elements and points of awareness that everyone in the company working on Generative Artificial Intelligence based technologies must adhere to. Since Copilot is the latest advancement and addition to the Generative AI toolkit at Microsoft, these documents were already found before the release of Copilot in 2023, adhering to all generative AI technologies that the company works on.

The current version of the Responsible AI Act was last updated in 2022 and covers six overarching goals which each contain several standards on a total of 27 pages, each of which entails a set of sub-goals with concrete requirements that guide employees through what exactly is meant by each of the concepts (Microsoft 2022).

Apart from the written guidelines, employees that work in any way in connection to the development of Copilot or Artificial Intelligence at Microsoft in general further receive responsible AI training in the form of online workshops, inspiration days that present cases from other teams at the company, and internal video materials. Within each team, there is one or more designated "responsible AI champs", which refers to employees that have extensive knowledge in the responsible AI guidelines, procedures, and regulations, and help to educate their peers further on the topic.

*"While our Standard is an important step in Microsoft's responsible AI journey, it is just one step. As we make progress with implementation, we expect to encounter challenges that require us to pause, reflect, and adjust. Our Standard will remain a living document, evolving to address new research, technologies, laws, and learnings from within and outside the company."* (Microsoft 2022).

The responsibility training within Microsoft therefore aims to make development processes flexible and open to changes when new technologies, laws or knowledge come into the picture. Responsibility in this view means to reflect and adjust when unforeseen challenges arise and to listen to new knowledge not only within the company but also from external actors. As the focus of the present investigation is on the development side of generative Artificial Intelligences, the research and collaboration revolved mainly around the interactions, experiences, and work processes that employees involved in development have with and towards the technology.

For the present study we collaborated with Microsoft specifically regarding their Copilot for Business Central product, and in particular their User Experience research team, a smaller division located in Kongens Lyngby, Copenhagen, Denmark, that works as a local branch of the larger UX team in Redmond, United States.

# 4. State of the Art

In order to derive insights on the different aspects connected to interaction between human and technology, we need to understand what constitutes these agents in discussion, which technical terms are used to talk about them, and how recent literature is discussing the state-of-the-art artifacts.

With the emergence of new technologies such as ChatGPT (OpenAI 2022), Claude (Anthropic 2024), Gemini (Google 2024) and Copilot (Microsoft 2024) as members of the family of chatbots and Artificial Intelligences, not only do the digital interfaces themselves but also a plethora of new concepts fill peoples' lives and much of the content they are presented with in (social) media.

New terms and ways to talk with and about technology found their way into everyday vocabulary, however stemming from scientific and very much technical fields, they require proper introduction for most people to be understood. Aiming to achieve exactly this, an introduction to the new language and landscape of Artificial Intelligence has been a core task to many scholars and professionals since the revolutionary release of ChatGPT in November 2022.

In order to provide guidance for end-users whose field of expertise lies outside Computer Science, IT, or Engineering, public organizations and other experts in the field continue to establish handbooks and dictionaries to explain and exemplify these new terms. An example is the release of the paper "*Sprogmodeller for Dummies - En intuitiv introduktion til teknologien bag ChatGPT*" ("*Langue Models for Dummies - An intuitive introduction to the technology behind ChatGPT*") written by Troels Jensen from Københavns Professionshøjskole (Copenhagen University College) to assist the local Danish audience in navigating technical terms and clear up with potential confusion and misconceptions surrounding the specific technology of ChatGPT (Jensen 2024).

Also, on an international basis the need to present and elaborate the various concepts that are associated with Artificial Intelligences and chatbots, remains to be on top of the academic agenda.

Latest publications such as "*A contemporary review on chatbots, AI-powered virtual conversational agents, ChatGPT: Applications, open challenges and future research directions*" (Casheekar et. al 2024), released in the current edition of the journal *Computer*

*Science Review* focus especially on ChatGPT as the most disruptive technological development in the area to date and challenges that occurred with its release, but also offers definitions of other associated concepts and sheds light on research gaps and ethical dilemmas that evolve(d) out of the emergence of Artificial Intelligences and chatbots (Casheekar et. al 2024).

Many of these terms such as Algorithms, Machine Learning, Large Language Models and Chatbot share technological DNA with "Artificial Intelligence" (AI) and are used under this umbrella term in everyday language. However, they refer to different elements of the technology, interaction with it, and associated phenomena and obstacles.

Breaking down the vast field of associated terms in this way already points to how broad the conceptual spectrum and technological field of Artificial Intelligence is, and for how many different fields within academia and broader society it has relevance.

For the present study we narrowed down the technological and conceptual field, to a space closely tied to the case study of Microsoft Copilot, to remain within the scope of concepts that are central to the problem formulation.

Throughout the following pages we therefore present the state-of-the-art technology through major related concepts of the domain of written prompt-based chatbots powered by artificial intelligence. We elaborate especially on the concepts of Algorithms, Large Language Models, Prompting, Machine Learning, (generative) Artificial Intelligence, Bias, Chatbots and Hallucination as the technological elements that are central to understanding the product Microsoft Copilot and thus map out different areas of interaction within these technological spaces.

Lastly, we present the latest social developments in terms of EU-regulations and concepts within policy making that impact discussions on Artificial Intelligence and development of tools using these technologies.

## 4.1 Algorithms

The basis for all Artificial Intelligence technologies and related concepts that are being explored throughout this paper is grounded on the mathematical elements of ***algorithms***.

As the term is used today, in the age of Artificial Intelligence, it can be defined as "*defined, computational instructions*", thus they are the mathematical sequences embedded in long lines of code that a computer follows as it runs a program (Jensen 2024).

However, the term has its origin way before the computer came to be and can therefore refer to something as simple as calculating a common divisor for two numbers (Birkholm 2018).

Today the concept is often used to refer to bigger pieces of computer software, but in fact often only make up parts of the complete software, not the computer program itself.

While computer programs do contain algorithms as parts of their many lines of code, others also include pure computational actions or commands and other coding "*materials*" (Birkholm 2018).

In their essence simply being mathematical equations, the concept algorithm has become a term with a multiverse of meanings throughout the past years. By now ranging from an almost independent agent to some, an instrument of power and politics and a meta-decision-making-blackbox for a large part of society (Ziewitz 2016).

The understanding and use of the term depends heavily on the specific actor that interacts with it. An end-user thus might interpret quite differently than what is meant when a software developer or a policy maker use the term "algorithm" as part of their work.

It has been criticized how the mere term algorithm has become *"somewhat of a modern myth"* (Ziewitz 2016, p.3) in recent years, as the concept is used increasingly in both daily speech and academia as an entity of its own that has the ability to *"rule, sort, govern, shape, or otherwise control our lives"* (Ziewitz 2016, p.3).

In that sense the algorithm has moved away from its essential connotation as a mathematical equation to an agent, and even further a multi-stable object that can be experienced and lived by - much like a culture of its own (Seaver 2017).

This is also due to larger societal developments that bring the technologies much closer to the human, as people interact with "their algorithm" on social media on a daily basis, associating it with entertainment, content generation, and computational sensemaking of the content they are presented with online (Seaver 2017).

It is important to account for these different connotations that "the algorithm" has received in latest discourse, since it informs how broader society and academic literature perceive technologies differently, and sometimes react differently or alter their behavior, when it is said that they are interacting with technology which runs based on this type of mathematical model.

Even more drastic, state-of-the-art literature talks about a form of anxiety that the mere term "algorithm" can evoke in people, depending on their knowledge and pre-assumptions on the concept (Seaver 2017).

This goes to show that the mere technical term that is part of a technology is perceived vastly differently by user and stakeholder groups and comes with connotations that exceed the terminology of technological components to social and cultural phenomena.

## 4.2 Large Language Models

The technology of Language Models (LM) or Large Language Models (LLM) takes the algorithmic technology a step further from basic mathematical equations with its computational abilities of generating text and spoken word, that mimic the human language, sometimes almost indistinguishable (Chang et al. 2024)

This is achieved through the model's transformative calculations that statistically predict what the likelihood of a word appearing next in a sentence on a given topic will be, based on pre-calculated probability which it retrieves from pre-trained datasets that is fed with as a knowledge base (Chang et al. 2024)

The models can thus generate new human-like text, based on probability of semantic occurrences based on the types of input that a user provides the tool with. The difference between Language Models (LM) and Large Language Models (LLM) resides in their size and capability. LLM are the more advanced versions of the regular LM with bigger parameter sizes and better learning capabilities. The big change from LM's to LLM's came with the introduction of the transformer modules, with makes it possible for the LLM's to do context based learning which learns from specific contexts or prompts given to the model. This and the aspect of human feedback makes the LLM's perfect for interactive and conversational use (Chang et al. 2024).

A way to interact with and test these models is with the use of prompt engineering where specific text is created and used as input for the model with the purpose of getting a specific type of answer output.

### 4.2.1 Prompting

Because large language models are based on human structured language with their transformer modules, the way to interact with them and give them input, is by writing full sentences, that the language model can then context analyze and give an output based on. These inputs are called prompts and can have different functions (Chang et al. 2024). The prompt that is given as input by the user is called a user-prompt, and is the specific instructions the model answers to. The output the model gives, will be a direct response to this prompt. The other type of prompt is the system-prompt, or a meta-prompt, which can be made by the developers. These types of prompts will influence all the following user-prompts and are often more general instructions of what type of answers the model should give (Chang et al. 2024). The meta-prompts often include restrictions on what the model cannot create answers based on.

## 4.3 Artificial Intelligence

The idea of robots or computers with human-like intelligence is a relatively old thought, as it has been the core of fiction since the 1800s, where computers can act, think, and empathize just like humans are able to (Gil de Zúñiga et al. 2024).

However this way of thinking artificial intelligence differs  from the technologies we call AI today. The concept of AI we see today was coined in 1955 when the mechanical procedure of machine learning was defined (Gil de Zúñiga et al. 2024).

Today's discourse on AI is not as clear cut as it is with many other technologies. This is partly because of the lack of a systematic scholarly definition of the term AI, and the debateability of the use of the word ‘intelligence’ in this context (Gil de Zúñiga et al. 2024, Birkholm 2018). Nonetheless, scholars have tried to use the interdisciplinary interest in this type of technology to create a meaningful definition of the terminology. One possible definition comes from de Gil de Zúñiga et al. who describes AI as:

“*the tangible real-world capability of non-human machines or artificial entities to perform, task solve, communicate, interact, and act logically as it occurs with biological humans*.”
(Gil de Zúñiga et al. 2024, p. 326)

In this definition they put focus on different types of AI depending on two dimensions: (1) level of performance and (2) level of autonomy. The different types of performances include performing tasks, taking decisions and making predictions and the level of autonomy is based on the level of human input, interaction and supervision. Other definitions before this includes Ninness and Ninness description:

*“learning algorithm used to approximate some form of intelligence operating within computing machines”*
(Ninness & Ninness 2020, p. 100)

Here, the interaction between humans and AI is not as clear. The main focus is rather on how algorithms can approximate human intelligence. In this sense, intelligence comes from how complex algorithms and computer programs can be in terms of how many operations can be done and how they are presented.

With both of these definitions the technology of AI does not have to actually be intelligent in the same way humans and other living organisms are, the focus is rather on approximating or mirroring human actions and abilities. One technological example of this is image recognition which is a technology under the umbrella of artificial intelligence, which has the ability to recognize the content of images, without the need for a human actor to annotate the picture beforehand.

## 4.4 Generative Artificial Intelligence

The most recent development to be observed within the domain of Artificial Intelligence in line with expanding computational power, to handle and organize data, is generative AI. Moving from retrieving information, piecing it together, and more or less “simply” recounting it for the user, artificial intelligence tools now generate unique content by piecing together smaller individual brackets of data, which it retrieves from its respective data sets it is trained on (Epstein & Hertzmann 2023). This is achieved with the aforementioned Large Language Models, which makes it possible for the technology to handle and express human-like language. The output often takes form in text in the case of communicative generative AI but can also be in the form of images and speech. A well-known example of a communicative generative AI is ChatGPT, a specific program based in Artificial Intelligence which creates a unique answer every time based on the specific input by the user. When a longer conversation thread is made with ChatGPT, all the former inputs and outputs are “remembered” and used for the creation of the newest output (Epstein & Hertzmann 2023). The AI thus learns from previous input. Just like actual human conversations, this way of calculating outputs is based on adding new information to previous human-language based input that is accumulated and taken into consideration for the further conversation, just like meeting a new person in real life and incorporating information gathered from physical conversation in future dialogue (Epstein & Hertzmann 2023).

### 4.4.1 Hallucination

The nature of Generative AI causes the model to always give an answer to the prompt given by the user. Because of how the model uses statistics to generate answers, it will always try to create the most fitting answer, if the input is not caught in any meta-prompt restrictions. While this gives users a certainty that they will receive some kind of output to their input, it does not secure the quality or truthfulness of the content of said output. The consequence of such improper content has been coined as AI Hallucination (Alkaissi & McFarlane 2023).

By hallucination it is referred to the process whereby the Generative AI derives to an answer that is related to the prompt as it consists of a semantically and grammatically sound arrangement of words and might even seem correct content wise at first glance, especially for someone without expertise in the area, but the content is factually wrong or not based on actual data points.

Alkaissi and McFarlane give examples of this in their paper *"Artificial Hallucination in Scientific Writing"* (Alkaissi & McFarlane 2023). In their example they consult ChatGPT about specific medical questions that require a complex answer. The researchers held the generated answer up against already existing literature on the subject and confirmed its legitimacy. They then asked the model to give the relevant literature that it draws its answers from, and here the model starts to hallucinate an answer. It creates titles and IDs of papers that are non-existing or completely unrelated to the topic (Alkaissi & McFarlane 2023). Here, the model answers the prompt with titles of papers that sound relevant because it uses relevant words in the titles, but the actual content of the answer is false and non-existing and therefore not actually relevant for the user.

## 4.5 Machine Learning

Artificial intelligence technologies as we know them today are created and trained with the use of Machine Learning, which refers to the computational process of using input data to achieve a desired outcome without giving the computer those specific commands through coding (Naqa & Murphy 2015). This can be done by giving the computer either a specific end goal or by manually telling it when it is right or wrong. This way, the computer will learn what to do over time and in the end be able to become better by itself (Naqa & Murphy 2015). These types of programs can be described as "soft coded" meaning that they are coded to learn by analyzing the relation between the input and the output, but not to do a specific task. This means that they are very flexible in their ability to complete tasks and that they can adapt to different scenarios and contexts. The adaptation happens through what is called training, where the algorithms are calibrated to be able not only to create the desired outcome from the input used in the training process but also from previously unseen input data (Naqa & Murphy 2015). The optimal goal of machine learning is to emulate the learning processes of human beings. This also means that when the complexity of the machine learning algorithm increases, the rationale of its outputs becomes increasingly harder to follow, which causes it to become what is called ‘blackboxed’.

### 4.5.1 Black Boxing

The term black box from a technical perspective comes from the field of cybernetics and other engineering fields, where it refers to a group of machine elements or commands that are generally well known within the specific field but is too complex to easily map out or explain in detail. The important elements of a black box are the input and the output (Olesen & Kroustrup 2007). This notion has also been adopted into fields like sociology, philosophy of science and Science and Technology Studies, where it refers to the things which inner workings, we as people do not have to actively think about or react to (Olesen & Kroustrup 2007).

However, in recent years the term has also become a concept in broader public conversation to point to different elements and the workings of emerging digital technologies that only a limited percentage of the population have knowledge on. Furthermore, as we will see in the latter part of the analysis, the concept also has a varying meaning depending on who is referring to which technologies. It thus becomes both a technical and socio-political concept that is used to discuss controversies emerging out of these discrepancies and differences in black boxing as a social phenomenon in the age of Artificial Intelligence (Olesen & Kroustrup 2007).

## 4.6 Bias

The type of output and the rationale behind the machine learning are based on the specific kind of data the model is trained on. Since each model is programmed by different people from different countries with different belief systems, knowledge and (economic) goals, there is a predetermined set of information and solutions for the model to choose and generate its answer from. This means the model can be said to have a prejudice towards every type of input, as its repository of answers is circumstantial (Gichoya et al. 2023). In the technological domain this is referred to as the bias of the model, to be seen as a value-free concept to begin with, as it is simply supposed to take into account that the model is not omniscient but limited to its data set. Bias within the field of AI and algorithms can be categorized as human bias or machine bias depending on if the bias comes from how the model is made by the developers or if it comes from, how the model handles the data, and what the training data is (Gichoya et al.

2023). However, within these two categories there is a range of different ways and times that bias can creep into an AI model.

A problem as a consequence of this is the lack of inclusion and availability of all existing data that can lead to output being generated that excludes certain groups, opinions, positions, facts etc, and thus seem to “pick a side” or only portray a certain picture given the specific prompt. This definition of the word bias is distinct to the pop-cultural connotation it has received in daily speech among the broad public since cases of harmful content created by Generative AI as the result of the models' biases have been publicly problematized. The identification and evaluation of bias in AI models are therefore also a way to discover new opportunities for use of the models (Roselli et al. 2019).

## 4.7 Chatbots

Chatbots are a specific type of generative technology that focuses on making the experience feel like a human conversation. Also referred to as a “conversational agent” this sort of software adds an interface to the generative computer program, which makes it possible to have a human-like dialogue with a computer (Adamopoulou & Moussiades 2020, Casheekar et al. 2024).

The technology of chatbots thus combines different elements of the technologies described above. It has the computational elements of generative AI in the sense it creates answers based on the input it gets. Depending on how advanced the chatbot is, the answers can be based on different types of algorithms or different sizes of language models.

For this reason the most text-based generative AI comes in the format of a chatbot. The main element that sets the chatbot apart from other generative AI in general is the message-like visual representation that hides the technological elements of the chatbot behind a user-friendly interface. This means that for the user the actual technology remains invisible and only the conversation, that is the chat window sometimes even equipped with an avatar or other sort of visual symbol, is visible. Chatbots fall under the field of human-computer interaction because of their interactive nature as a user is directly engaging in a written conversation with the interface (Adamopoulou & Moussiades 2020).

The word “chatbot” as in “chatting” and “robot” as well as the term “conversational agent” already imply the humanized appearance of a technology that characterizes this specific sort of tool. The highly individual output that the software can produce thanks to the Artificial Intelligence mimics a human conversation (in some cases) on an impressively natural and relational level (Casheekar et al 2024). This goes especially for “social bots” , the specific type of chatbot that is programed to *“engage users in casual conversations for entertainment or emotional support”* (Casheekar et al 2024, p. 8)

Studies on users’ perception of chatbots already from 2018 showed that most users interpret chatbots as friendly companions rather than a simple tool or program for assistance. Furthermore, this can be seen in an evaluation of an investigation from 2017 that looked at the type of request users typed into a specific chatbot. Almost half of the reviewed requests showed that conversations were based on emotional rather than informational content, adding to the overall sentiment of a chatbot being a conversation partner, rather than a search engine or work tool (Adamopoulou & Moussiades 2020).

However, there are many categories and types of chatbots whose output and level of conversation depends on the intended type of conversation or task they are programmed to accomplish, such as education or information for knowledge-based chatbots, or customer-service and assistance for task-oriented chatbots (Casheekar et al. 2024, p.8).

## 4.8 EU AI Act

The range of technical terms needed to understand the functions of the AI is one indicator of how complex the technology is. And with this technological complexity and opportunities of use comes also a complex matter of risks. And exactly because of the uncertainty of the potentials and the risks the European Parliament has made regulations on how this technology should be used, in the form of the AI Act. The aim of the regulations is to give the developers of AI technology better conditions to create AI that has the potential to benefit society, but with the possible risks in mind. To make sure this happens the European parliament focuses on elements of transparency and traceability and that the AI systems are non-discriminatory and environmentally friendly. They further put emphasis on how the technology should be overseen by humans and not automation (European Parliament 2023).

Different rules have been put in place for different levels of risks. The European parliament has identified a range of AI systems and categorized them in (1) “*unacceptable risk*” (2) “*high risk*” (European Parliament 2023 p.8) and (3) “*transparency requirements*” (European Parliament 2023, p.9). The AI systems they have categorized as having unacceptable risk includes cognitive behavioral manipulation of people or vulnerable groups, social scoring and biometric identification like face recognition. These systems are seen as a clear threat and will be banned in the EU. In the high-risk category, they describe AI systems used in medical

devices, law enforcement, immigration and interpretation and application of the law among others. These systems are assessed before being put on the market. The European parliament does for example not evaluate generative AI technologies like ChatGPT as high risk, but they do however demand that OpenAI, the creators of ChatGPT, comply with transparency requirements, where content created by the AI has to be labeled as such and the data used for training is summarized publicly (European Parliament 2023).
As a last part of the act, the Parliament wants to support the technological innovations of AI systems, the act therefore requires local authorities to provide companies with a testing ground that simulates the real-world environment and conditions, so that these companies can train and develop their models before releasing them to the public (European Parliament 2023).

To sum the introduced concepts up from the broadest to the smallest entity that goes into the final user interface, there are the underlying mathematical formulas as the algorithms that govern the sorting and rearranging of information. The larger computational compilation of these equations into Large Language Models constitutes the generative Artificial Intelligence that is based on Machine Learning which refers to the ability of a computer to generate output based on pre-trained data in an almost human-like semantic manner. The output that is generated is thus dependent on the language and context that the language model handles, which further is influenced by external parameters such as biases through the content of the training data and regulatory systems such as the EU AI Act. These terms and their influence on each other in connection with the theoretical framework presented in the following chapter, will inform the analysis of empirical materials on Microsoft Copilot as a specific type of program that integrates all above mentioned elements.

# 5. Theory

We map out the theoretical landscape that the present inquiry builds upon by introducing meta-, general-, and specific-theoretical concepts structured from broader ontological and epistemological considerations to specific theories and concepts (Egholm 2014).

We start with an outline of the framework within philosophy of science building on Karen Barad and Andrew Feenberg. After that, we move from the very broad epistemological and ontological considerations to meta-theoretical thoughts on mediation and intentionality in Science and Technology Studies (STS) grounded on the work by Olya Kudina as an extension of the work by Peter Paul Verbeek.

From this general level we move to specific theory on Human-Computer-Interaction and specifically Feminist-Human-Computer-Interaction following the work of Ann Light where we focus on the concepts of users and non-users and blackboxing, that are central to the discussion of the problem statement in later chapters.

Lastly, we move to the most concrete level by diving into the specific theoretical approaches from the sub-field of Human-Machine-Communication following latest theoretical considerations by Guzman and Lewis, that are introduced directly in relation to the case of Copilot as a chatbot technology.

## 5.1 Epistemological & Ontological Framework

### 5.1.1 Phenomena of Intra-actions

The meta-theory, as the epistemological and ontological base-elements, build upon the work of physicist and scholar Karen Barad whose work is inspired by experiments in quantum physics conducted by renowned physicist Niels Bohr. Barads' research took point of departure in natural sciences, with later incorporation and adaptations for social sciences and feminist philosophy of science (Juelskaer & Schwennesen 2012, p.10).

Her analysis and conceptualization of abstract reflections from quantum physics that Bohr explored, inspire the epistemological standpoint that "phenomena" - as in any phenomenon in any domain of research - are to be seen as "*not objects-in themselves, or as perceived objects, but as specific intra-actions*" (Barad 2007, p.128).

Barad deconstructs Bohrs' critiques of objectivism and determinism that used to ascribe objects specific capabilities, properties and meanings *"independent of the necessary conditions needed to resolve the inherent indeterminacies"* (Barad 2007, p. 127).

Her review of Bohrs' assumptions results in the ontology that rejects the notion of objects' static and independent existence that is only "disturbed" by the scientist engaging with them, but rather that what we call the "phenomenon" is informed by the entanglement and interaction with it, which underlines its subjectivity and explains both the nature and emergence of objects (Barad 2007, p.119). A phenomenon in this view is the "*specific intra-action of an 'object' and the 'measuring agencies'* " (Barad 2007, p.128). Barad thus comes to the conclusion that no matter of investigation is to be considered a static entity that remains in a certain shape and appears as the same in every state to everyone who engages with it, but rather that the engagement with any matter, tangible or intangible, causes, influences and changes its mere existence and is therefore a fluid and changeable process rather than an independent object.

Drawing on experiments theorizing the physics of natural light as waves that can be measured by making them visible, it is further exemplified through Barad's interpretation of Bohrs' critique on classical Newtonian assumptions, that it does heavily matter who measures, what is measured, by which terms, under which circumstances, and along which trajectories. She underlines the notion of inseparability between phenomenon and interaction, thus between "*words and things*" (Barad 2007, p.107), and the abortion of the assumption of objects being tied to predetermined and defined boundaries. "*Knower and known*" (Barad 2007, p.107) are inseparable, and a phenomenon is always influenced and shaped by the interaction of the measuring entity and method. More so, the phenomenon as such only emerges from the interaction between measuring entity, tool, and agenda which she calls the **intra-action** (Barad 2007, p.128).

The size or reach of such impact is not the main concern, but rather the acknowledgement that the interaction, no matter how small, interferes (interference not connotated neither negative nor positive) with the phenomenon. This leads to the realization that "*the conditions which define the possible types of predictions constitute an inherent element of the description of any phenomenon*" (Barad 2007, p.128).

In essence, this comes down to the epistemological rationale that knowledge is created ideographically, that is through the intra-actions that are shaped by subjectivity as well as the inherent subjectivity itself that is underlying for processes happening through and within the phenomenon (Egholm 2014, p.28, Barad 2007, p.128). Barad manifests this meta-theoretical understanding by underlining the relational and interactional layer that constitutes a phenomenon as she clarifies how an "*interaction thus forms an inseparable part of the phenomenon*" (Barad 2007, p.119).

Her review of Bohrs' observations are thus to be summarized as the two concepts – intra-action and interaction, that exist simultaneously within, and in relation to all phenomena.

Intra-action refers to the actions that happen within the creation and development that constitute the phenomenon. At the same time there is interaction between the "*measuring agency*" and the phenomenon (Barad 2007, p. 107). While the concepts as defined by Barad stem from natural sciences and are grounded on the specific theories in quantum physics, they can be adapted to a Science and Technology Studies (STS) terminology and applied in a Techno-Anthropological context as well.

Throughout the following chapters we will use the term intra-action as an analytical concept that refers to the phenomena that describe the processes and active engagement of different actors with Copilot taking place as specific and unique situations and entanglements between human and technology. It is used to point to processes that have an influence on and are themselves influenced by the actual interaction and mediation of the Copilot technology on the human actor. Thus, **intra-action** is distinct from **inter-action** as the two terms refer to different types of processes, where the intra-action as a singular phenomenon contains processes of mediation and interaction, but is dependent on specific spatial, temporal, technical, social, cultural and professional circumstances.

To concretize the distinct meaning of each concept for the present inquiry it can be said that actions constitute the smallest entity of processes, where an action is directed from somebody or something (the actor) to something or somebody, an interaction referring to several actions that can also proceed back and forth within a given environment and an intra-action as a distinct set of interactions that is specific to circumstances and can only occur in its nature of a phenomenon that emerges out of said process only once and is not reproducible as the circumstances shaping it also change through this shaping process itself.

### 5.1.2 Political Level of Objects and Instruments

The notions of inseparability and impacts of interaction (and intra-action) between the human and the phenomenon are part of the theoretical considerations by researcher Andrew Feenberg, who takes point of departure in social sciences and adds a layer of politics embedded in interaction with technology.

In a joint discussion with French researcher Bernard Stiegler and the French scholar Gabriel Rockhill participating in a Critical Theory Workshop at EHESS in 2018, Feenberg recalls historical developments in Science and Technology studies that leads to his argumentation and call for a politicization of technology (Feenberg 2018).

He recounts how technological development has historically been viewed as a singular, linear trajectory, starting with the first human and a rock as a tool that over time develops step by step into the modern, technology-using individual. However, as he points out, this linear view of technological evolution is heavily over-simplified and flawed, since multiple trajectories have been followed with, alongside, and due to technological developments (Feenberg 2018).

Feenberg criticizes the deterministic evolutionary schemes and trajectory along which technology used to be seen. Adjusting these outdated linear ways of thinking about the singularity of a technological trajectory, he moves away from the question of whether a technology is good or bad, but rather which technology is looked at and how. Using the pharmacological example of any drug impacting the human body, he points out how every substance may be either a “medicine or a poison” depending on circumstances, dosage and intended use. Projecting this way of looking at the phenomena onto technology, they are no longer objects categorized as “either - or" within set boundaries, but interactional phenomena shaped by politics and varying, simultaneously existing trajectories (Feenberg 2018).

On the social side of the problem, he observes a collective misunderstanding of what “politics” means. His base of argumentation for the politicization of technology is the assumption that the concept of politics should not only be interpreted as the voting at presidential elections, where technology is a means to an end in campaigning and collecting votes, as it is often perceived by the broader public. Rather, politics are an element of agency in peoples’ lives, the guiding principles in everyday life, and the thoughts that influence our understanding of the world. By looking at technology with this perspective in mind it becomes clear how every technology has a political dimension to it (Feenberg 2018).

He takes these assumptions to the socio-political level by adding the perspective that every technology inherently contains a form of political agenda and political relations by design. To his definition, there is no way that technology is not political, since the development, (non-)

use and (non-) communication with and around all technologies always result in an influence on the human.

Feenberg uses the example of a smartphone clearly being a technological device but, in his view, also possessing political power and elements. This understanding underlines the importance of incorporating thoughts on potential consequences embedded in the design of technologies. Practical examples such as this one further portray the explicit and implicit power-structures that the human as a social being lives by as a result of and is a (re-)producer of in modern days which Feenberg refers to as the Anthropocene we are experiencing today (Feenberg 2018).

He essentially breaks with classic boundaries between technologies as objects and instruments and ascribes them their political power, which goes along the lines of thought that Barad brings up when abstracting Bohrs' argumentation of the inseparability of object, knower and interaction (Barad 2007, p.128).

Combining the theories by Feenberg with the ontological and epistemological accounts presented by Barad is useful to derive somewhat more concrete general-theoretical examples on present-day phenomena from the abstract meta-theoretical level.

From these very abstract conceptualizations from natural sciences melting together with social sciences, we further draw on a more specific theory that revolves around interaction and intra-action between different entities when looking at humans and technology.

### 5.1.3 Relations of Human and Technology

In traditional approaches to communication, humans and non-humans have historically, both in broader societal understanding and in the field of philosophy, been thought of as oppositions that are clearly separate and ontologically different. It was agreed that culture and nature, object and subject, were on opposite sides (Latour 1993).

As we have seen, the gap between humans and technology, and thus the notion of object and subject, has changed in past years, among other things through contributions like the introduction of the "cyborg" as a view on technology that makes technologies an extension of the human (Haraway 1991, Verbeek 2008). In this view humans and non-humans become interconnected entities rather than polar opposites and thus become hybrids – the cyborg (Haraway 1991).

In line with this way of seeing humans and technology, philosopher Peter-Paul Verbeek explores the idea of technological intentionality in terms of this hybrid thought. He looks at how there is an interplay between human intentionality and technological intentionality in human-technology relations (Verbeek 2008). The idea of intentionality is rooted in the phenomenological thought that human actions and perception is directed towards something, there is always something at the other end. It is a human-world relation.

This notion is based on the post-phenomenological thought of Don Ihde, who describes technology as a mediator of human perception. Where in the traditional phenomenological view a world and a human exist, technology now becomes part of the equation. Verbeek elaborates further in his work how intentionality plays into Ihdes preceding mediating relations. The intention that Verbeek talks about, therefore becomes mediated by a non-human element. The human acts into the world and then experiences or perceives the world around them. Based on the specific relation either the actions and or the experiences are mediated through specific technologies (figure 1).

| | |
|---|---|
| embodiment relation | (human – technology) → world |
| hermeneutic relation | human → (technology – world) |
| alterity relation | human → technology (- world) |
| background relation | human (– technology – world) |

Figure 1: The original relations by Don Ihde (Verbeek 2008)

Through this way of looking at technological relations, the intentionality is one-directional, from the human towards the world. According to Ihde, the specific type of relation depends on how the human interacts with the technology and how the technology mediates the human intentionality.

The figure depicts the four main relations that Ihde presents. They are visualized in a way that shows the specific relation of the technology with both the human and the world. It shows how the technological mediation of the world happens in different ways in different technological contexts.

Verbeek takes up these examples by Ihde and points out that the relationships are not exclusively human experiences, and that a mediating element is needed for the human to even have experiences. A technology is for example required to have the experience of a telephone conversation, because it physically enables the experience and the intention to have a long-distance communication. He goes on to describe how human intentionality can further be categorized depending on how technology mediates the relation, meaning that each type of relation also entails a certain intentionality. More specifically, Verbeek describes three different kinds of intentionalities: The mediated intentionality (described above), hybrid intentionality, and composite intentionality (Verbeek 2008).

Other than the mediated intentionality as seen in the embodiment relation, Verbeek also puts forth the hybrid intentionality, which is relevant when the technology goes beyond embodied and instead merges with the human. Prosthetic limbs exemplify this in the sense that the perception of the world through, for example the sense of touch, is shared between the human and the technology.  This shows how the role of technology changes from being specifically a mediator between the human and the world to something that changes the human character itself (Verbeek 2008).

As the third type, he points to a composite intentionality, where the intentionality is not only directed from the human to the technology, but also directed from the technology to the human. This idea of technological intentionality refers to how technology can be directed and how directedness is an essential part of the technological functions. With a composite intentionality the technology can both be directed at the world around it but can also constitute a reality with its directedness out of its own. This can be exemplified by technology that has the ability to translate something invisible to visible, such as when radio telescopes have the ability to translate, for the human invisible, radiation into visible images on a screen (Verbeek 2008). With inspiration in the composite intentionality Verbeek also calls for a new type of relation which he coins the *Immersion Relation*. It expands upon the background relation, where the technology merges with the world rather than the human and makes it possible for the world to interact back on the human. An example of this comes in the form of smart-home technologies, where the technology is integrated into the world around and gives feedback, data and information to the user as a reaction on how the technologies are used (Verbeek 2011).

One criticism of this post-phenomenological view on human- technology relations, is the lack of emphasis on the sociocultural elements of technological appropriation that is problematized by researcher Olya Kudina. Through her investigation of sensemaking of voice recognition technologies he elaborates further on the models of Don Ihde and Peter-Paul Verbeek by introducing the notion of lemniscate into technological mediation (Figure 2, Kudina 2021).

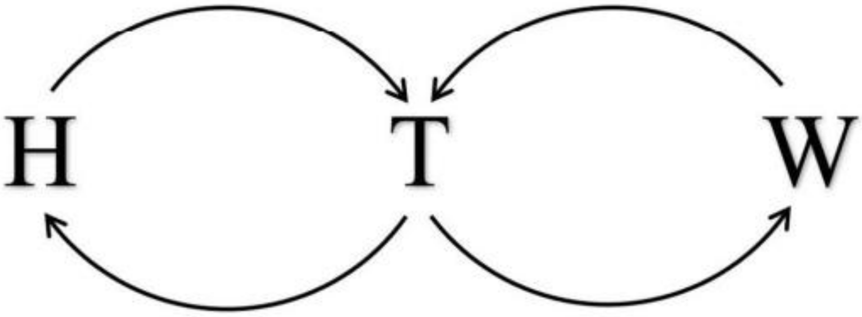


Figure 2: The Human-Technology-World Lemniscate (Kudina 2021)

The basic structure of the former models as introduced by Verbeek are preserved, but the relations between them are altered. They encompass the active human appropriation of the technology, how the technology actually gets embedded into the physical world and how the world becomes meaningful to the individual through this appropriation (Kudina 2021).

She thus introduces the mediated world as sociocultural; the relation between human and technology is more than just a mediation of the physical world. It becomes a way to make sense of and interact with the world. One aspect of this interaction Kudina specifically points out, is the responsibility of the user. She ascribes responsibility to the user to explore the practices that the technology offers, to explore how technology makes certain actions more or less possible, and how the interaction with technology also shapes interactions with others (Kudina 2021).

## 5.2 Human-Computer-Interaction

Moving from the meta-theoretical perspective to a more general theory of STS, the domain of Human-Computer-Interaction is focuses on interactional processes that happen when a human (the measuring agency) interacts with a very specific kind of "phenomenon" - the "machine" or the "computer". Human-Computer-Interaction (HCI) as a field of study is concerned with a variety of elements that concern how a human behaves around, with, through and because of digital technologies. HCI can be broadly described as "*a fast-growing, multidisciplinary field that, broadly, explores how we can design digital technologies to better meet the needs of* *people, and, dyadically, attempts to better understand people through the use of digital technologies*" (Bellini et al. 2022, p. 45).

However, in recent years a clear definition of the domain appears to be increasingly difficult to give as HCI becomes more and more multi-facetted and entangled with several research fields such as psychology, sociology and not least politics.

The multi-disciplinary field of study finds itself amidst a revolution, even called out by some scholars as a paradigm shift since "*there is no longer a coherent set of aims or goals, or accepted classification of contributing disciplines. It seems anything goes and anyone can join in*" (Kuutti & Bannon 2014, p.3543).

Part of this multi-dimensional entanglement with several domains within Humaniora is the direct impact that HCI has on humans as social beings, which points to underlying aspects of agency and politicization. This political notion is in line with what Feenberg is voicing; the need to focus on the politics that every technology and developments thereof entail. It is based on the ontological and epistemological assumptions of subjectivity and intertwined-ness of human and technology, which further can be seen as the underlying assumptions that Ann Light takes up when conceptualizing "Queer HCI" and "feminist HCI" in her text "*HCI as heterodoxy: Technologies of identity and the queering of interaction with computers*" (Light 2011).

Just like Feenberg demands to account for the political agenda in technology and technological design, Light also voices the explicit need to politicize Human-Computer-Interaction in particular, which she reasons similarly to Feenberg's argument, with its ever changing, processual nature by stating "*What HCI cannot afford to be is apolitical and ahistorical, since this suggests there is no change, whereas, in fact, there is no fixity to design for*" (Light 2011, p. 431).

Light provides a rather radical and activist approach to achieving such political action through the introduction of several methods that she conceptualizes under the broader term "queering", drawing on the Greek definition of the word to "*treat it obliquely, to cross it, to go in and adverse or opposite direction*" and by doing so to "*problematiz(e) apparently structural and foundational relationships with critical intent*" (Light 2011, p.432)

Her take on HCI thus employs a feminist standpoint, a particular attitude towards humans, technologies, and their relation, which further points to underlying intentions and internal agendas that earlier philosophers have theorized through the concept of “intentionality” (Verbeek 2008).

When problematizing the Human-Computer-Interaction and the intentions, political stances, and power structures embedded in these interactional processes, it is further important to look at how the human as an entity is defined and positioned as their entanglement increases. Within academic research, technology and users have often been seen as two separate fields or entities. However, more recent research fields like Science and Technology Studies have looked at them as two sides of the same coin. They point to the idea that users and technology are part of a co-construction where they influence each other, and where users are not only passive consumers of technology (Oudshoorn & Pinch 2003). One of the first approaches to put focus on users and break with the traditional way of seeing users within technology studies is Social Construction of Technology (SCOT). Here, users are categorized as social groups that have an influence on the construction of technologies. Different social groups can inscribe different meanings into the same technology and are therefore active participants in shaping the use and evolution of a technology (Oudshoorn & Pinch 2003).

This approach has however been criticized for ending the user investigation too early and leaving out specific “non-relevant” social groups, where there might be invisible work done or hidden meanings embedded.

Drawing back to the notion of feminist HCI, exactly this is a central point of discussion that the feminist approaches have been addressing and analyzing with the focus on how technology can also affect user groups that are not immediately deemed relevant for the design of an artifact.

To put emphasis on these invisible actors and the power relations between user groups, feminist scholars differentiated between *end-users* and *implicated users*, where implicated users are those users that might be affected by a technology without being the intended end-user or even being taken into account in the design process. These users are also referred to as *non-users* (Oudshoorn & Pinch 2003).

This means that the notion of users is broader than the immediate group of people that a technology might have been designed for.

With the terms of users and non-users it is clear that some people are actively included when a user group is thought of and some are not. It is for this reason relevant to look at the notion of inclusion in the development of technology and specifically artificial intelligence. Inclusion and diversity are terms that are used and discussed in a range of different fields like education,

organizational science, and political science and with the explosion in technological and digital innovations in the last fifty years, has the discussion also been relevant within technology development (Enslin & Hedge 2010).
Inclusion is however more than merely creating technology and functions for specific users, it is also about making it possible to use for as many different people as well. This can for one be done by giving both intended and unintended users external tools that guide them in the use of the technology (Aghdam et al. 2022).

Technological inclusion is not necessarily bound to being able to physically use the technology, but also to have the knowledge and skills to understand and reflect on the use. With this it becomes clear that the lack of inclusion is not the same as exclusion. Both the developers and the user might not know that there is a lack of proper understanding of the technology and therefore also a lack of inclusion (Aghdam et al. 2022). In many fields the main focus of inclusion is to create diversity in terms of gender, ethnicity and age. The two notions are therefore closely connected, which is also the case in the present paper. The focus is however mostly on the inclusion, because of the developer perspective of the paper. The perspective is on the developers and related organizations, and how they can be more inclusive in the tools they create. The focus is therefore on inclusion and where the responsibility of it lies.

## 5.3 Human-Machine-Communication

One particular way of interacting with technology, or computers, that lies within the broader spectrum of HCI, is Human-Machine-Communication (HMC). As a rather recent field of study, HMC conceptualizes how humans communicate and converse with technology (Guzman & Lewis 2020).

Previously, natural and social sciences have been separated into distinguished faculties in an academic context. Guzman and Lewis clear these outdated assumptions up in their text "*Artificial intelligence and communication: A Human-Machine Communication research agenda*" (2020). They explain how this new theoretical framework bridges the pre-existing gap between classic human-human communication theory and emerging technologies, and how it is distinct from overall HCI: "*What sets HMC apart is its focus on people's interaction with technologies designed as communicative subjects, instead of interactive objects*" (Guzman & Lewis 2020, p.71).

Their designated explanation of the difference between the two shows the development within recent years, especially in light of generative AI-based communication tools that constitute "*communicative subjects*", moving even further away from the traditional notion of machines or computers as objects, but now to be seen as "phenomena" in line with Barad's understanding of the concept (Barad 2007). As touched upon earlier, the way humans relate to the emerging field of communicative technologies, such as generative AI, does not fit directly into the traditional interactive or relational categories already put forth by the aforementioned scholars. This is highlighted by Guzman and Lewis in their account of Human-Machine Communication. They point to emerging technology, especially Artificial Intelligence, as something that breaks down the divide between human and machine even further:

*The anthropocentric definition of communication is predicated on a larger cultural conceptualization of communication as a uniquely human trait"*
(Guzman & Lewis 2020, p. 73).

The way we communicate and interact with these technologies changes, as responses to the new level of circular communication that communicative AI brings. This does however not mean, that the technology loses its mediating capabilities, but rather that they do not only fit on specific purposes but have multiple roles within an interaction.

*"To be clear, HMC scholars are not arguing that machines are no longer mediators. What is important within HMC is that technology is not relegated to only one role within communication(...)"*
(Guzman & Lewis 2020, p.74)

Guzman and Lewis thereby point to a change in the way we regard technology and our communication - thus intra- and inter-action with it. Observing the shifting roles in these relations and intentions marks a critical point in theoretical development that this investigation wishes to address.

Starting from the meta-theoretical considerations and moving down the ladder of abstraction to the case-near Human-Machine-Communication, we take the presented assembly of theoretical elements as the baseline for analysis of the intra- and inter-actional processes that make up the empirical data as described in the following chapters.

# 6. Methods

The overall strategy when designing the research was to build an exploratory investigation taking point of departure in Techno-Anthropological epistemology, meaning that we approached the research process with a strategy to cover all corners of the Techno-Anthropological triangle (Botin 2013, p.50), to get diverse and holistic insights as to how responsibility and inclusion are being incorporated in the development of Copilot.

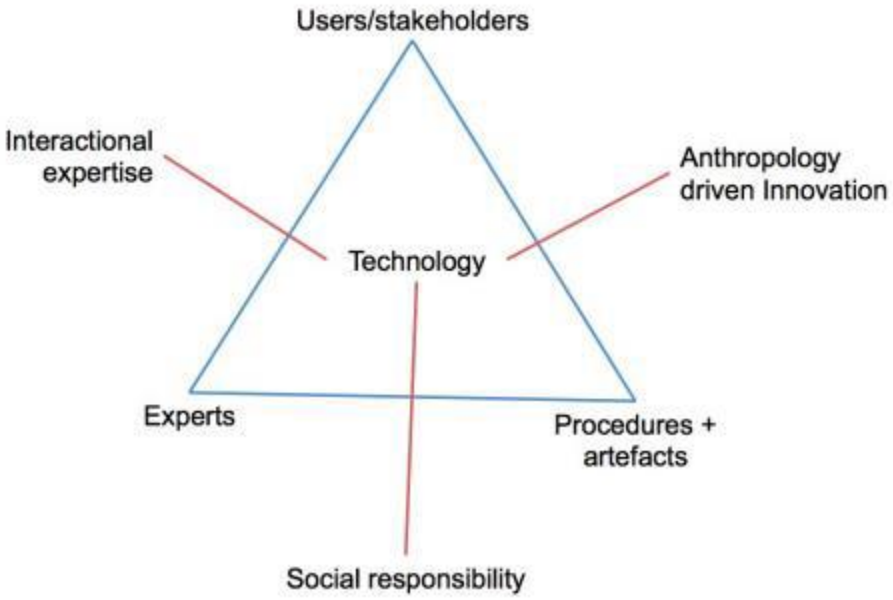


Figure 3: The Techno-Anthropological Triangle (Botin 2013)

The triangular model describes the relations between the users, developers and artifacts of a technology and how the techno-anthropologist is a professional in bridging the gaps between these three. One of the ways this is done is by employing interactional expertise between different stakeholders with different expertise and by bridging between technological experts and lay-people. The Techno-Anthropologist works as an interactional expert by having the competencies, skills and knowledge to speak the languages of different experts and translate between them without being a designated expert appointed to one specific field themselves. (Børsen 2013, Botin 2013). Entering the investigation with such socio-material understanding, we focus on how both human and non-human actors are connected with each other, and therefore do not see any of the actors as independent or isolated entities, despite dedicating specific research methods to each of them to dive deeper into their socio-material construction. We built the research design primarily on the interactional expertise to bridge the gaps between users, producers and Copilot by looking at which processes are taking place and how they influence each other. We aimed to involve all necessary parties that are the technology, the expert, and the user, therefore designing a fieldwork process that included methods that would cater to all three entities and combine them into one analysis later on.

## 6.2 Accessing the Field

We came across Microsoft's public offering for collaboration on Thesis projects for students after establishing the first preliminary problem formulation and were instantly keen to reach out since Microsoft is one of the world's most resourceful entities in regard to development of generative AI tools.

After engaging in initial conversation about aims and potential value for both sides, we derived a contract with focus on the newly released Copilot product as the specific technology and object of case study (Thesis agreement Appendix 16, p. 187).

The reason and an advantage to focus on Copilot specifically is the fact that it is one of the few programs being developed locally in Denmark and therefore accessible to us in the sense that we could get access to the development team in person, talk to developers here in Copenhagen, and partake in the organizational events related to the product. Apart from those research factors, the product is the prime example in terms of timeliness at the pulse of state-of-the-art technology within the generative AI domain and is under constant development, making an excellent case to investigate and an agile, dynamic phenomenon with room to 'experiment'.

After concretizing the research scope and project outline, we were thus appointed an internal supervisor which Microsoft offers all students as part of thesis collaborations. She not only acted as the gatekeeper, referring to the classical Anthropological definition of a person who "control(s) access resources" (Schensul & LeCompte p.39). In our case the appointed contact person also enabled access to internal information, participant recruitment, and was the communicator connecting us to all other departments and contacts within Microsoft. Apart from that she also is an expert herself as she is part of the User Experience research team at Business Central with an academic background in Techno-Anthropology making her ideal for the role as the organizational supervisor.

The “field” itself was comprised of a multi-sited, digital as well as analogue space centered around the product Copilot, more specifically the chatbot interface that users can access on their desktop, that uses written input to generate written output. This is not limited to text but can also be the format of a written prompt, e-mail, formatting etc, just not image or sound as of March 2024 (Microsoft 2024).

Due to the case study on Copilot specifically, we also chose to narrow our scope of methods to investigate interaction that is based on written input to the chatbot. While strategies in Digital Anthropology can follow a plethora of ways to engage with technology, such as using auditory or visual methods, we chose to focus on methods to investigate interaction based on written input and prompting.

There is no clear distinction between the different fields of Anthropology of media, digital Anthropology or origins from material culture studies, rather they overlap and create “*newish worlds*” across and along the established field-lines (Miller 2018, p.5). By focusing on written interaction with a specific technology we therefore aim to gain more concrete insights about a narrower space of investigation, using interviews, digital auto-ethnography, and participant observation, rather than vague tendencies on a very broad spectrum of engagement by looking at various ways of interacting with AI.

## 6.3 Ethnographic Methods

### 6.3.1 Sampling

When deriving the sample of informants for the qualitative research, we sought after people who had experience in working with Artificial Intelligence tools and Chatbots and were interested in sharing their thoughts, use cases and experiences with chatbots and LLM’s.

Drawing back on the Techno-Anthropological triangle, we wanted to include both experts and users as stakeholder groups, and thus get varying angles on the technology, use, and thoughts on responsibility and inclusion.

By not only interviewing Microsoft employees as the expert voices exclusively, we engaged a more diverse group of experts that share a common professional experience and interest with the technology, without necessarily being at the forefront of their production (engineering, coding etc) but have each their critical analysis and standpoint of working with, for, and on generative artificial intelligences.

This led to two initial groups of informants for interviews: The “Microsoft internal” group and the “external group”.
The participants from Microsoft were recruited through an open written invite that was posted on an internal chat and sent out by e-mail to Microsoft employees in Kongens Lyngby, without limitation towards field of expertise, academic background, current project, seniority level etc. The only requirement from the researcher side was a certain connection to, and experience with working on or with GenAI at Microsoft and experience and knowledge towards Microsofts' responsible AI and inclusion acts and workshops. The invitation was shared through our company intern supervisor and via her mail account.
The result of the open invite was a list of 8 employees who had contacted our supervisor expressing interest to participate in an interview. All of these eight have been contacted by the researchers personally, of which five replied to our personal mails.

Five of the interviewees were Microsoft employees attached to the Copilot product and responsible AI training in different ways, while four of them are attached to the Lyngby location, we also included an expert from the US due to their chief research position and as a comparative element to see potential differences between Denmark and the US:

- Informant 2: Manager of User Experience team in Business Central
- Informant 7: Product Manager at Business Central
- Informant 6: Chief UX Researcher at Microsoft, US
- Informant 3: UX researcher at Copilot
- Informant 5: Software Engineer at Copilot

The recruitment for the “external group” was executed by the researchers themselves through direct mail correspondence and LinkedIn. The two expert interviews were held with informants that had contacted the researchers previously through 1: participation of the researchers in a research project initiated by Teknologirådet to investigate young audiences' behavior and use around chatbots and 2: a previous semester project on ChatGPT in the Danish Education system. The two informants roles were thus:

- Informant 1: Technological expert and part of the development team of “TITAN” chatbot, a socratic large language model. The informant is also employed at

Teknologirådet and specialized in projects concerning the integration of generative AI and chatbots in education and organizational contexts.

- Informant 4: A researcher and teacher at Copenhagen University College, this informant is specialized in ChatGPT its use, employment, and education of fellow educators at the University College in its potential, risks and regulations in the education system.

Both experts from outside Microsoft are thus employed in a public organization each, and in the broadest sense have a stake towards and interest in policy making from a Danish perspective. Therefore, as far as the qualitative representation of user opinion goes, these two experts can also be seen as spokespeople for the broader public and private users outside of business or academic use context. By sampling in this way, we not only regard the end-user as a private person, but also the Microsoft developer, the education expert and the representative by Teknologirådet as users themselves.

### 6.3.2 Explorative Semi-structured Interviews

All interviews were in-depth, exploratory, open-ended interviews (Schensul & LeCompte 2012, p.135).

Prior to the very first interview, we prepared an overarching *interview guide* with some of the main (broad and open-ended) questions we wished to touch upon, letting the interviewee naturally take the conversation in their direction of expertise (Interview Guide Appendix 9, p. 161).

The first expert interview served as a baseline, from which we adapted the interview guide for the second and third one. Furthermore, we wished to explore different aspects of working with the technology and adjusted the interview guides to each area of expertise (product management, UX research) of each of our interviewees respectively, whereby the questions on responsible AI training, measures and experiences remained the same for all. It was important to us to capture the unique nuances that each field of expertise brought to the overall study through individual lived experience, which speaks to the ethnographic nature of the interviews. Giving room and time to let each expert direct the way of the conversation towards their specific field of interest led to the natural feel and interactional, conversational style that characterizes the ethnographic interview (Munz 2018, p.3-4).

For the expert interviews outside Microsoft, we adjusted the interview guide by removing all company specific questions, rather asking into their respective field of expertise and how they relate to working with generative AI and their (critical) stance towards it. In addition to that we asked about experience and opinions on inclusion, bias, and accessibility of Artificial Intelligence tools, which featured Microsoft's technologies as examples but were not exclusively used as reference points (Interview Guide Appendix 11, p. 167, Interview Guide Appendix 12, p. 169).

## 6.4 Participant Observation

The majority of Microsoft employees work in a hybrid fashion; almost equally online from their home office as well as from the office in Lyngby. This also meant that physical participant observation in the sense of following employees through their workday was neither possible nor feasible if the object and processes of study happened digitally. Therefore, we largely employed digital qualitative methods, such as digital ethnography and online interviews (Birkbak & Munk 2017).

However, towards the latter part of the data collection phase, there was an opportunity to partake in a physical event as we were invited to physically join the "Generative AI Day of Learning" at Microsoft in Lyngby, Copenhagen, that also featured parts of the responsible AI training. The event consisted of six shorter presentations by different teams, including but not limited to Copilot for Business Central, and their latest adaptations to the Copilot features and how they are integrating existing and new features into larger applications. Thus, there was a broad spectrum of viewpoints and places of incorporation of Copilot in different programs, which gave insights to different use cases from various angles and further, how each of the teams and products take measures and conduct research in product development to ensure responsible, successful, and safe interaction for end users.

The event was not directly part of the responsible AI training that is mandatory for employees to attend, but featured elements of the responsibility agenda and explanations on how these have been realized in latest updates. The actual responsibility training happens mostly online in the form of a video series which all employees who work with or directly on generative artificial intelligence have to watch in a type of online-lesson. The field day further included meeting several of our interviewees and our supervisor in person, accompanying them to the presentation and taking part in the Question & Answer sessions after each presentation.

Observations and notes were captured in a written fieldnotes file that both researchers shared in a separate digital document (Fieldnotes, Appendix 13).

## 6.5 Auto-Ethnography on Generative AI

Part of the investigation was to engage with the technology itself. As it is increasingly the case with Anthropology in general, and with Digital- and Techno-Anthropology specifically, the classic discipline of ethnography is increasingly moving to digital spaces, where digital technology both becomes the means of research as well as the field site (Miller 2018, p.10).

Therefore, when researching generative AI technologies, and chatbots specifically, it comes only naturally to engage in auto-ethnographic methods that involve the use and conversation with chatbots as part of the investigation.

We did so by working with two different chatbots specifically: The first one is TITAN chatbot, a distinct technology as part of a EU-funded research project facilitated by Teknologirådet, and Copilot as the central technology and case study of the present inquiry itself (Teknologirådet 2024, Microsoft 2024).

The reason to engage with two different chatbots was the structure of the research process, whereby we established contact to the expert from Teknologirådet prior to our collaboration with Microsoft, and therefore gained access to their chatbot before focusing the research on Copilot as the specific technology.

Furthermore, involving two different chatbots was useful for comparing qualitative insights. By engaging in conversation with two chatbots who run on vastly different technological as well as political and economic agendas, it was a helpful method to understand the differences and nuances between individual Large Language Models and user-interfaces.

As indicated above, we were in touch with Teknologirådet and their TITAN program prior to our collaboration with Microsoft. As part of the interview with their expert who works on the development of several generative AI-based projects for both private users and professionals that work with programming chatbots, we were invited to partake in their study on a newly launched chatbot based on a Socratic model with the title “TITAN”.

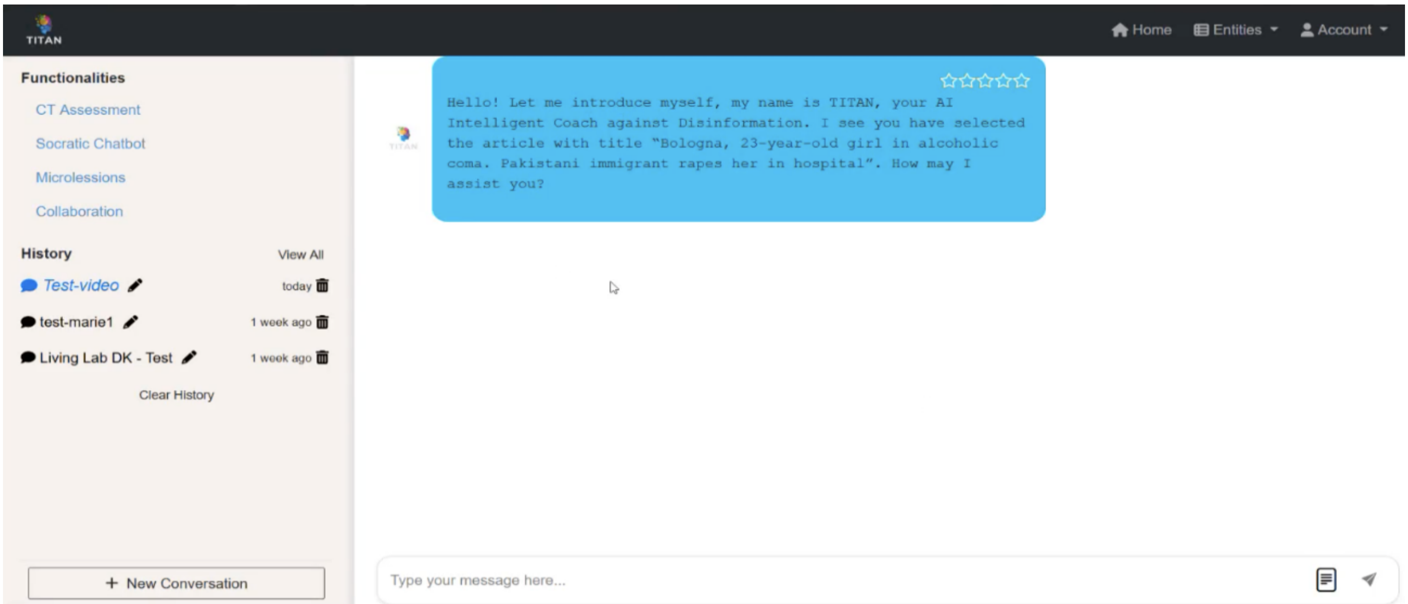


Illustration 3: Chat Interface "TITAN" Socratic chatbot (Screenshot, Teknologirådet 2024)

The project and technological tools are a EU-funded project to research and prevent the spread of misinformation, especially among young users between the ages 14 and 25 (Test Invitation Appendix 15, p. 185).

Our participation in their project therefore resulted in a short reciprocal research collaboration, whereby we contributed to TITAN's study in testing and engaging with the TITAN chatbot, while the engagement and conversation with the chatbot served as a digital ethnographic method for our project as well.

After entering the partnership with Microsoft, we focused on Copilot as the primary space for Digital Ethnography. Since the chatbot and attached programs are part of the university's Microsoft suite package, we were able to access Copilot ourselves.

Engaging in auto-ethnographic research meant to use the chatbot function and interact with the tool in the sense of testing several ways to prompt the chatbot in its three different default settings "more creative", "more precise", and "more balanced" (Microsoft 2024).

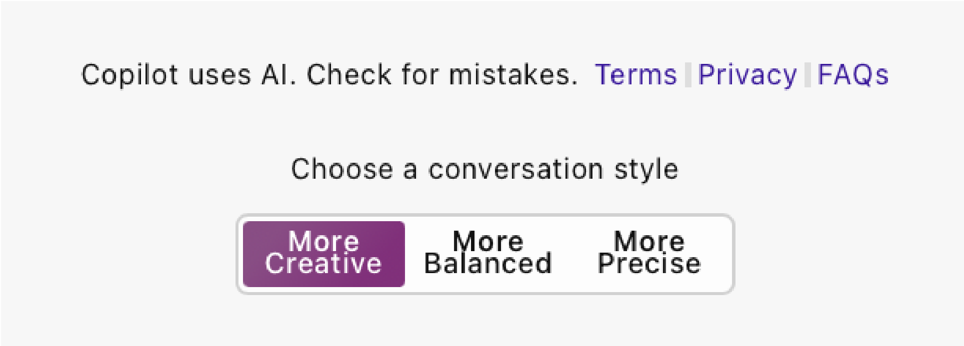

Illustration 4: Microsoft Copilot AI stamp in User Interface (Microsoft 2024)

Exploring the technology in this way covered the area of the artifact and procedures of the Techno-Anthropological triangle (Botin & Børsen 2016, p.50), as we are not only talking ‘about’ it to other stakeholders but also ‘with it’ as an immersive technology. Furthermore, this enabled the more holistic connection and integration of insights between informants and their experiences with the technology, as it “*serves as a means of linking the study of new digital technologies with an assessment of their consequences for populations*” as a key element to STS based research (Miller 2018, p.10).

## 6.6 A Survey

As described in fundamental publications on newer ethnographic methods as the result of the emergence of Digital Anthropology as a field of its own, there are different responsibilities within this domain that the Anthropologist must account for. Despite much of the investigation, the field itself “*exist(ing) entirely as a result of digital developments and the wider digital impact of the digital*” (Miller 2018, p.6) and thus the majority of methodology possibly being carried out online, it is important to account for the the group of the human, as the end-users at the user / stakeholder corner of the Techno-Anthropological triangle.

We must account for “t*he consequences of all these developments upon the everyday lives of ordinary people*” (Miller 2018, p.6). This is why we wanted to establish a knowledge baseline of the impact that such technology has on the ‘ordinary’ user.

This was achieved by supplementing the qualitative investigations with an online survey established through the tool Survey Exact, that is part of Aalborg University’s digital package for students (Survey, Appendix 17).

The total of 17 survey questions were both multiple choice and qualitative and aimed at understanding which platforms / programs people make use of, how, when, how often, and out of which motivation they interact with generative AI and specifically how they interpret the generated content and make sense of it. A total of 136 respondents answered the survey (Survey, Appendix 14).

Despite the main focus of the methodology were qualitative insights and also because quantitative surveys tend to be rather unpopular due to potential superficiality, our use of this method is reasoned by a mixed-method approach, whereby the quantitative study supports the qualitative investigations.

It is to be regarded as an additional, digital tool that accumulates a large number of qualitative responses in a convenient and time-efficient manner, that gives a broad overview over a large data space that gives context to the deeper, more focused, and smaller sample of qualitative methods.

## 6.7 Triangulation of Data

Holding the empirical findings from qualitative interviews up against a quantitative baseline served as a form of triangulation. In its essence triangulation is a method in of itself that is used to *"increase the credibility and validity of research findings"* (Noble & Heale 2019, p.67) by combining different elements of research in a study. The multiplicity of methods decreases the risk of bias by balancing out sources and covering a more diverse spectrum of interlocutors, methods and research materials (Noble & Heale 2019, p.67).

It was thus the aim to ensure validity in the sense of the extent to which the expected behavior and knowledge professionals and developers assume ordinary people to have towards generative AI tools. Triangulation in terms of methodology therefore meant comparing the attitudes and uses that professionals expected to see from ordinary people, with their own interpretations and attitudes towards the technology to the answers of the quantitative survey and our own ethnographic accounts from the participant observation and digital auto-ethnography of working with the TITAN and Copilot chatbots.

## 6.8 Coding & Analysis: Diagrams and visualizations

When moving from the empirical data collection to coding the materials, we initially transcribed all interviews using AI-based tools to generate text that could be analyzed and referenced from (Good Tape 2024). We employed a traditional approach to coding by reading through the transcripts, proofreading their accuracy and language and then moving to marking and categorizing pieces of transcript according to predetermined overarching themes (Williams & Moser 2019). Marking them in each transcript, collecting them to overarching groups of topics, and comparing the content to analyze the insights from each topic.

However, since the analysis was largely focused on making sense of processes taking place between different (non-) human entities and the particular situational description many informants offered about their interaction with Copilot or Generative AI in general, we were interested in making the multiple intra-actions, interactions, effects and intersections visually understandable.

We thus decided to supplement the text-based coding with a visual method, mapping out the interactional processes and their relationality by sorting entities, processes and effects with inspiration from Elaine Gans approach to diagrams as an Anthropological method to in her text "*Making Multispecies Temporalities Visible*" (Gan 2021). Producing a visualization in the shape of an intra- /interaction scheme as a visual model had three main reasons and methodological advantages.

First, the process of developing a visual model is reflective, a method in of itself and can give structure and comprehension to dynamics that are complex to elaborate solely by text. The curation of a visual model therefore is a method, analytical strategy and intervention combined. "*Drawing diagrams is a slow art and science that mediates between field work, analysis, and writing. It is an iterative, provisional, playful practice for working ethnographically"* (Gan 2021, p. 117).

Secondly, the diagram makes it possible to connect the explorative nature of the empirical fieldwork with the *"They open up a provisional space where empirical and theoretical engagements might start to come together or work in tension. Diagrams are great to think with."* (Gan 2021, p. 107).

Thirdly, "*Diagrams (...) are always in play, never just a fragment or memory aid, never a finished story or complete statement*" (Gan 2021, p. 107). The diagram can and must be re-iterated, altered, adapted, added to or changed as insights proceed. This goes both for the research process and the project period as well as beyond the temporal scope of the thesis, as the technological and social developments are proceeding rapidly, and it can serve Microsoft's work after the finished collaboration to keep track of the intra-actional processes and parties involved.

Due to the reflection on the usability of such visual model, we decided to not only as a method for the analysis, but also offer use the final iteration at the end of the project to Microsoft as a blue print for intra- and interactional processes, which may serve as the baseline of evaluation of similar processes in other contexts and for the future assessment of the tool and UX research. The theoretical and graphic inspiration for the final visualization was the adaptation of the lemniscate model by Olya Kudina (see chapter 5). The reason to take point of departure in this model in particular was its theoretical basis for the analysis as well as the fact that it entailed all entities that we identified as being present in our analysis as well.

Gan's way of visualizing her fieldnotes and insights from her ethnographic fieldwork are translatable to the present case study in the sense that they connect and show *"interplays, coordination, and encounters"* (Gan 2021, p.110) between different entities at different situational and interactional stages, which is why we choose to visualize our analysis of these findings in a diagram as it can serve as a practical, visually explanatory method to show these implications.

## 6.9 Reflections on the role as a Techno-Anthropologist

As a Techno-Anthropologist that collaborates with a company, the researcher can find her or himself in a split position between being a consultant to the organization and a researcher and student with a question and research agenda independent of economic or organizational goals. There thus emerges a balance act that the researcher needs to be aware of as it impacts the research and shapes the scope and methodology.

We have in this chapter explored the range of methodologies used in the investigation of Copilot and other relevant actors and tools around it. The methodologies range from exploratory interviews with developers at Microsoft and external technical experts to participant observations of an AI training day, to a qualitative survey on the relation between end-users and the technology.

There remains a professional distance to Microsoft as a company because despite working in a partnership, and researching the technology from their point of view, we are not obliged to limit the research to company internal or even promotional ambitions. We keep a neutrality that allows and calls for critical assessment and discussion of all elements and parties involved, which is also shown in the involvement of public organizations and personal ethnographic accounts of the technologies in question.

# 7. Analysis

In this chapter we present our findings from the empirical data collection based on qualitative interviews, ethnographic fieldwork and the quantitative survey.

When reviewing the empirical data, we identify several actions, processes and engagements between the different entities involved in Human-Computer-Interaction processes between human and Microsoft Copilot that have an impact on the ultimate output that Generative AI produces. Thus, we are presenting and analyzing each process and interaction where measures or actions are taken that matter in terms of responsibility and inclusion when using Copilot.

We start by presenting and analyzing the central phenomena of processes that we identify as **intra-actions**. These engagements both in conversation with, and in the development of chatbots, are based on the concept as introduced by Karen Barad and adaptations and expansion of the technological mediation model by Olya Kudina.

We dive deeper into elements of communication, social relations, language, and political power and agency, that emerge out of and as a consequence to intra-actions of human and technology, drawing on socio-political implications by Andrew Feenberg.

Lastly, we combine all elements into the existing lemniscate model by Kudina and analyze how the expansion of her model of technological mediation helps to understand and illustrate the multiple processes of intra- and inter-actions that are taking place in Human-Computer-Interaction with the Copilot chatbot.

## 7.1 What Constitutes the Intra-action

Comparing the empirical materials on the way users and developers engage with Copilot, there are a multitude of processes that constitute the actual inter-action, and the way people make sense of the technology. From the empirical data it quickly became clear that what constitutes the inter-active processes of humans and Copilot is not a singular stream of interaction, but rather a multitude of intra-actional processes happening simultaneously, which in turn inform the larger inter-actional processes happening between human, technology and world.

To elaborate on these findings, we firstly have to revisit the model of the "hermeneutic lemniscate" established by Olya Kudina:

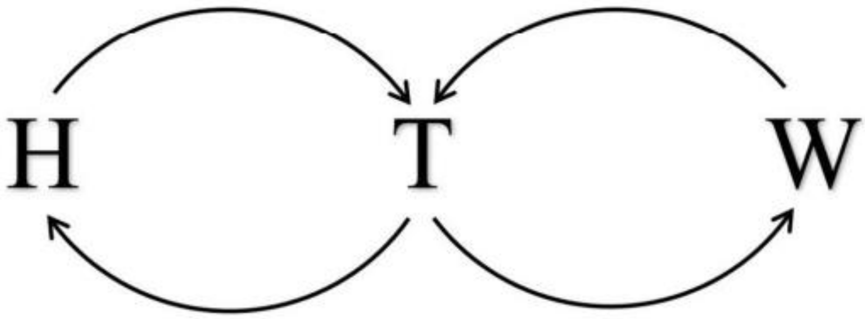


Figure 2: The Human-Technology-World Lemniscate (Kudina 2021)

When Kudina introduces her model, she uses the case of AI-based digital voice assistants as definite entities and consistent interfaces, that may appear differently in the way they are designed but remain the same stable technology with a predefined set of functions and abilities throughout the process of their usage. Her way of identifying the technology in question refers to a way of seeing it as a tool. Technology itself functions as a mediating **object**:

*"The hardware and software, type of interface, physical appearance and algorithmic underpinning all enable specific technological practices, magnifying some aspects of reality or suggesting certain patterns, while reducing the visibility of the alternatives"*
(Kudina 2021, p. 235)

She thus focuses more on the outer, tangible existence of the technology that people interact with, rather than the AI that is in the background of the "*AI-based DVA's*" (Kudina 2021). The technology as the object that has mediating capabilities shapes and changes the way the human perceives and interacts with the world, however the artifact itself remains constant and appears as a closed entity of a tool in Kudinas' analysis. Following the words of Verbeek whose theory precedes that of Kudina, the mediating effect that takes place can thus be described as from the human through technology to the world, and back (Verbeek 2008).

We argue however, that in the case of Copilot, the same identification of an Artificial Intelligence as such a stable object, the entity "T" in Kudinas model, is no longer given. This means, Kudinas' model is not falsified, as mediating processes are still taking place, but rather that there is a step missing when talking about Copilot as a Communicative Generative Artificial Intelligence.

The specific AI technology is not of the same stable, tangible, unchanged nature that a technology from a "tool perspective" employs, there is even more to the technology that Kudina calls the "*virtual agent*" in her work (Kudina 2021). Emphasizing the **agent** part of the concept, we come to identify the technology as pertaining to more 'agency', following Barad's definition of the word, than other technologies are capable of (Barad 2007). The technology no longer works purely as a mediator of perceptions, understanding, and values, but rather than that the medium itself changes and is influenced by and through the interaction with it, making the mediation process - and thereby also the technological artifact - unique and irreproducible with each interaction. The phenomenon emerges out of the interaction with it, which shapes the perception and interaction of and with the world.

To approach this complexity, we use the socio-technical adaptation of Karen Barad's concept of **intra-action** as defined above, which refers to these processes as phenomena in themselves rather than perceived objects (Barad 2007). With this intra-action it is possible to explore more of the several nuanced levels within different actions and processes that take place simultaneously than just referring to them as **the human** and **the technology**. With the example of a user interaction with Copilot in its appearance as a chatbot, the intra-action contains both, the concrete text exchanges between the user and the language model, but also a plethora of different elements that influence both the input and the output that is being created. These elements include among others the language used, the current version of the language model, and system- and meta-prompts. All these technological, social and cultural elements play varying roles in how the intra-actions are formed. What we identify as an **intra-action** in the use and development of Copilot, what constitutes such intra-action and what makes it distinct from an **inter-action**, becomes clear when looking at Kudinas' model of technological mediation and incorporation Barad's notion of intra-action into it.

This is the reason why we analyze the mediating and interactive processes as an extension of the mediating model derived by Olya Kudina, and adding to the relations she presents with the intra-action dimension. We find through the different actions taking place, as part of more encompassing action-structures, that both theoretical schemes hold, none is replacing or criticizing the other, rather the combination of both models makes it possible to visualize and understand the influence Generative AI has as an agent in the conversational and interactional space between human and technology.

The intra-action between user and Copilot is the specific phenomenon that emerges, when the user engages with the technology. This specific intra-action is different to other intra-actions created when other users interact with the technology. When one intra-action is completed, that specific intra-action will not happen exactly again in the same way because any other intra-action will also have its immediate impact on the given situation. This shows the fluidity of phenomena and their entanglement with the different actions and the world. The intra-actions are all separate but not completely blocked off from each other, as they still take place within the entangled processes of the world. Both user and technology take knowledge from the outside into the intra-action.

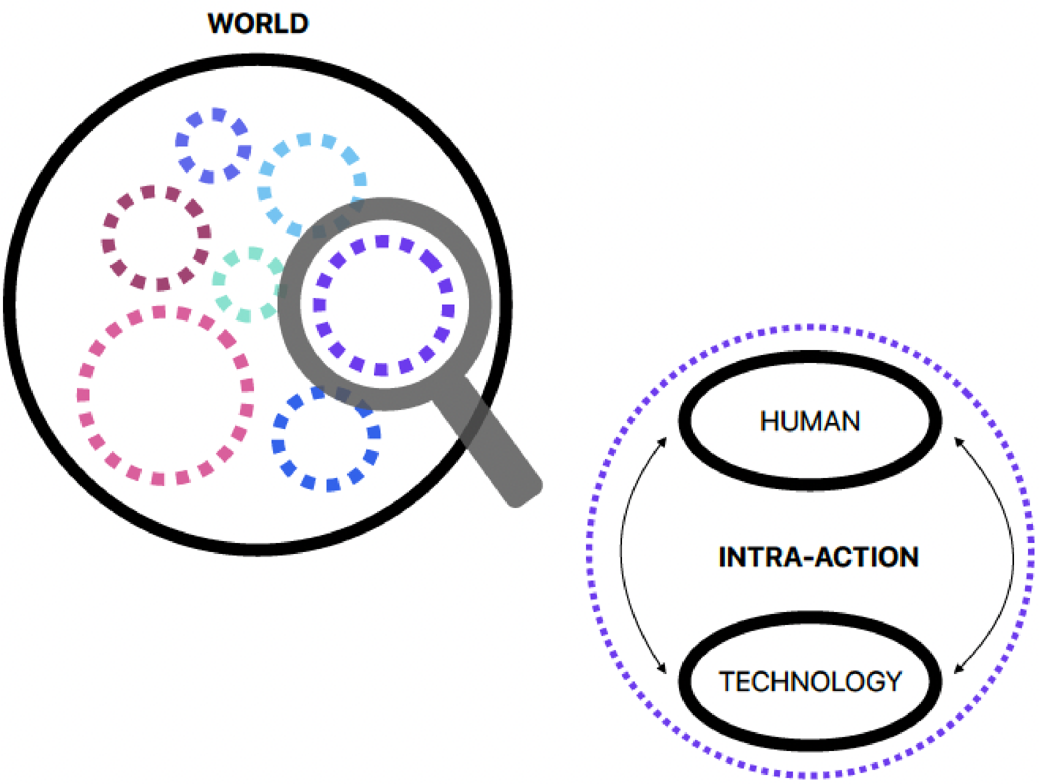


Figure 4: Human & World & Technology Intra-Action (Feldes & Nielsen 2024)

As described in the methodological approach, the intra- and inter-actional processes are best understood and explained when visualized. To put Barad and Kudina's approaches together in plain words as Elaine Gan describes it: *"no plant ever grows alone"* (Gan 2021, p.109). Just like Gan refers to her observations from her ethnographic fieldwork on how different circumstances impact the nature and growth of different plants individually, but simultaneously are part of a larger system of interactions (Gan 2021, p110), an intra- and inter-action between a human and a chatbot never occurs isolated either. There is no one-dimensionality in interaction with generative AI, as the environment, situation, circumstances and previous knowledge are different to each and every user and each chatbot or LLM is programmed differently, leading to the situationally constrained output. Therefore, the model on mediation through technology by Kudina requires adaptation and further development. As Guzman and Lewis explain, the mediation aspect is still contained within the new modular way of thinking, Kudina is not to be falsified, but extended (Guzman & Lewis 2020).

Due to several intra-actions happening as part of the relation between user and chatbot, there are several layers of active and passive processes added to the overall inter-action. Barad notes that the intra-action of the user actions and considerations create the interaction that ultimately constitutes the generative AI conversation, therefore understanding different intra-actions helps to understand the more encompassing inter-action.

## 7.2 Types of Intra-actions with Copilot

Moving from the overall analysis of the existence of intra-actions, we identify three main intra-actions from the empirical material, each existing as phenomena in of themselves, but related to each other and each of them depending on different actors and background factors. Each intra-action entails multiple actions which constitute the smallest entity of process that is one-directional within the larger phenomenon. The combined actions inform an intra-action that displays a unique way in which the phenomenon of Copilot is understood, acted upon and experienced by the different stakeholders.

### 7.2.1 Intra-action of Developer and Copilot

One intra-action we identify can be described as the processes between the developers and software engineers and Copilot when they are working with the technology and on its technical development.

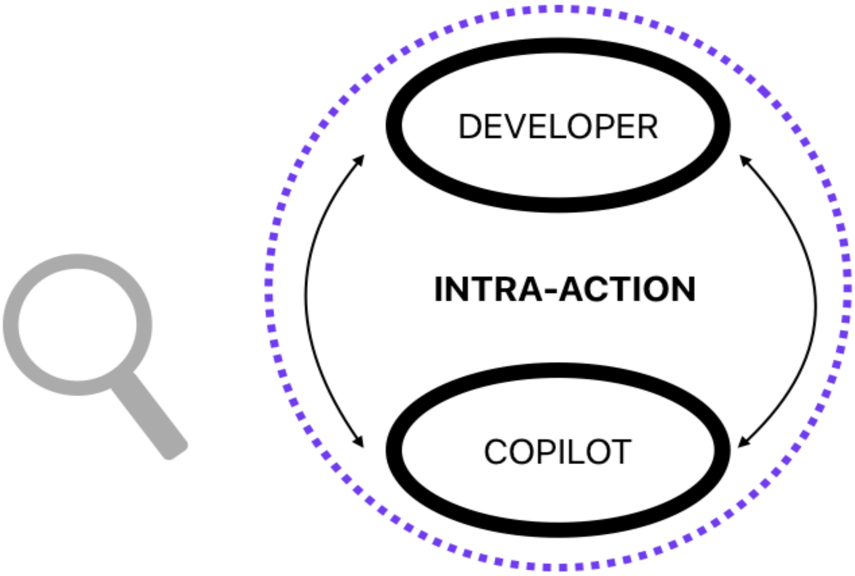


Figure 5: Developer & Copilot Intra-action (Feldes & Nielsen 2024)

The developers at Microsoft interact with Copilot from the backend. They work with the production side of the actual technology and interfaces, as they work directly on the code and the datasets that it is built on. They further create the meta- and system prompts as the overarching technical restrictions which define the input the model works on. The system prompts work as filters around the model, as they dictate what kind of information Copilot can and cannot share and which specific words and phrases it may or may not use in its responses:

*"If people ask the LLM, how do I build a bomb? There's checks in place. So the model doesn't answer that, even though it probably knows how to build a bomb.*
*And similarly, in terms of replying in a way that is consistent with the principles around diversity in terms of gender and race and all these here,*
*there's a lot of checks that have been built into the output from these models that safeguard this."*

(Informant 5, Appendix 5, p. 44)

These filters are not visible to the user but are underlying technical instructions for the model. The filters influence what kind of inputs the model can create outputs based on, and how that output is created. The filters will only be apparent to the user when they make a request which the system prompts is filtering out. As a result of this request the chatbot will actively respond to the user with a message that tells the user that it cannot fulfill the request.

The user prompt is therefore not the only instruction that the model receives to generate an output, it is also influenced by the system prompts, technical instructions and filters imposed indirectly by means of production. With this example it becomes apparent how the technical capabilities and developer choices are individual actions that shape not only the intra-action of developers with Copilot, but also the intra-action of users and Copilot. In addition to this direct action that the developer has on the technology, there are a number of external interventions that influence their action. Developers refer to the responsible AI act that is created within the company to make it possible to check and iterate where potential harmful content can occur which may be prevented through additional meta- or system prompts (Microsoft 2022).

Another external factor that influences the intra-action with Copilot which adds a geo-political dimension is the EU AI act, where developers have to take into consideration which rules and regulations impact Copilot and its functions. The purpose of generative AI technologies like Copilot is to create outputs that users can interact with or use in external contexts, which makes transparency an important part of risk management.

> *"Certain AI systems intended to interact with natural persons or to generate content may pose specific risks of impersonation or deception, irrespective of whether they qualify as high-risk AI systems or not. Such systems are subject to information and transparency requirements. Users must be made aware that they interact with chatbots."*
> (European Parliament 2023)

Because of Copilot being categorized as a general Artificial Intelligence, it does not fall under any specific risk factor according to EU law (European Parliament 2023). Rather, there are set certain transparency requirements, which for example require providers to make users aware that they are interacting with a chatbot. This furthermore shows how each action connected to the phenomena of AI generally, and Copilot specifically, affects the different intra-actions. The developers of Microsoft interact with the AI act from the EU which then influences the intra-action they have with Copilot in the way that they have to make specific features and design choices based on these regulations. As the EU AI act describes, Microsoft as the provider of an AI system has to make sure that the users are aware of the artificial intelligence elements of the program. Microsoft has met this criteria by creating visible AI elements in the user-interface and describing Copilot as *"Your everyday companion with artificial intelligence"* (Microsoft 2024) among other things.

Further, these technological choices have an influence on how users see and understand Copilot and therefore their intra-action with it. One specific action within the phenomena has consequences for the whole web of actions and intra-actions.

### 7.2.2 Intra-action of User and Copilot

The second intra-action we identify are the processes between an end-user and the chatbot. Users in this case can refer to the Microsoft Copilot customers as SME employees, managers etc, or private people in their overall use of the program. When a user interacts with Copilot by sending prompts to the model, they do so by writing text pieces in a chat format into the chatbot interface. The user interface is a consistent design that looks the same for all users. However, despite the visual consistency, individual and specific intra-actions emerge as consequences of unique actions, that form how they perceive the phenomenon Copilot.

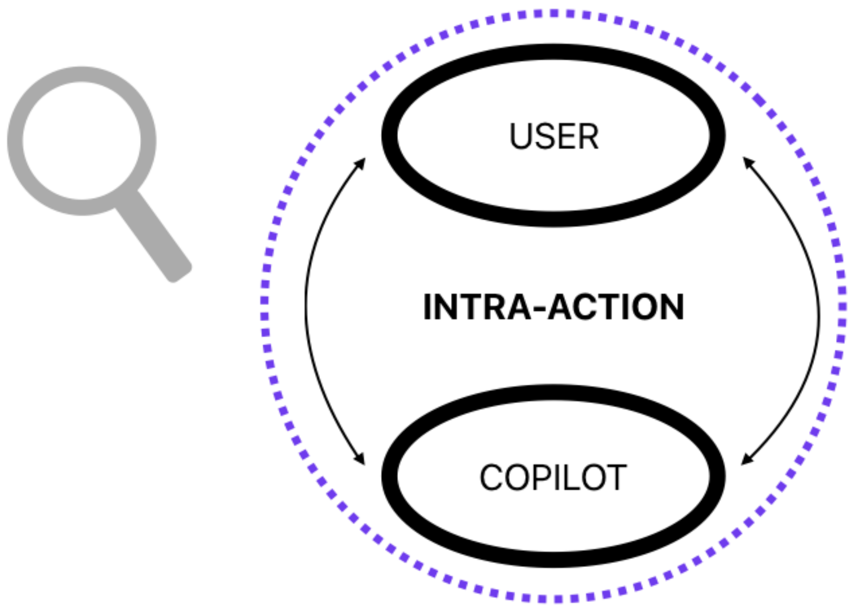


Figure 6: User & Copilot Intra-action (Feldes & Nielsen 2024)

The intra-action between the user and Copilot contains a range of different actions that are more complex than an interaction because of the entanglement of several actions and intra-actions that happen within the same phenomenon. Each prompt made both by the user and the model are used in the creation of the next following prompts, again both by the user and Copilot. When Copilot generates a new answer, it does so by computing all of the preexisting prompts in the thread both from the user and the model itself. This action, influenced indirectly by the developers through meta-prompts, creates an intra-action where the output will vary depending on how many prompts have been inserted before and how they are formulated. This means that the output changes as the conversation proceeds, because the type, quality, and content of the prompt influences the content of the output, which can have consequences for the quality, as one of the informants explains:

*"And they (teachers) often have the approach, if they ask a badly formulated question, they then get an answer that is not very good. They ask a question where they have some kind of domain knowledge, and then they get a non-optimal answer."*

(Informant 4, Appendix 4, p. 80, translated from Danish)

This means that the user's knowledge on how the model reacts to specific questions and formulation plays a large role in the kind of answer that is given by Copilot. The user is coming into the intra-action with specific knowledge on specific fields and topics, which the model does not necessarily have access to or can read from the user-prompt given. If the user has field-specific knowledge that they do not explicitly state in the prompt, the output from Copilot might then not meet the expectations of the user. This is one example of why prompting knowledge is relevant in the context of generative AI.

Another example of how knowledge about the model is important, is technological hallucination. As described earlier, the phenomenon of hallucination can happen in the use of generative AI. It occurs when the language model of the AI is not trained on data needed to answer a specific user-prompt correctly. It then fills out the knowledge-gaps with the information that matches the request the best according to the semantic likelihood of the next following word (Chang et al. 2024). This means that if users request very specific information that the model cannot access, the answer it gives is completely incorrect, but appears by the prompt to be the correct information. It hallucinates an answer that is not grounded in the real world (Alkaissi & McFarlane 2023). The less a user knows about these possible effects, the more dangerous the interaction and conversation with the chatbot becomes, as users may take false output for the truth without critically asking or checking for its validity. As observed by an informant who often works with more and less technical literate users:

*"But I also experience teachers using it in a way they definitely should not. Where they are not aware of the phenomena of hallucination. We have for example had two cases with teachers, who have put assignments from students into ChatGPT and asked: 'Did you write this assignment?' And it then answered yes or no, and they then used that as argumentation"*
(Informant 4, Appendix 4, p. 80, translated from Danish)

This becomes another layer in the intra-action between user and the generative AI Copilot, because the user either receives wrong information or has to verify the information prompted by the model, depending on technological knowledge and skill that the individual user has. With the risk of AI hallucination on top of the interaction between human and technology, the entanglement of actions and reactions becomes much more complicated and intertwined than a mere technological mediation. Taking Kudinas' critique on earlier theories of mediation to the next level, intra-actions elucidate that the mediation of the world does not happen through the technology as a tool, but especially in the case of generative AI, the technology is an actor in itself that changes not externally, but internally with each action it is prompted by. The actions between technology and humans become an entanglement of knowledge, critical thinking, and language that exceed the mediation aspect of a tool as being ready-to-use and having a consistent, stable way of employment for a user.

The developers of AI systems like Copilot have the ability to interact with the technology in ways that influence the end-users' intra-action with the technology. This also happens from the users themselves, even if it is in another way than programming the technology. An educator at the University College of Copenhagen who himself is a user of the technology, has made a guide on the use of generative AI targeted at other users without technical knowledge on the technology:

*"An attempt at demystify. It was also just about giving people some conceptual frameworks for understanding. Which makes it easier to interact with them. Try to understand that it just predicts the most probable next word. And therefore I should formulate my sentences in a specific way to get a good answer."*
(Informant 4, Appendix 4 p. 84, translated from Danish)

The goal of his guide was to give users knowledge about how the model functions so that they can create their prompts and form their interaction to give the best possible output. Instead of making a technical solution that changes how the technology presents its information, the educator creates an external element that helps users in their interaction with the technology as it is. This external guide, then becomes a part of the entanglements of the intra-actions by pointing to technological elements that are not immediately apparent to all users.

## 7.3 The who is who in Intra- and Inter-actional processes

When referring to users, we initially thought about a rather binary separation of interviewees in users (private people, non-Microsoft employees) and developers (Microsoft employees). In this notion Microsoft experts were not thought of as users in the same way as private people, as we had an anticipation or prejudice, about them having extensive experience with the use of Copilot or similar technologies as end-users in their professional and private life.

However, regardless of the division in terms of sampling and our expectation towards their knowledge, skills and experience of use, we were still as much interested in how the "developers" refer to the technologies in question, make sense of them, and use them in their daily life themselves and how that impacts their interaction and view on these technologies at work. It showed, that despite working on the production of Generative AI in a direct or indirect way, did not mean extensive experience or prior use of similar technologies that would make them experts in the use of Copilot before engaging with it as part of their profession:

*"We were as, you could say, as new to this as the rest of the world was. So we were also first introduced to this whole new, new paradigm of computing and needed to understand what it can do, what are the opportunities, what are the challenges, how does it work?"*

(Informant 2, Appendix 2, p. 33)

At the same time, many of the users from "outside Microsoft" had as much experience with Artificial Intelligence and knowledge about the technological background as Microsoft employees. Some of them also describe continuously using the tools in their daily work or in private as opposed to many of the Microsoft employees. Even more so, one informant described to engage in the technical releases and reports by companies such as OpenAI to follow along the development processes, despite not having a direct stake in their development, but rather out of a personal and semi-professional interest:

> *"We often use ChatGPT as a didactic tool. And we are trying to explore some of the possibilities it has by for example creating these custom GPTs in terms of teaching preparation"*
>
> (Informant 4, Appendix 4, p. 79, translated from Danish)

We thus reflected on the separation, or rather labeling, of different groups and came to realize that just like trying to define the technology as one consistent entity, defining the human does not make sense either. As mentioned above, the difference in the output that is generated is dependent on the phenomenon that is the intra-action itself, which in turn can be influenced by many factors that affect the human in her or his actions.

Reflecting further on the concrete terminology that we used throughout the research**,** we chose to refrain from a separation in users and experts, as it became evident that they are not different analytical entities in the way we address and characterize them throughout our research.

Instead of a methodological distinction, the varying degrees of "use" or "non-use", that is the actual intra-actions taking place, turned out to be an empirical insight and therefore an analytical distinction. Especially in regard to the interactional expertise depicted by the Techno-Anthropological triangle, an interesting shift or rather dichotomy in roles seemed to take place when looking at the intra- and interactional processes in the development and use of Copilot.

Revisiting the roles of "expert" and "user", the successful use and interaction with the technology is not a given for those who work on developing and improving the generative AI for others. More so, it is dependent on the individual intra-action, constituted of actions based on mediated use and values, that make up the intra-action which in turn informs the interaction of people with the technology - whether that is the user or developer side.

The question of user and non-user thus becomes a matter of agency as Barad points towards when defining the phenomenon in her analysis: *"Agency is not an attribute but the ongoing reconfigurings of the world. The universe is agential intra-activity in its becoming."* (Barad 2007, p. 141)

Stepping away from a binary classification of experts versus lay people into users and non-users in each their domains and from their individual intra-actional perspective thus gives room to explore how exactly such agential intra-activity can emerge, or rather already is emerging through peoples' interaction with Copilot in intra-actional processes.

*"More often than not, they're far more knowledgeable than I am. So it's this notion of trying to learn. (...) more often than not i am encountering the customer that i'm working with is encountering the phenomena that you're sort of describing it's a black box to them"*
(Informant 6, Appendix 6, p. 118)

*"Another big one was how familiar are they with LLMs? Do they understand what their capabilities are? And so they're learning what they can do and they can't do. And so what we're finding is that folks tend to be more successful with Copilot, getting valuable output out of it, the more familiar they are with sort of the bounds or the guardrails of what LLMs can and cannot do"*
(Informant 6, Appendix 6, p. 119)

It becomes evident that the agency is much more extensive for those who have practical experience with the technologies in any shape or form to begin with, thus the more cycles of intra-actions they have been through, the higher the degree of agency that emerges as a result. This on the other hand also means that less experienced "experts" may not have significantly more knowledge or skills that qualify judgment over what is "good or bad" just because of their position at Microsoft. As Andrew Feenberg points out in his piece, this comes down to decision-making power being given to employees or developers based on the general assumptions that due to their position within the organization they must be knowledgeable about what should go into the LLM and what should not.

"*Technical experts and the representatives of the institutions which employ them base many of their decisions on determinant judgments in which general categories subsume particulars"*
(Feenberg 2017, p.164)

Quite the opposite is the case when looking at some of the informants from outside the organization, who engage actively with Copilot and similar technologies, being extremely knowledgeable individuals with reflections, stakes and use cases that showcase shortcomings of the way technologies are available to them at this point.

But what we see from the empirical data from within Microsoft, is a large focus on the revision of user-generated data, not the actual user or their interaction with technologies on the development side. Internal lists of harms are established and used by Microsoft internally, and reports on user-testing are shared and followed as protocols as part of the development processes, but there is no actual interaction between developer and user in the development or user-research process.

*"Then we have internal resources where we actually guide people through responsible AI assessment. Because it is the training. Because for each harm, we have a link to a separate page or like training material, which can explain like what it is, how it can work. For example, like this prompt injection. (..) So like then we explained mitigation, etc (...) And then people can learn about each harm, see if it's applicable for their feature or if they need to mitigate it."*

(Informant 7, Appendix 7, p. 149)

To take up Feenbergs notion of judgment yet again, we see here the general application of a system of judgment that is based on the universal - the internal lists towards harmful content that are the same for the entire organization and everyone in it who is working on or with Artificial Intelligence.

*"We have like a Word document that's got like tons of sections and yeah we go over through each like thing so it's of like categorizing the types of harms that can be produced by these models what are you doing to like make sure this doesn't happen"*

(Informant 5, Appendix 5, p. 102)

Part of the reason the intra- and inter-actional processes are handled this way is the time-efficiency, competition, and therefore economic aspect that forces companies like Microsoft to align aims of deep user insights with time constraints and shortening research and development processes as much as possible to be able to release new products as the first ones in the highly competitive market of AI technologies.

*"And that is what is harder now with AI. Because everything is moving so fast. (...) People want to be the first. As soon as you have gotten an idea, then it just needs to be built. Because else someone else will beat you to it."*

(Informant 3, Appendix 3, p. 50, translated from Danish)

The intra-actional processes that people outside of Microsoft engage in are even more diverse. As the empirical data gathered at Københavns Professionshøjskole shows, there is a vast difference in intra-action and engagement with generative AI in general on the user side. While some employees are extremely proactive and positive towards incorporating the technology in their work, others are far behind their peers and do not know how to use the technologies appropriately, or refuse to make use of them entirely:

*"I don't experience a lot of maturity about how people use it. (...) It is astonishing how few teachers actually use it. It is not very widespread among the teachers at KP."*

(Informant 4, Appendix 4, p. 80, translated from Danish)

*"Where they do not have the knowledge about how to use the technology, is where they can't use it. That is something we encounter relatively often. But I don't experience a lot of advancement in what the technology can and cannot do."*

(Informant 4, Appendix 4, p. 81, translated from Danish)

The informant compares their own experiences and intra-actions with those they observe among colleagues who theoretically have access to the exact same technological tools if they wish to use them, however skills and knowledge do vary enormously.

### 7.3.1 The layers of blackboxing

Despite the examples showing that there are "user" elements to both sides, it still makes sense to analytically split users and developers and look at their intra-actions separately, as there are not only matters of direct action defining their intra-actions, but also socio-political and structural circumstances. One example of such differences is the transparency of the models to developers, whereas Generative AI is blackboxed on different levels for users, depending on the individual intra-action.

*"I feel like we've really gone back to square one with a lot of these things, especially because it is effectively like a black box, you're throwing inputs in and you get something out. (...) you have this mysterious thing in the middle doing something. So really the only way you can assert that it's doing what you want is you have all these test runs that you run against it."*

(Informant 5, Appendix 5, p.99)

This quote, looking at Generative AI from a technical development perspective, presents Copilot a blackbox to users (figure 7), because the specific actions the algorithm does upon receiving input is not fully known even for the developers. The algorithm is trained on specific data and has gotten general instructions, but each individual action is black boxed. Because of this technical embedded black box, the developers have to create filtering on the model, to make sure that it does not create unintended outputs. Despite there being technical barriers that create a black box for developers, which is the actual language model itself, the black box is much more extensive for an end-user as they are made "victims" of additionally created black boxes that encompass the actual algorithm and make everything behind the actual chatbot interface invisible, inaccessible, and therefore not understandable to them. Where the developers know what the training data is, how the model is built and what kinds of filtering that is put in place, the users only see the user interface put on top of the actual model. They can therefore only interact with the generative AI through this interface. The concept of black boxing becomes both something technological and something socio-technical depending on which perspective is taken (or given).

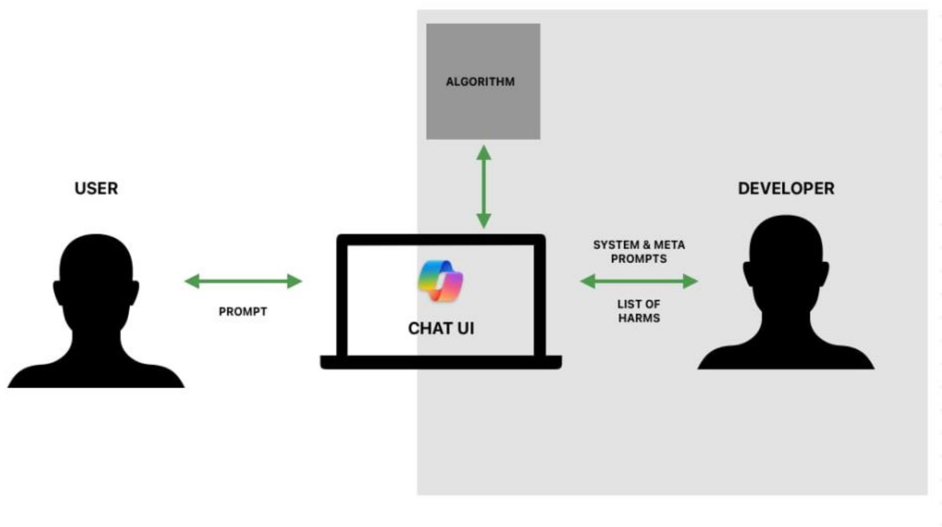


Figure 7: Different layers of black boxes (Feldes & Nielsen 2024)

*"So it's not like that, that you have sort of unfiltered access to the models that there's safety built into all Microsoft's public AI functions so that users don't harm themselves, if you will. (...) But exactly how the input filtering is done and how the output filtering is done is different from Microsoft to open AI, from open AI to Google, from Google to Facebook and so on."*
(Informant 2, Appendix 2, p.45)

The filtering changes the way that users can interact with the model and how the model replies back and therefore influences the complete intra-action both through inputs and outputs and in the way that the user perceives the phenomenon of Copilot. Furthermore, the perception will change depending on which generative AI model is used, since applied filters differ between each language model, and even within Copilot, this will change depending on which version of Copilot is used.

By exploring the user intra-action that is happening more directly, and analyzing how it is impacting the generated output by Copilot can give Microsoft insights into the possibilities to make the product even more intuitive to use and thus more inclusive for broader user groups. If agency is intra-actionally derived and shaped, then investigating exactly these processes is worth focussing on.

### 7.3.2 Technological Agency

The paragraphs above describe how all users, no matter their technical knowledge and skills, have agency in the use of this technology. However, their level of technical knowledge and experience with generative AI influences how much agency the technology itself has. Barad describes this when she explains how agency is not something that is given, but rather something that happens in the process of intra-actions (Barad 2007).

This agency is applicable to all entities involved in an intra-action, therefore the chatbot itself is also ascribed a certain agency that emerges, and is perceived, as somewhat independent of the human agency.

The agency that generative AI technologies are ascribed is connected to the intransparency of the technology. A chatbot will always produce a form of output that will make somewhat sense to the reader, as it will be grammatically and semantically sound. While Copilot offers the user a list of sources its output is based on, not all chatbots necessarily give a direct reference to sources, reasoning or context. This differentiates the conversation from a search engine that points to sources which use the specific words in their content, whereas in a chat with Copilot, there will be a generated answer written out that is convincing to those who cannot prove or do not question the validity of output.

Answers from our survey show how the majority of users get outputs that are matching their expectation at least 68% of the time, yet the majority of users are not aware of where the output comes from or how it is derived.

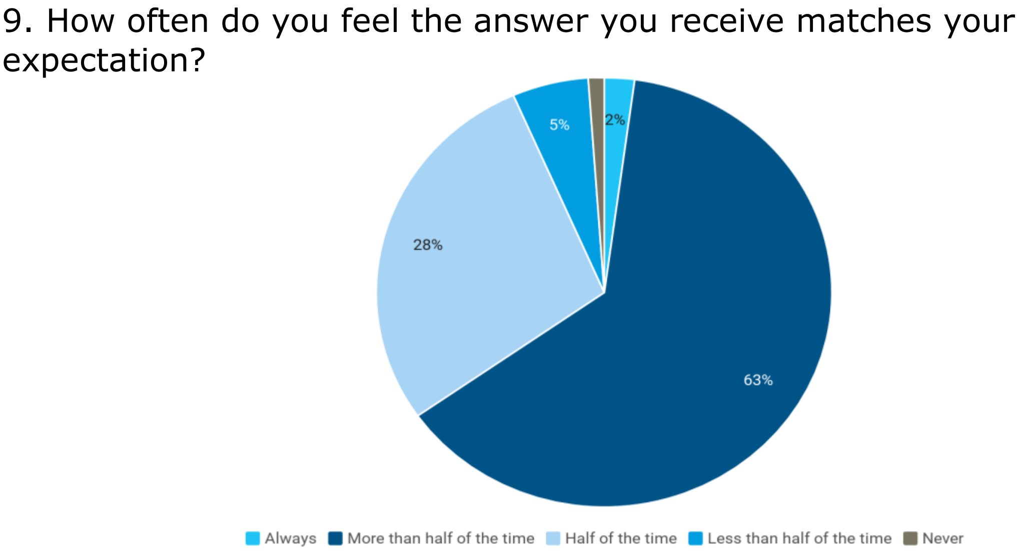


Figure 8: How often do you feel the answer you receive matches your expectation? (Survey, Appendix 14, p.181)

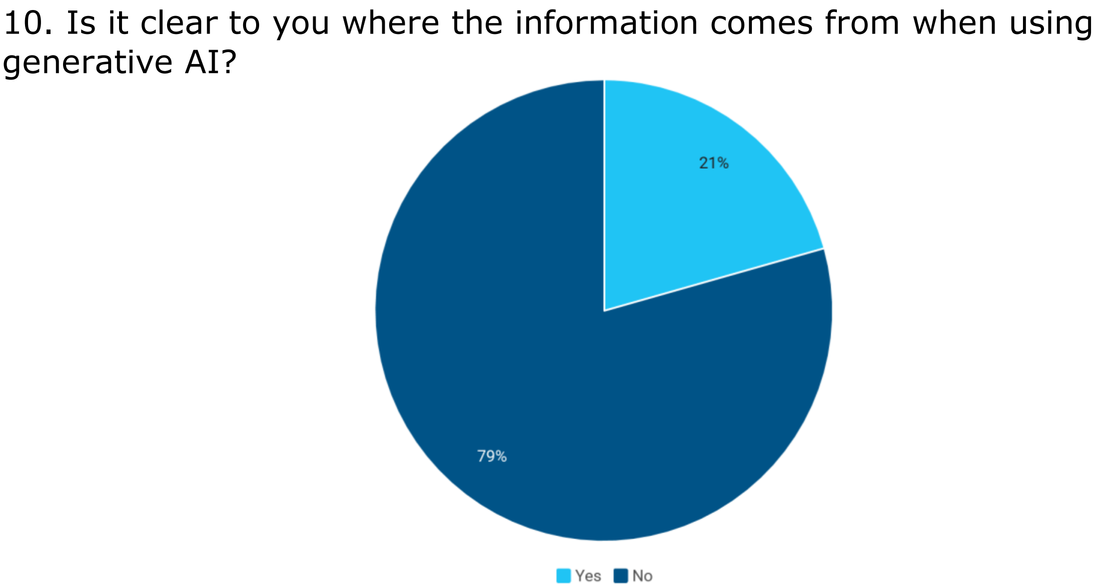


Figure 9: Is it clear to you where the information comes from when using generative AI? (Survey, Appendix 14, p.181)

The graphs point to how there is ascribed agency to the technology itself, as it mimics a conversation with a human. People receive an answer to whatever input, but there is technically no way to prove at an instant that the conversation partners' answer is truthful or why the conversation partner chooses to answer this way, let alone on which exact sources the answer is grounded on. Again, this information varies between different chatbots, where Copilot shows direct links to sources, other interfaces do not offer the same transparency.

With this comparison and the earlier point of accurate prompting being influential for the creation of output, it further becomes clear that the intra-action with the generative AI is influenced by how much (technical) knowledge the user has.

Just as it is the case with disinformation in analogue conversations between humans, the less knowledgeable a person is on a topic or a phenomenon, the easier it is to convince them of a "fact" or at least not have them question the output:

*Misinformation also existed before Generative AI, and it is a challenge for the democratic conversation. It can cause some kind of polarization between people in society in relation to a political election."*

(Informant 1, Appendix 1, p. 2, translated from Danish)

The same goes for the technological agency of chatbots. The less people know about how the algorithm works, how prompting is properly done, or how and why the chatbot derives the answer it gives, the harder it is to decode fact or fad, and therefore question the agency or "authority" of the technology.

*When you have generative AI that can increase the amount of misinformation and the quality of the misinformation (...) You are scrolling and only looking at headlines, so you don't necessarily think about whether you believe it or not, but then you go out and tell it to a friend anyways, and then it becomes a reality."*

(Informant 1, Appendix 1, p. 3, translated from Danish)

Using the example from tests with the Socratic chatbot TITAN here exemplifies the problematic that can occur as the consequence of perceived technological agency.
In its essence this comes down to what we literally term the "artificial intelligence". As described in the literature review, the generative AI technology as opposed to previous computing technologies has individualizing and personalizing abilities based on computational

memory (Epstein & Hertzmann 2023).

This leads to its appearance as a human-like working brain that is ascribed and perceived by many as containing an agency - as a way to autonomously act out of own reasoning that is not explicated to the recipient.

The technology is built in a way that its output is based on the individual prompt, but this in turn does not mean that one specific prompt always gives the same output. In addition to the base dataset, the chatbot will derive its output on previous prompts, building a sort of memory within the conversation which adds to the individuality and non-reproducibility of each conversation. The output and the way to derive it technologically remains unpredictable. This technological action is what can be described as a technological agency within intra-actions. A specifically important aspect that influences this phenomenon is the languages used when prompting.

## 7.4 Language and Terms used in the Intra- & Interactions

In review and comparison of the empirical material it becomes evident that the language that is used by different stakeholder groups has a large impact on the intra- and interactions between human and Copilot, and more generally that language has a defining meaning to the way people experience the communicational processes and therefore the relation they have with the technology.

There are different ways that users talk to the chatbot. In fact, when asking private users in our survey, the responses as to how they address the chatbot are quite diverse:

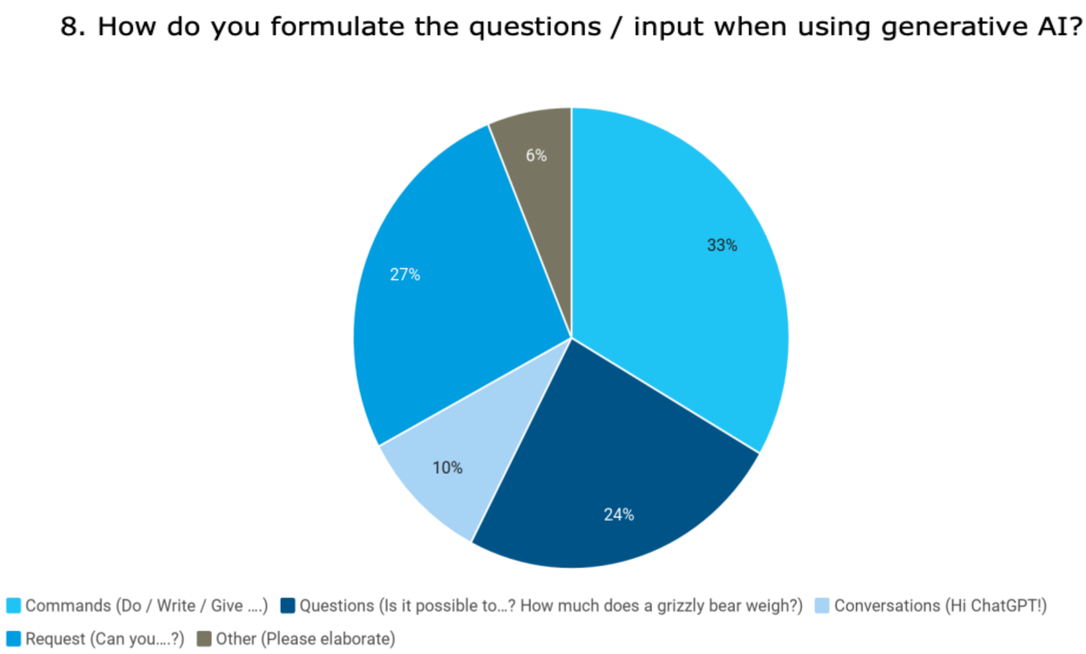


Figure 10: How do you formulate the questions/input when using generative AI (Survey Appendix 14, p.180)

As the graph shows, there is no perceived universal way to prompt the chatbot with.

Depending on the individual program, users are asked to "speak" to the chatbot like they were speaking to an actual human, more so like someone less knowledgeable about the topic in question than themselves. They are advised by Microsoft to "*think of it as if you were talking to a helpful coworker*" (Microsoft 2023), or to use the words of one of the informants: to describe the LLM's task as "*if they were talking to an intern on their first day*" (Informant 4, Appendix 4, p.82)

For the Microsoft Copilot there even is an official instruction manual that guides the user through the "conversation" with the chatbot and advises them how - or how not - to talk to the technology. Their document "Microsoft 365 Copilot: The art and science of prompting" shows users how to construct sentences and includes "do's and don'ts" for types of grammatical and semantic structures (Microsoft 2023).

At the same time, there is increasing attention drawn to making users aware of the limitations or boundaries of a chatbot, and remaining clear of those, by sticking to an instructive language rather than conversation with a companion:

*"For these scenarios of the marketing text junctions, bank reconciliation, sales line suggestion. (...) I think it's very obvious that it is a machine you're dealing with. (...) But ChatGPT is kind of the can of worms. (...) And then, yeah, it's super important that users are aware that it's not a human. So again, every chat bubble, we stamp: 'it's AI generated, handle with care'."*

(Informant 5, Appendix 5, p.113)

There is a dichotomy in the conversation with chatbots. People are supposed to be made aware of the fact that there is no human effectively engaged in the conversation taking place and that the result is not to be taken for truth as no human has checked the content before release to the end-user. At the same time, they are supposed to use formulas and engage in empathetic conversation just like with a real human counterpart.

This creates confusion among users. They have to navigate between trying to understand how the systems work technically so they can give prompts that yield useful results but at the same time supposed to show empathy, even told to use polite formulations to ensure best possible output, and almost build a social relation to the chatbot.

*"You will get a better output if you don't use the word 'please'. (...) And how should you know that? There are a lot of small tricks that you need to know. And it gives you a lot in terms of getting a better output from the language models, if you know them."*

(Informant 4 Appendix 4, p. 81, translated from Danish)

*"The thing with talking to it as if it was a human, that is just a hack in some form. To get a better output. Because it is trained on human written data."*

(Informant 4, Appendix 4, p. 86, translated from Danish)

Since there is no direct interaction between humans that connects the intra-action of each human with the technology it becomes hard especially for "less knowledgeable" end users to make sense of how they are supposed to interact with the technology and what they can expect from the output, which is also mirrored in the survey results that show that people do "speak" to the chatbot in many different ways.

And at the same time, no matter how they approach the conversation, they do not trust the output or take it a credible source in of itself:

12. Do you use the AI generated output directly or modify it?

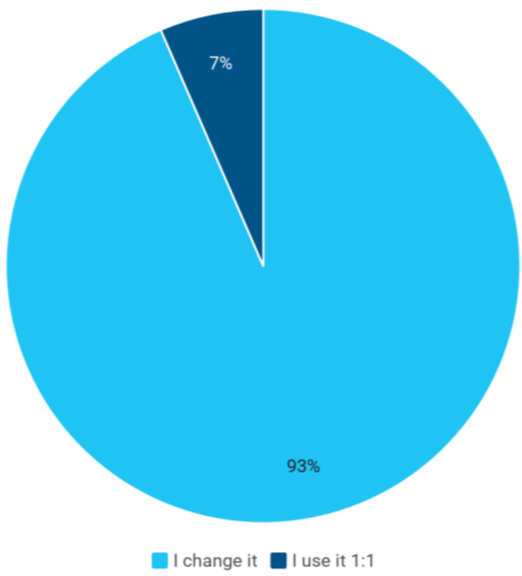


Figure 11: Do you use the AI generated output directly or modify it? (Appendix 14, p. 182)

What this further shows, is an imbalance in knowledge and skill among users which in turn reveals and points to an imbalance in the underlying power structures that are part of the use of generative AI technologies such as Copilot. An additional point to this is the makes this the individuality of the language model-based tools, as users cannot directly translate the use and interaction with one chatbot to another and expect the same result or quality of results:

*“And it is very inconvenient, because it often is for the specific language model, how it should be considered. How you should speak to it and get an answer. It is not something you can generalize for all language models. Because it's dependent on the dataset it is trained on.”*

(Informant 4, Appendix 4, p. 81, Translated from Danish)

This means that the skill to physically use a chatbot and have a conversation is not sufficient to use it successfully and neither are these skills universally applicable to all chatbots or generative AI tools that exist, but that there is specific knowledge - and access to this knowledge required.

This problem is also acknowledged on developer side, as the differences in each language model has implications for their production as well:

*"So there's a commonality between all these, uh, AI services in that they are built on the large language model. But exactly how the input filtering is done is different from Microsoft to open AI, from open AI to Google, from Google to Facebook and so on each, each."*
(Informant 2, Appendix 2, p. 45)

The solution that Microsoft is working with, is making the separation between human and machine visually visible as much as possible through interfaces looking differently to the users as well as written statements in the chatbot interface that clearly state their interaction with an algorithm, not a human:

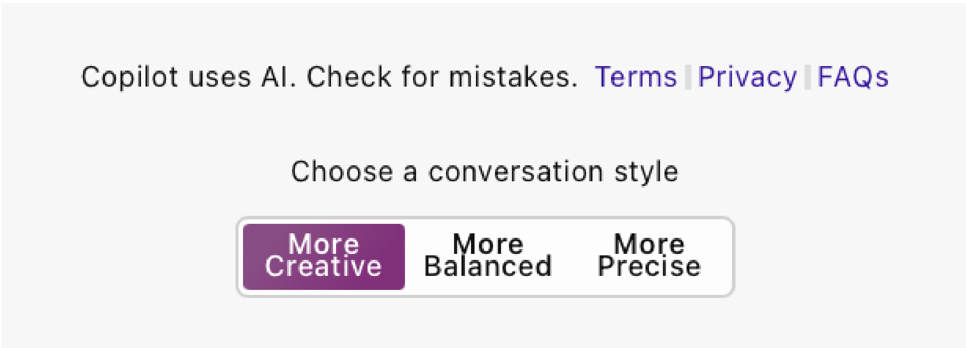


Illustration 3: Microsoft Copilot AI stamp in User Interface (Microsoft 2024)

The reason to integrate these clear statements is to create awareness and attention to the difference in the conversation is explained further by one of the informants:

*"So it needed to surface in the UI in a new dialogue and a new journey, sort of a mini journey for how you are reviewing these transactions. So the user was able to say, OK, now I am accepting a match that was made by generative AI. And we state clearly AI generated content may be incorrect. It's right there in that dialogue. So we're trying to put some terminology and some context. And we're trying to put some content in that experience that makes it transparent to the user what they are dealing with."*
(Informant 2, Appendix 2, p. 39)

As the quote exemplifies, speech, conversation, and a clear language are used to point out which conversational relations and processes are taking place when interacting with a chatbot, however this is a matter of creating awareness of the problem, not offering an active way of influencing the outcome or conversation for users.

Comparing the different levels on which language determines the outcome and success of the intra- and therefore ultimately interaction with Copilot, shows that language is associated with knowledge, skill, and competency.

Language to instruct the chatbot by using the right phrasing is an integral part of the conversation, just as is the knowledge on how to master the art of prompting and the competency to prompt the chatbot with the "right" content to derive the most valuable and factually sound output.

## 7.5 Merging the Models

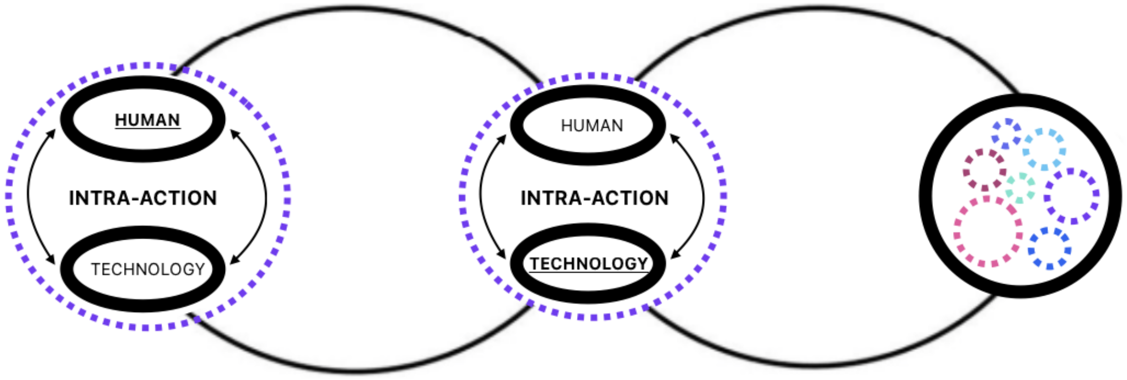


Figure 12: Merged model of intra- and interactions (Feldes & Nielsen 2024)

Ultimately, when combining all intra-actions as analyzed above into the original lemniscate by Kudina, we derive this extended version of her visualization.

What we see is the integration of Barads' notions into the existing mediation framework by Kudina. As unfolded above, the H for human and T for Technology in Kudinas' visualization are not to be understood as stable, binary, static, or reproducible entities anymore. Following Barad's notions, the phenomenon that is either the human or the technology are both shaped and changed through the technology during the use process.

All entities of course still exist in their tangible and physical representation, that is the chatbot interface will look the same to the end-user with each new prompt and will also look the same for all users who engage with it. Just as much will the physical human on the other side of the screen remain in their physical nature. But the internal processes that are happening throughout the intra-action lead to a change in intangible existence of all entities, making them the 'phenomena' in their situational uniqueness, that Barad describes.

As we have shown through each of the individual intra-actional processes throughout the analysis, there are multiple action-points for both users and developers that influence the output a chatbot generates. Whether they are language or technology based, and which side of the chatbot (that is user or developer) a human is located, makes a difference for the use of the technology and the results it produces.

As becomes visible through the elaboration on the different intra-actions as well as the combined figure above, the actionable space for developers is much more encompassing than it is for users. While developers can manipulate the chatbot directly through technological design choices, users are "limited" to the interaction with the interfaces that are made available to them. This further means that only a minimal part of the technology itself is visible and accessible for them, establishing a larger blackbox around the technology as a whole, as opposed to the algorithmic blackbox that developers are confronted with. These findings open up a space for discussion of the socio-political and structural implications that we unfold in the following chapter.

# 8. Discussion

There are four main points emerging out of the empirical findings that we wish to discuss and elaborate on from different angles:

(1) We presented a number of intra-actions that inform larger interactions but discuss the lack of direct interaction between users and developers as two human counterparts.

(2) Power relations become visible from the degrees of agency that emerge as a part of the intra-actions and we therefore discuss these power relations with regard to several implications for access, regulation, use and development of Copilot.

(3) As a consequence of these power structures and imbalances a number of black boxes appear in the different processes that add to the mystification of Copilot and generative AI in general. Therefore, it must be discussed which possibilities of demystifying this technology in terms of power structures and blackboxes there are.

(4) The interactional expertise of the Techno-Anthropologist must be discussed as well in light of the changes and developments in dynamics between the parties. We argue for extension and alterations in the academic discipline to make room for emerging technologies such as generative AI as an actor on their own.

## 8.1 No Intra-action of User and Developer

What the above elaborated intra- and inter-actions also show is the absence of human interaction. Out of the empirical data gathered, none of the developers, that is the engineering teams who are coding the actual models and system prompts, mention direct interaction with a user as part of their work. Rather, the “list” of harmful content gives the developers insight to adaptations needed to be made in the LLM itself, which is followed by test-runs. The list contains requirements like:

*“Identify and prioritize demographic groups, including marginalized groups, that may be at risk of being subject to stereotyping, demeaning, or erasing outputs of the system.”*
(Microsoft 2022)

These points are framed as checklists and tasks that developers reflect upon and hold the models or work in progress up against, it does however not mean that they have to directly interact with the “demographic groups” they are proving their model for (Microsoft 2022). The developers are aware of potential harms on specific user groups, which they handle in a user-centered approach rather than a participatory manner (Sanders & Stappers 2007).

That means throughout the entire development process of the chatbot, no direct interaction between developers as the software engineer programming the model and user interface and (potential) user is taking place.

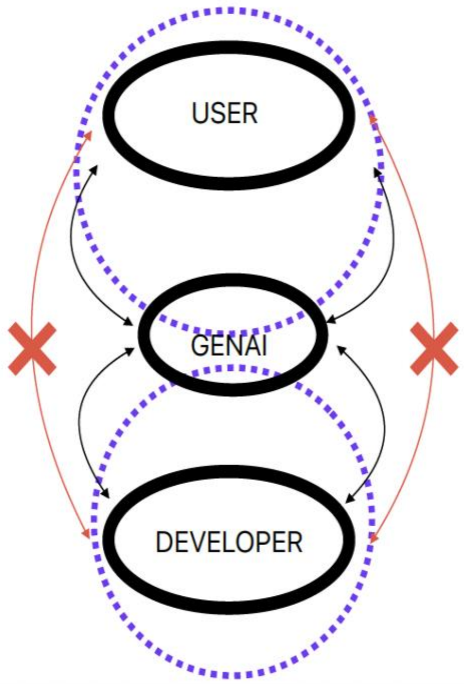


Figure 13: The missing interaction between users and developers (Feldes & Nielsen 2024).

There is however an important human component involved in the that intersects with both developers and users, but not as much the technology itself, which is the designated UX research. During the product development and it has been released, there are test runs and UX research conducted to investigate how end users interact with the technology, where they fail, and where there are blind spots for developers because of their knowledge and skills in handling the technology is preventing them from misuse or unsuccessful prompting, that can rather easily happen to a “novice” that most end users are.

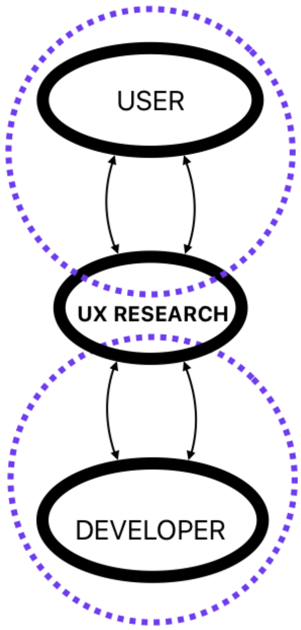


Figure 14: Interaction of UX Researchers with users and developers (Feldes & Nielsen 2024)

Being the only direct human link between experts and users underlines the importance of UX research as part of the development process, which in many ways roots on Techno-Anthropological methods and theories. As the informant describes, the company is well aware of the importance of this kind of qualitative research:

*"They were thirsty for research. And they understood their users better. I needed to establish the process of gathering knowledge and doing it at a good time also"*
(Informant 3, Appendix 3, p. 50, translated from Danish)

But taking the notions by Barad into consideration here again, what happens when 'measuring' these entities, that is understand, study, research or evaluate either elements of the lemniscate, there is no way of reproducing them exactly like a user would have done, as external circumstances will differ and different knowledge and skills of users impact the conversation. The technology changes during the use processes when it generates new knowledge by letting the algorithm 'learn' by incorporation previous prompts (Adamopoulou & Moussiades 2020). It therefore must be taken into consideration that the perceived agency of the technology itself as a consequence of the intra-action will further impact the user experience on the technical level.

The same procedure holds for the ‘phenomenon’ that is the human who is equally shaped and influenced by each interaction with the chatbot.

This further means, that by means of trying to reproduce an intra-action between an end-user and Copilot, UX researchers at Microsoft will not be able to evaluate accurately from pre-established lists of harmful inputs, whether the actual intra-actional process indeed is a harm to the end-user, or a matter of an individual prompt situated in a specific entanglement of circumstances and user-skills / knowledge, that are negatively, rather than positively influenced if manipulated by meta- or system prompts.

What is needed is direct conversation and research with and among users, to get as close as possible to the direct intra-actional experience as possible without interfering.

A challenge that comes with this type of research and classic anthropological research as a whole, is the inevitable interference with the subject of study upon entering the field. In the specific case of investigating users in their intra-action with Copilot there are obstacles are both the access to user information like prompts due to GDPR regulations that prohibit the sharing of actual user data, as well as the challenge of finding eligible candidates as test-users that have the necessary technical literacy to reflect on their intra- and inter-actions with the chatbot in a way that can be used for UX research.

The discussion shows the difficulties of involving all necessary entities into the research and development phase, which becomes even more complicated if the technology in the question is not a stable and fixed tool, but a fluid, dynamic phenomenon that brings yet another agency and active participant into the entanglement of processes.

## 8.2 Power Structures Between Actors

Taking the roles and challenges for each of the involved parties further, it is important to discuss their degrees of power and influence on the dynamics when investigating inclusivity into the development process of new AI based technologies. The empirical data on both developer and user side gave insight into the kinds and levels of agency that emerges in intra-actional processes with the technology. We have further analyzed how these types and degrees of agency vary depending on the stakeholder group.

Developers have the technological power and agency to implement meta prompts that are invisible to users but influence their prompts. This invisibility creates a power structure, as it limits how much knowledge about the generative AI a user can have as opposed to developers. The reason for ‘hiding’ these meta prompts from the user are however not intended to keep information from the user in the sense that they should be restricted in optimal use, but rather to ensure that the use of the AI is within the companies own agenda. When the user does not know the specific meta prompts and 'security checks' that are put in place, they do not have the same tools to go around those technological restrictions. In this way, the developers use their power and agency on the technology to make sure that it is used in a way the company deems ethical.

Taking up the concept of blackboxing again, its meaning and impact on the intra- and interactions with Copilot can be discussed as a result of the intra-actional processes.

As we have identified by using Karen Barads' definition of the concept, agency - that is both Copilots' and the human agency, is a result of the intra-action. Agency develops out of intra-actions, it is not a given attribute or ascribed asset of an agent (Barad 2007).

The black box has a technical dimension and is tied to the actual algorithm. Developers have the power to program and technically influence and guide all processes and actions that go into a language model by putting up the guard rails of system- and meta-prompts:

*"You can control what you put into the model. (...) but Microsoft are adding some additional stuff to what you say. So it's not like that you have sort of unfiltered access to the models that there's safety built into all Microsoft's public AI functions so that users don't harm themselves, if you will. But that's a different level of security than OpenAI are providing or that Google are providing to their services."*

(Informant 2, Appendix 2, p. 45)

As the quote describes, the access to the backend of Copilot allows for control over the model. To developers, the black box is the reasoning process inside the actual algorithm that makes it impossible to foresee which exact choices a language model will make and what reason it has to do so. All other processes and technological design choices that influence how the chatbot will look like, what input it will "accept" and which output is released are transparent and accessible to them. They are so to speak in power to gatekeep the access for user input and also chatbot-output. The black box therefore is a technological phenomenon from their point of view and intra-action.

This further also means that developers have the power to determine what is transparent, or "hidden away" from the users through their intra-actions. The companys' evaluation of what constitutes harmful input or not, guides the design and programming choices that influence the chatbot and its output that users ultimately engage with:

*"The design folks and people in Redmond have from day one been very focused on broadly the agenda of responsible AI. (...) And then there is the bias that are sort of part of these models. What are they trained on? What does that mean for the answers that come back in terms of all the different biases that can be hiding in the training?*
*And the responsible AI is a broader agenda than just bias. It is also related to how do you annotate what comes out. How do you get out of the model so people know the limitations of it? How do you avoid that users are, that the AI is suggesting actions that the user takes and then bad things happen? So, responsible AI is a very broad umbrella."*
(Informant 2, Appendix 2, p. 36)

As the informant describes this not only sets an agenda for the language and wording that is supposed to be used but also the content and regulations that they "allow" for users to prompt the model with or receive a result on.
Talking about the *bias* that goes into the development of a model means setting the political agenda of what is deemed right or wrong, allowed or prohibited, when working with Copilot. The bias, as understood in the origin of the concept as the pre-understanding and point of departure for decision-making based on previous knowledge and experiences, informs what is added to the list of "harmful input" of Microsoft, and will thus look different than the "internal lists" of other companies.

*"(...) formal bias is imposed through the rational procedures that govern the "world of things."*
*"Identifying bias in the world of things is difficult and persuading others of its existence even more so. The individual confronted with such bias must reason "reflectively" from his or her own particular case toward a possible universal under which it stands, and then communicate that perception publicly."*
*(Feenberg 2017, p.166)*

Feenberg hereby gives a more abstract description of different biases that are at play simultaneously, yet hard to identify and understand across individuals and the broader public due to their individuality. It points to a part of the problem of agency and power structures whereby the awareness and visibility of individual bias is hard to point out in existing formal systems. Translating these insights to the case of Copilot, it can be said that Microsofts' bias is amplified through their chatbot as the organizational (formal) bias, informed by a political agenda that materializes in the list of harms, informs what can be accessed or prevented from the Large Language Model.

Due to this power, user bias can be impacted, through system- and meta-prompts that restrict or dictate which input and output can and will be handled by the language model and in which way. A system- or meta-prompt in this way can be seen as an additional layer of formal bias that is put before and after the technological blackbox of the algorithm itself.

With these instruments of power in their hands, this further means that the concept of blackboxing becomes much more of a socio-political phenomenon than choices of product design and usability.

## 8.3 Ways to Demystify Existing Blackboxes

Looking at the blackboxing from a user perspective, it is perceived by those who consider themselves active users and experts through the emerging agency and ongoing entanglement with the technology, as if mystifying measures are actively employed as a way to exclude them from being successful users and to ensure their agency does not exceed the economic or political goals of the organizations producing them.

> *"I think it was a giant step backwards. When the GPT-4 model was launched last year, without a related academic paper. I was waiting that evening, and we were reading the communicative paper that they always released with their older models. And then it was a completely empty technical rapport. Where they did not write anything about how it was made. And since then, have we not gotten a proper research paper alongside the big models. And that makes it a closed field."* (Informant 4, Appendix 4, p. 94)

With this experience of being purposely left out of the conversation, not having an official impact on the developments that highly influence peoples' lives as users of emerging technologies, we can ask ourselves what alternative ways to have an impact as a user remain.

We discuss two main ways in the latter part of this chapter:

1. Independent organizations using their resources to educate the broader public
2. The Techno-Anthropological Intervention: Including third parties with intra- and interactional expertise to bridge the perceived trenches that divide entities.

### 8.3.1 Education as independent means for Competence Development

A way to increase the inclusion of different users in the development and usage of these technologies is the sharing of knowledge. This knowledge sharing can happen from the distributors of the technologies, but also from tech savvy users. An example for such efforts has been demonstrated by external actors in relation to generative AI.
As we have previously introduced, there were two representatives from public organizations involved in this research as external experts on generative AI technologies and chatbots. The first is a project manager at the Danish Technology Council and works on an EU funded AI support tool named TITAN. The goal with the tool is to combat misinformation created by other AI tools and make citizens more aware of the risks of misinformation when using generative AI.

*"So it will ask you questions about what the article you have read is about, what can the intention behind it be, what kind of things does it emphasize and what could the source have of interest?"*

(Informant 1, Appendix 1, p. 5)

The tool is a chatbot itself and makes the user think critically about an article by asking different questions about content, biases and motives. The user can then reflect and answer those questions and the tool will keep asking questions and never give any other type of output. Here, external organizations take action in informing citizens about the potential risks of using generative AI and giving them a tool to become more reflexive and knowledgeable in the world of AI. By doing this, they try to make the technology of AI more inclusive by giving more people the abilities to get actual value out of the technology.
Another example of external actors making AI technologies more inclusive is our other informant who is a consultant at the department for Digital Competency Lift at the University College Copenhagen. He has made a textual guide for non-technical users to understand the

basics about generative AI, how they function and what kind of data they create.

> *"For me it was important to remember how confusing it was to enter this field in the beginning. (…) And how many intermediate steps are needed to understand that it is not magic. It was to try to demystify it. "*
>
> (Informant 4, Appendix 4, p. 84)

The quote showcases the difficulties that end users encounter upon introduction to generative AI tools and how a sort of guidance like instruction manuals and handbooks functions as another type of assistance to give users more knowledge and information about the basis for generative AI and therefore helps to demystify the blackbox.

### 8.3.2 Techno-Anthropology as an Intra-actional Expertise

The discussion on intra-actional entanglements and their influence on what defines a user, non-user, developer, expert or non-expert evidently has an impact on the way we view and engage with Artificial Intelligences in this still rather new era of Human-Computer-Interaction / Human-Technology Relation.

As Techno-Anthropologists, researchers with focus on exactly these meeting points between human and technology, we must stay in touch with these developments by actively reflecting and iterating our own methods and views alongside these developments. As Donald Schön points out in his text "The reflective practitioner", we are influenced by the immediate response of a situation we engage in and adapt theories and methods accordingly, throughout the actual research process and in evaluation of research processes in hindsight (Schön 1983).
This means we not only have to discuss the implications out of a Techno-Anthropological perspective looking from the inside of academia out to organizations and societies, but we also need to take into consideration what implications these developments have for the academic discipline itself.
There are contributions in the academic field that take potential extensions of the Techno-Anthropological triangle that the framework is grounded on into consideration.
Techno-Anthropologists Anne Marie Kanstrup and Pernille Bertelsen describe how a participatory dimension can be put onto the model adding a third dimension to the triangular figure, turning it into a Techno-Anthropological pyramide.

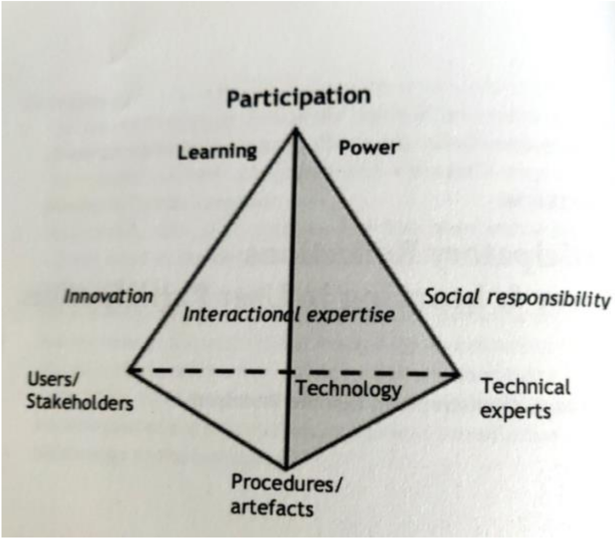


Figure 15: The Expansion of the Techno-Anthropological Triangle (Kanstrup & Bertelsen 2013)

This new dimension emphasizes the importance of social and cultural aspects such as power and learning in design processes. Their goal is to make sure that the interactional expertise is sensitive to the difference in power and learning between the different stakeholders and expertise (Kanstrup & Bertelsen 2016). Taking this expansion to the yet another level, we argue for another revision of the model, where it is being accounted for intra-actional expertise that informs and shapes the role of the interactional expert.

In the way the Techno-Anthropological triangle is currently employed, the interactional expertise does not involve creating a bridge of knowledge between human actors and the technology itself, in the sense that the artifact appears as a static object that is not seen as an expert in itself.

The technology is more to be understood as a tool, a tangible entity as unfolded in the analysis. As of now, the area of the triangle between the technology and experts instead is the social responsibility and for the area between artifact and users the focus is on anthropology driven innovation. However, with the introduction of intra-active entanglements through generative AI technologies like Copilot, technologies' role as an agent changes. As we have shown above, this also means the model should be reiterated to account for these technological developments in the Techno-Anthropological discipline as well. Thus, the triangle should be altered to encompass the agency, language and expertise that artifacts are themselves ascribed as well.

Generative AI technologies are more than mediators of a user's perceived world. They are rather complete phenomena in themselves that appear and transform in intra-action with users, developers and other artifacts. Artificial Intelligence in this sense does not mean that the AI has natural human-like intelligence, an inherently human attribute, rather the processes that we identify as intra-actions are mimicking human-like interaction. Rather a third dimension opens: a chatbot that is communicatively and processual close to the human, enabling intra-action that changes the phenomenon through the engagement with it, but remains the same “tool” algorithmically and a “machine” in terms of technical interface.

The interactive nature of technology is already embedded into the model in how technology is in the middle of different relations between actors. The geometric shape and idea of the triangle however restricts the types of relations and therefore relevant competencies between the three corners. When the side of interactional expertise is opposite the corner of the artifacts, it appears as though there is never a need to translate and bridge the expertise of the technology to any other actor. This is not necessarily the case, with the introduction of intra-actions and the intra-actional expertise. We argue that three specific actions in the intra-action with technologies like Copilot are not accounted for in the model as it is visualized at this point

1. The technological agency present in the intra-action with communicative generative AI
2. The language of prompting in the intra-action with communicative generative AI
3. The “expertise” of the communicative generative AI that is based on and informed by

Based on these actions and how they are a prominent part of the AI intra-action, we call for an intra-actional expertise in addition to an interactional expert. The difference between the two, comes visually in how many relations each actor of the triangle has. With the interactional expertise, one side, or one specific competency bridges each corner. The bridging between users and artifacts happens with anthropological driven innovation, where the focus is on how classical anthropological methods can find gaps of knowledge and interests in the interactions between different relevant actors and stakeholders with the focus on a specific technological solution.

We argue that because of the intra-active nature of emerging technological phenomena, there are even more competencies at play. The notion of technological hallucination illustrates the importance of the bridging of knowledge between artifact and users. A generative AI hallucinates by creating wrongful outputs about a specific topic because it has no data on that topic. Hallucination here becomes an example of how awareness of the expertise at play is important both from the AI and the user perspective.

The Techno-Anthropological triangle points to social responsibility as the core competency of a Techno-Anthropologist in the relation between experts and artifacts. Here the focus is on what responsibility the technological experts have when developing and distributing new technology. We here argue that the intra-actional expertise should be another connection between experts and artifacts, because of the blackboxed nature of the technology.

*"And so they're [developers] learning what they can do and they can't do. And so what we're finding is that folks tend to be more successful with Copilot, getting valuable output out of it, the more familiar they are with sort of the bounds or the guardrails of what LLMs can and cannot do. They sort of, prompt engineering, they sort of speak in LLMs. They sort of understand how to communicate to get more value.*"

(Informant 6, Appendix 6, p. 119)

The specific action of the model is blackboxed even for the technical experts, however the overall function of the model as well as the data used for training is known to the developers. This means that even though the developers have created the technology and know how it operates, there are still parts of the technology that are mystified to them. This causes developers to also have a learning curve in the same vein as the end-users. As the quote above points to is skills like prompting something the developers also have to learn and explore. The blackboxed nature and the rapid growth of generative AI increasingly puts developers into a user position, where they at the same time as end-users have to learn the expertise of the technology and get a sense of how to interact with it. We therefore see an opportunity for the intra-actional expert to also play a role in the communication and understanding between technical experts and the technological artifacts.

Since new technologies emerge as complex phenomena with a range of entangled actions within intra-actions, the view on the Techno-Anthropological role in technology development should also be altered to a more entangled profile of competencies. We argue that Techno-Anthropological methodological and theoretical tools should be used more dynamically between the actors of experts, users and artifacts because the borders of these entities become less clear with the introduction of communicative generative AI, in the sense, that the technology itself has a certain agency and a perceived expertise which makes the divide between user and expert more fluid. Exactly this entanglement between the three actors is a relation that Feenberg also points to:

*"Technical experts play a role in any technical change, but they are not alone. Users and victims judge the technical systems in which they are involved on the basis of their experience*

*and their participant interests. These are "oriented" reflective judgments, shared both with the general public through universalizing procedures such as warnings and rights claims, and with experts who can translate them into technically rational language and design"* (Feenberg 2017, p.167)

He describes how it is not only technical experts that are forming the technology, but it is also other actors around in interaction with the technology itself. The practical example of air pollution is used to illustrate how non-experts' experiences and perceptions can point to reflections and judgements that must be taken into consideration when working on and releasing new technologies:

*"For example, the engineers and automobile companies in Detroit were not the first to identify smog as a problem. It was the citizens of Los Angeles who initiated political demands for better automotive technology. Citizens arrived at this conclusion not through consulting manuals of engineering or medical textbooks, nor from deduction from principles, but from direct experience of the inconvenience and hazards associated with dirty air."* (Feenberg 2017, p.166)

It is on this basis that we argue that existing Techno-Anthropological competencies altered with the introduction of an intra-actional expertise are relevant and valuable for the development of generative AI technologies and tools. Techno-Anthropologists have the competences and expertise to bridge between users and experts as well as between the technological elements and relevant stakeholders through knowledge and acknowledgement of intra-actional processes that influence each of the entities and their intra- and interactions with each other.

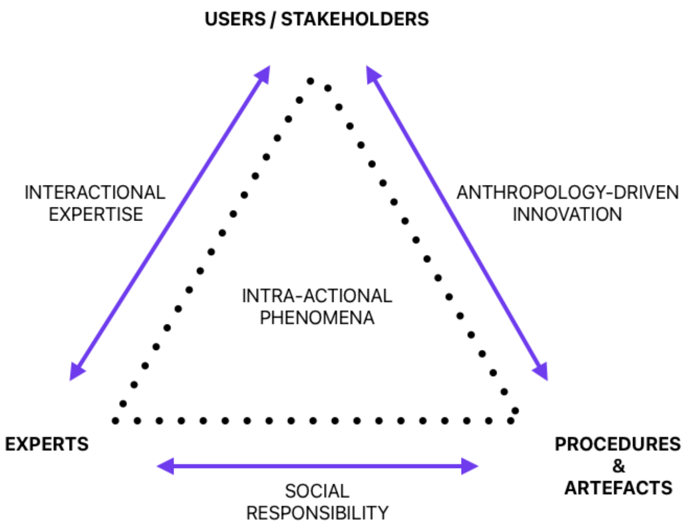


Figure 16: The intra-actional expertise included in the Techno-anthropological triangle (Feldes & Nielsen 2024)

The altered version of the TAN triangle takes into account the competency needed to interact with technologies that have increasing agency of their own. It is no longer purely about the practical knowledge and translation of instructions between user, expert, and artifact but rather the techno-anthropologist is necessary to facilitate the competencies to engage with the different entities.

## 8.4 Intervention

When entering a thesis collaboration with an external organization, the aim is not only to gain insights that add to research within the academic discipline and assist to be on top of developments within the Techno-Anthropological domain, but also to be able to derive some sort of product or tangible outcome that the organization can take further to use for internal purposes, such as inspiration, evaluation, quality assurance and further developments of their products or services. We therefore want to make it clear to AI developers like Microsoft how a more inclusive process and technology can be created in the field of Generative AI.
Giving the power to influence on the actual development of the tools only to those with technical access to the technology is exclusive to those who wish to contribute and can offer personal accounts as end-users that have rich potentials for the company as well. By trying to develop inclusive and safe technologies secluded and hidden in a blackbox to the public eye,

Microsoft can run the risk of producing involuntary exclusion of readily available user knowledge, experience, and power that the company can profit from.

Through analysis of intra-actions we can understand the nuances in mediation that are essential for the inclusive development of generative AI. Including users more directly and earlier on in the development processes does therefore not only yield potential for organizations to reach their inclusivity and responsibility goals more efficiently and close to the actual lifeworlds of their users, but also assist them in integrating these user needs and experiences into their products in line with the development processes, speeding up the development process which ultimately can facilitate faster release on the market. While iterative User Experience already does so to a certain degree, through research cycles based on secondary collected feedback and partial involvement of users in test-runs of technologies, the more direct involvement through participatory efforts using co-design artifacts can be a valuable addition to the qualitative research in the department.

An expanded version of the Techno-Anthropological idea of interactional expertise, where users, developers and technological artifacts are included on an equal playing field, can be valuable for companies like Microsoft to create a more inclusive development phase and technology.

Including these three actors in both the development phase and evaluation phase could help to reach the goals that Microsoft set for themselves in their responsible AI act in  a valuable and efficient way:

> *"We believe that industry, academia, civil society, and government need to collaborate to advance the state-of-the-art and learn from one another. Together, we need to answer open research questions, close measurement gaps, and design new practices, patterns, resources, and tools."*
>
> *(Microsoft 2022, p.3)*

The intra-actions between humans and these kinds of technologies are complex and include a range of different actions which requires a range of different actors to understand.

We argue that these problems of missing conversation, interaction and involuntary exclusion of potentially valuable user capacities and knowledge call for means to bridge the gap that exists and widens between users and developers. One way of doing so is the integration of professionals by means of consultation from the exterior of the company which the Techno-Anthropologist is an obvious choice for. However, we further argue that there must be considerations and evaluations of the Techno-Anthropological discipline to do so in the most holistic and valuable way.

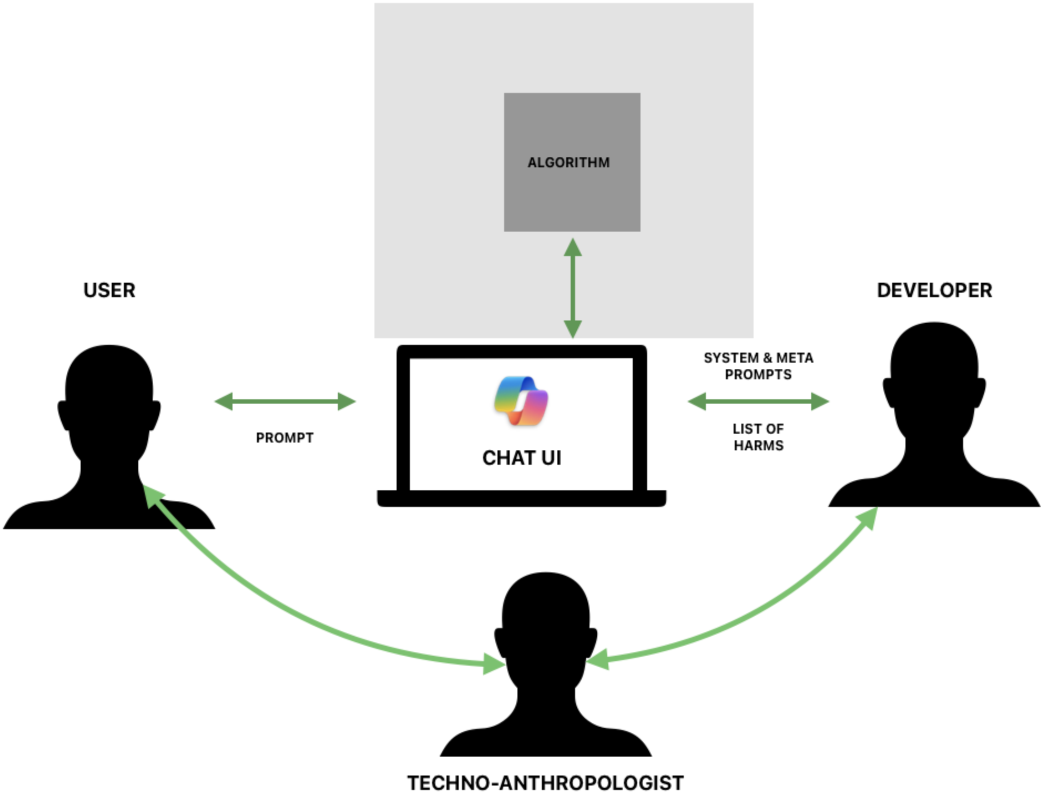


Figure 17: Demystifying the chatbot blackbox (Feldes & Nielsen 2024)

# 9. Limitations and Future Research

As indicated in previous chapters, the analysis and recommendations for intervention are highly individual and these specific actions and intra-actions are only applicable to this case of Microsoft Copilot. While the processes of intra- and inter-actions, as well as mechanisms to influence them as analyzed above also hold for other technologies, any other chatbot interface or software that is based on generative AI will come a different output and variations as the phenomenon and all its facets in terms of user experience emerge from the intra-actions that are taking place. That means future research must be done on different cases based on similar technologies and compare insights across technologies in order to derive universal regulations that are inclusive and closer to users' lifeworlds than they are at this point. The theoretical framework of technologies as phenomena of intra-actions and the methodologies of a techno-anthropologist can elucidate these technologies in a similar way to the Copilot case. In addition to this, the integration of the future version of the LLM will, and in its current version already does, appear differently in the user interface when accessed as the chatbot in the browser function compared to the integrated prompt interface in Business Central applications.

On the 13th of May 2024 OpenAI came out with a press release announcing the new version of the GPT model, the GPT-4o (OpenAI 2024). This latest version of their chatbot has the ability to derive its information not only from text but also audio and computational vision. This model will also be compatible with the Microsoft Copilot AI.

Compared to its preceding version, the model is faster and more accurate and has the ability to recognize voices and elements on a computer screen. Furthermore, the verbal part of the model is made more 'humanlike' in the sense that it has the ability to imitate both singing and laughter (OpenAI 2024).

With these improvements and new features to the technology, the intra-actions that have been analyzed in the present inquiry will change as well. Intra-actions will transform and expand, not necessarily in the sense that the inra-actions are completely different than described above but rather in terms of complexity and amount of and types of actions within each intra-action., which can have further impact on the agency, perception and use cases that users make as a result of conversing with the chatbot.

The timing of the launch of this newest version, only few months after the release of the GPT4 version, the integration in the existing Microsoft Copilot chatbot and the rapid speed of further development call for further investigation of the intra-actions and the effects that these developments have on users, their behavior, competences, experiences and attitudes towards Copilot and similar technologies.

It lies beyond the scope of the present paper to go into depth with these recent technological iterations and their impact on the human but should be subject of inquiry to following research efforts within the field.

This does not only make a case for further research in the academic field of Techno-Anthropology but also points to the constant and increasing need to involve Techno-Anthropologists as well as users themselves in research and development processes, making out a vast field of employment for professionals that will emerge further in the close future.

# 10. Conclusion

We began the present inquiry asking the question:

**How can considerations on user inclusion and blackboxing in development of generative Artificial Intelligence in the case of Microsoft Copilot be addressed and demystified through the concept of intra-action and why does Techno-Anthropological inquiry and problem solution support user inclusion?**

In order to derive an answer, we designed an exploratory research based on ethnographic as well as literature review methods.
Through qualitative interviews, digital auto-ethnography, participant observation, and a quantitative survey we build a quali-quantitative study around the case of Microsoft Copilot chatbot with focus on the User Experience Research and development team at Microsoft in Copenhagen, Denmark.

The theoretical approach grounded on the theoretical concepts of intra-action by Karen Barad, technological mediation theory by Olya Kudina, political science by Andrew Feenberg, and feminist Human-Computer-Interaction by Ann Light.
By holding up the empirical data of the Copilot case against the concept of intra-action, we identified the need to understand and acknowledge the existence of intra- as well as inter-actional processes between different stakeholders involved in use and development of generative artificial intelligence and chatbots based on these Large Language Model algorithms.
We found that the entanglements of human and chatbot are iterative phenomena that involve a plethora of intra-actions in of themselves and entail various conversational processes, agency and power structures on different levels.
We particularly identify an intra-action between developers and Copilot where elements like system-prompts and EU regulations influence how developers act with and on the technology. The second type of intra-action is the relation of user and Copilot, where users' technical competencies and understanding of the technology influence their interaction with Copilot. As part of these intra-actions, we observe varying levels of power and agency in relation to the technology, which are influenced by the language used in the interaction, technical literacy, competencies, and access to the knowledge of functions of a chatbot, which further makes

blackboxing, language and ways of conversation a crucial part of receiving valuable output when working with Copilot.

We discussed and accounted for the key to demystification of emerging AI technologies to be the awareness of the differences between technical and socially established blackboxes for different stakeholders which impact each their intra- and inter-action and thus use, perception, and opinion towards such technologies.

We have argued that increased inclusivity in the form of user-involvement in development processes and knowledge exchange between individual human actors can support demystification and inclusion, as well as further development of generative AI based chatbots. We have further demonstrated and argued that Techno-Anthropological competencies play a large role in these efforts and should be adapted to the influences of emerging technologies, because sharing of expertises not only relates to human actors but also non-human actors. Acknowledging these insights by including them in existing Techno-Anthropological theory and method supports bridging the gap between user and developer in design, development, and user experience research of Copilot and similar technologies. The involvement of the Techno-Anthropologist as an intra- and inter-actional expert can further add to the theoretical and academic domain that investigates state of the art technologies and how they can be understood and incorporated in large organizations such as Microsoft.

All in all, we come to the conclusion that demystification and inclusion can be addressed through the acknowledgement of intra-actions and their effect on humans and technologies as shaping mechanisms for use and development of AI technologies today and in the future. The Techno-Anthropological discipline is needed to facilitate these processes because it has the skills and competencies, when adjusted to new technological standards in regard to agency of humans and artifacts, to accomplish those.

# 11. References

### Figures


Figure 1: Verbeek, P. P. (2008). Cyborg intentionality: rethinking the phenomenology of human- technology relations. *Phenomenology and the cognitive sciences*, *7*(3), 387-395.

Figure 2: Kudina, O. (2021). "Alexa, who am I?": voice assistants and hermeneutic lemniscate as the technologically mediated sense-making. *Human Studies*, *44*(2), 233-253.

Figure 3: Botin, L. (2013). Chapter 2: Techno-Anthropology: Betweeness and hybridization. In Børsen, T. & Botin, L. *What is Techno-Anthropology?* pp. 35-66. Aalborg University Press.

Figure 4: Feldes, S. & Nielsen, T. D. (2024). Opening the Blackbox of Generative AI with Microsoft Copilot: A Techno-Anthropological study on demystification of relations & processes between humans and chatbots.

Figure 5: Feldes, S. & Nielsen, T. D. (2024). Opening the Blackbox of Generative AI with Microsoft Copilot: A Techno-Anthropological study on demystification of relations & processes between humans and chatbots.

Figure 6: Feldes, S. & Nielsen, T. D. (2024). Opening the Blackbox of Generative AI with Microsoft Copilot: A Techno-Anthropological study on demystification of relations & processes between humans and chatbots.

Figure 7: Feldes, S. & Nielsen, T. D. (2024). Opening the Blackbox of Generative AI with Microsoft Copilot: A Techno-Anthropological study on demystification of relations & processes between humans and chatbots.

Figure 8: Appendix 14: Survey Answers, p. 181

Figure 9: Appendix 14: Survey Answers, p. 181

Figure 10: Appendix 14: Survey Answers, p. 180

Figure 11: Appendix 14: Survey Answers, p. 184

Figure 12: Feldes, S. & Nielsen, T. D. (2024). Opening the Blackbox of Generative AI with Microsoft Copilot: A Techno-Anthropological study on demystification of relations & processes between humans and chatbots.

Figure 13:  Feldes, S. & Nielsen, T. D. (2024). Opening the Blackbox of Generative AI with Microsoft Copilot: A Techno-Anthropological study on demystification of relations & processes between humans and chatbots.

Figure 14:  Feldes, S. & Nielsen, T. D. (2024). Opening the Blackbox of Generative AI with Microsoft Copilot: A Techno-Anthropological study on demystification of relations & processes between humans and chatbots.

Figure 15: Kanstrup, A. M., & Bertelsen, P. (2013). Chapter 16: Participatory Reflections: Power and Learning in User Participation. In T. Børsen, & L. Botin (Eds.), *What is Techno-Anthropology?* (pp. 405-430). Aalborg Universitetsforlag. http://forlag.aau.dk/Shop/medier-og-informatik/what-is-techno-anthropology.aspx

Figure 16:  Feldes, S. & Nielsen, T. D. (2024). Opening the Blackbox of Generative AI with Microsoft Copilot: A Techno-Anthropological study on demystification of relations & processes between humans and chatbots.

Figure 17:  Feldes, S. & Nielsen, T. D. (2024). Opening the Blackbox of Generative AI with Microsoft Copilot: A Techno-Anthropological study on demystification of relations & processes between humans and chatbots.

## Illustrations

Illustration 1: Microsoft. (2024). Copilot. https://copilot.microsoft.com/ Accessed May 2024.

Illustration 2: Microsoft. (2024). Copilot. https://copilot.microsoft.com/ Accessed May 2024.

Illustration 3: Teknologirådet. (2024). Titan - Kunstig intelligens til bekæmpelse af desinformation.  https://tekno.dk/project/titan/ . Assessed May 2024.

Illustration 4: Microsoft. (2024). Copilot. https://copilot.microsoft.com/ Accessed May 2024.

## Literature


1. Abburi, H., Suesserman, M., Pudota, N., Veeramani, B., Bowen, E., & Bhattacharya, S. (2023). Generative ai text classification using ensemble llm approaches. *arXiv preprint arXiv:2309.07755*.

2. Adamopoulou, E., & Moussiades, L. (2020). Chatbots: History, technology, and applications. *Machine Learning with applications*, *2*, 100006.

3. Aghdam, A., Birungi, C., Duncan, D., Ghosh, P. K., Kalra, R., Mareels, I., Marimuthu, R., & Pasik-Duncan, B. (2022). Diversity & Inclusion in Universal Access to Technology - A Perspective. *IFAC-PapersOnLine*, *55*(39), 123-128. https://doi.org/10.1016/j.ifacol.2022.12.022

4. Alkaissi H, McFarlane S. I. (2023). Artificial Hallucinations in ChatGPT: Implications in Scientific Writing. Cureus. doi: 10.7759/cureus.35179. PMID: 36811129; PMCID: PMC9939079.

5. Anthropic. (2024). Write with Claude. https://claude.ai/login?returnTo=%2F%3F. Accessed May 2024.

6. Bandi, A., Adapa, P. V. S. R, Kuchi, Y. E. V. P. K. (2023). The Power of Generative AI: A Review of Requirements, Models, Input–Output Formats, Evaluation Metrics, and Challenges. *Future Internet*.; 15(8):260. https://doi.org/10.3390/fi15080260

7. Barad, K. (2007). Chapter 3: Niels Bohr’s Philosophy-Physics: Quantum Physics and the Nature of Knowledge and Reality. In *Meeting the Universe halfway: Quantum Physics and the Entanglement of Matter and Meaning*. London, Duke University Press.

8. Bellini, R., Meissner, J., Finnigan, S. M., & Strohmayer, A. (2022). Feminist human–computer interaction: Struggles for past, contemporary and futuristic feminist theories in digital innovation. Feminist Theory, 23(2), 143-149. https://doi.org/10.1177/14647001221082291

9. Birkbak, A & Munk, A. K. (2017). Kapitel 6: Digitale metoder i sammenhæng. In Birkbak, A & Munk, A. K. Digitale metoder. pp.167-190. Hans Reitzels Forlag.

10. Birkholm, K. (2018). Algoritmens autoritet - en etisk og politisk udfordring. In Birkholm K. & Frølich N. *De skjulte algoritmer - Teknoantropologiske perspektiver.* 1. Edition. Djøf Forlag.

11. Botin, L. (2013). Chapter 2: Techno-Anthropology: Betweeness and hybridization. In Børsen, T. & Botin, L. *What is Techno-Anthropology?* pp. 35-66. Aalborg University Press.

12. Børsen, T. (2013). Chapter 1: Identifying Interdisciplinary Core Competencies in Tech-Anthropology. In Børsen, T. & Botin, L. *What is Techno-Anthropology?* pp. 35-66. Aalborg University Press.

13. Casheekar, A., Lahiri, A., Rath, K., Prabhakar, K. S., & Srinivasan, K. (2024). A contemporary review on chatbots, AI-powered virtual conversational agents, ChatGPT: Applications, open challenges and future research directions. *Computer Science Review.*, *52*. https://doi.org/10.1016%2fj.cosrev.2024.100632

14. Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., ... & Xie, X. (2023). A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology*.

15. Chiu, T. K. (2023). The impact of Generative AI (GenAI) on practices, policies and research direction in education: A case of ChatGPT and Midjourney. *Interactive Learning Environments*, 1-17.

16. Egholm, L. (2014). Philosophy of Science: Perspectives on Organisations and Society. 1st Edition. Hans Reitzels Forlag.

17. Enslin, P., & Hedge, N. (2010). Inclusion and diversity. *The Sage handbook of philosophy of education*, 385-400.

18. Epstein, Z. Hertzmann, A. (2023). Art and the science of generative AI. *Science* 380. 1110-111.DOI:10.1126/science.adh4451

19. European Parliament. (2023). EU AI Act: first regulation on artificial intelligence. https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence Accesed: 29/04-2024.

20. Feenberg, A. (2018). A. Feenberg & B. Stiegler, "Technology at the End of the World". Critical Theory Workshop. Accessed March 2024, via: https://www.youtube.com/watch?v=XvSzA10SENg&t=1177s

21. Gan, E. (2021). 8. Diagrams: Making Multispecies Temporalities Visible. In A. Ballestero & B. Winthereik (Ed.), *Experimenting with Ethnography: A Companion to Analysis* (pp. 106-120). New York, USA: Duke University Press. https://doi.org/10.1515/9781478091691-011

22. Gichoya JW, Thomas K, Celi LA, Safdar N, Banerjee I, Banja JD, et al. (2023). AI pitfalls and what not to do: mitigating bias in AI. Br J Radiol. 10.1259/bjr.20230023.

23. Gil de Zúñiga, H., Goyanes, M., & Durotoye, T. (2024). A scholarly definition of artificial intelligence (AI): advancing AI as a conceptual framework in communication research. *Political communication*, *41*(2), 317-334.

24. Good Tape. (2024). Good Tape. https://goodtape.io Accessed: May 2024. Zetland.

25. Google. (2024). Gemini: Boost din kreativitet. https://gemini.google.com/. Accessed May 2024.

26. Gupta, M., Akiri C., Aryal, K., Parker E. & L. Praharaj. (2023). From ChatGPT to ThreatGPT: Impact of Generative AI in Cybersecurity and Privacy. In *IEEE Access*, vol. 11, pp. 80218-80245. (2023). doi: 10.1109/ACCESS.2023.3300381.

27. Guzman, A.L. & Lewis, S.C. (2020). Artificial intelligence and communication: A Human-Machine Communication research agenda. New Media & Society, 22 (1). Sage Publications.

28. Haraway, D. (1991). A cyborg manifesto: science, technology, and socialist-feminism in the late twentieth century. In D. Haraway (Ed.), *Simians, cyborgs and women: the reinvention of nature* (pp. 149–181). New York: Routledge.

29. Jensen, T. (2024). Sprogmodeller for Dummies: En intuitiv introduktion til teknologien bag ChatGPT. Københavns Professionshøjskole.

30. Jeong, C. (2023). A Study on the Implementation of Generative AI Services Using an Enterprise Data-Based LLM Application Architecture. *arXiv preprint arXiv:2309.01105*.

31. Juleskjær, M. & N. Schwennesen. (2012). Intra-active Entanglements - An Interview with Karen Barad. Kvinder, Køn & Forskning. (1-2).

32. Kanstrup, A. M., & Bertelsen, P. (2013). Chapter 16: Participatory Reflections: Power and Learning in User Participation. In T. Børsen, & L. Botin (Eds.), *What is Techno-Anthropology?* (pp. 405-430). Aalborg Universitetsforlag. http://forlag.aau.dk/Shop/medier-og-informatik/what-is-techno-anthropology.aspx

33. Kudina, O. (2021). "Alexa, who am I?": voice assistants and hermeneutic lemniscate as the technologically mediated sense-making. *Human Studies*, *44* (2), 233-253.

34. Kuutti, K., & Bannon, L. (2014). The turn to practice in HCI: towards a research agenda. *Conference on Human Factors in Computing Systems - Proceedings*, 3543–3552. https://doi.org/10.1145/2556288.2557111

35. Latour, B. (1993). We have never been modern (C. Porter, Trans.). pp. 13-30. Harvard University Press.

36. Light, A. (2011). HCI as heterodoxy: Technologies of identity and queering of interaction with computers. Interacting with Computers (23). Elsevier.

37. Mauro, G & Schellmann, H. (2023). 'There is no standard': investigation finds AI algorithms objectify women's bodies. In the Guardian. https://www.theguardian.com/technology/2023/feb/08/biased-ai-algorithms-racy-women-bodies Accessed: June 2024.

38. Microsoft. (2024). Copilot. https://copilot.microsoft.com/ Accessed May 2024.

39. Microsoft. (2022). Microsoft Responsible AI standard, V2. General Requirements. Accessed 2024. https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE5cmFl?culture=en-us&country=us

40. Microsoft. (2023). Microsoft 365 Copilot: The art and science of prompting. https://adoption.microsoft.com/files/copilot/Prompt-ingredients-one-pager.pdf Accessed April 2024.

41. Miller, D. (2018). "Digital Anthropology". In The Open Encyclopedia of Anthropology, edited by Felix Stein. Facsimile of the first edition in The Cambridge Encyclopedia of Anthropology. Online: http://doi.org/10.29164/18digital

42. Munz, E. (2017). Ethnographic interview. In *The SAGE Encyclopedia of Communication Research Methods*(Vol. 4, pp. 455-457). SAGE Publications, Inc. https://doi.org/10.4135/9781483381411

43. Naqa, I., Murphy, M.J. (2015). What Is Machine Learning?. In: El Naqa, I., Li, R., Murphy, M. (eds) Machine Learning in Radiation Oncology. Springer, Cham.

44. Ninness, C., & Ninness, S. K. (2020). Emergent virtual analytics: Artificial intelligence and human-computer interactions. Behavior & Social Issues, 29 (1), 100–118. https://doi.org/10.1007/ s42822-020-00031-1

45. Noble, H. & Heale, R. (2019). *Triangulation in research, with examples*. Research made simple. Evid Based Nurs 22 (3).

46. Olesen, F. & Kroustrup, J. (2007). ANT - Beskrivelsen af heterogene aktør-netværk. In Jensen, C. B., Lauritsen, P., & Olesen, F. (Eds.). (2007). *Introduktion til STS: Science, Technology, Society*. Hans Reitzels Forlag.

47. Oudshoorn, N. & Pinch, T. (2003). Introduction: How Users and Non-Users Matter. 10.7551/mitpress/3592.003.0002. In Oudshoorn, N. & Pinch, T. (2003). *How users matter. The co-construction of Users and Technology.*

48. OpenAI. (2022). ChatGPT. https://chatgpt.com/auth/login Accessed May 2024.

49. OpenAI. (2024). Hello GPT-4o.https://openai.com/index/hello-gpt-4o/. Accessed May 2024.

50. Roselli, D., Matthews, J., and Talagala, N,. (2019). Managing Bias in AI. In Companion Proceedings of The 2019 World Wide Web Conference (WWW '19). Association for Computing Machinery, New York, NY, USA, 539–544. https://doi.org/10.1145/3308560.3317590

51. Sanders, E. B. N., & Stappers, P. J. (2008). Co-creation and the new landscapes of design. *CoDesign*, *4*(1), 5–18. https://doi.org/10.1080/15710880701875068

52. Schensul, J. J., & LeCompte, M. D. (2012). Chapter 6: In depth, open-ended Exploration Interviewing. In Schensul, J. J., & LeCompte, M. D. *Essential ethnographic methods: A mixed methods approach* (Vol. 3). Pp. 134-167. Rowman Altamira.

53. Schön, D. A. (1983). Chapter 3: Design as a Reflective Conversation with the Situation. In Schön, D. A. (1993) *The Reflective Practitioner: How Professionals Think in Action*. pp. 76-100 https://doi.org/10.4324/9781315237473

54. Seaver, N. (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. Big data & society, 4(2), 2053951717738104.

55. Teknologirådet. (2024). Titan - Kunstig intelligens til bekæmpelse af desinformation. https://tekno.dk/project/titan/ . Assessed May 2024.

56. Thompson, R., Knyazev, B., Ghalebi, E., Kim, J., & Taylor, G. W. (2022). On evaluation metrics for graph generative models. *arXiv preprint arXiv:2201.09871*.

57. Verbeek, P. P. (2008). Cyborg intentionality : rethinking the phenomenology of human- technology relations. *Phenomenology and the cognitive sciences*, *7* (3), 387-395.

58. Verbeek, P. P. (2011). Moralizing technology: Understanding and designing the morality of things. University of Chicago Press.

59. Williams, M., & Moser, T. (2019). The art of coding and thematic exploration in qualitative research. *International management review*, *15*(1), pp. 45-55.

60. Zhu, G., Sudarshan, V., Kow, J. F., & Ong, Y. S. (2024). Human-Generative AI Collaborative Problem Solving Who Leads and How Students Perceive the Interactions. *arXiv preprint arXiv:2405.13048*.

61. Ziewitz, M. (2016). Governing Algorithms: Myth, Mess, and Methods. *Science, Technology, & Human Values*, *41*(1), 3–16. https://doi.org/10.1177/0162243915608948

62. Zúñiga, H. G., Goyanes, M. & Durotoye, T. (2024). A Scholarly Definition of Artificial Intelligence (AI): Advancing AI as a Conceptual Framework in Communication Research, Political Communication, 41:2, 317-334.