AALBORG UNIVERSITY

STUDENT REPORT

BEYOND MISINFORMATION: EXPLORING THE POSITIVE AND ETHICAL POTENTIAL OF DEEPFAKE TECHNOLOGY

A Techno-Anthropology thesis

28/05-2025

Title:

Beyond misinformation: Exploring the positive and ethical potential of Deepfake technology

Theme:

Scientific Theme

Project Period:

TAN10 - Spring semester 2025

Project Group:

Group 14

Participant(s):

Duy Hoang le Ha (dha23@student.aau.dk)

Supervisor(s):

Mette Ebbesen (mettede@plan.aau.dk)

Copies: 1

Page Numbers: 28,5

Date of Completion: 28/05-2025

Abstract:

The thesis aims to explore the positive applications of Deepfake through ethical guidelines. However, the primary use of Deepfake has created a stigma around the term, which overshadows the potential opportunities the technology may offer. A combination of online ethnography and interviews, conducted within the framework of postphenomenology and ethical principles, suggests that fear and concerns exist among social media users. It is important to regulate and restrict the use of Deepfake technology to ensure the safety of social media users. The findings emphasize a deeper insight into the relationship between Deepfake technology and human-world interactions—insight that could improve technical mediation and thereby enable future implementations of Deepfake technology across various fields.

Contents

Introduction

Technology made its first appearance in the 17th century and was used as a term to describe discussions related to applied arts only. The applied arts themselves later became the object of designation. In the early 20th century, the term “technology” was not only associated with processes and ideas but also tools and machines (Buchanan, 2025). In the mid-century the definition of technology was no longer mere objects that existed independently of human senses, in other words a noumenon (Kant, 1894). Technology was instead described as activities or tools that individuals sought to manipulate or change their environments. The definition raised concerns amongst observers who highlighted the challenge of differentiating scientific inquiry from technology activity (Buchanan, 2025). The history of technology has shown how its meaning has gradually changed – which does leave us with the question: What does technology mean currently?

Technology can be considered the advancement of systematic techniques for creating and accomplishing tasks. Essentially, the techniques can be considered as methods of creating tools and the products of the tools. The capacity to create these tools is determined by humanlike species. To elaborate on why it only applies to humanlike species, we will use other species as examples:

- Spiders make web to catch their preys
- Paper wasps construct their nests with various of materials (Evans, 2023)
- Ant construct ant hills to use as their homes

While these traits are befitting of the current definition of technology, they are only a result of instinctive behaviour patterns thus not being able to adjust quickly to unexpected changes. Contrary to other species, human beings do not possess the same level of instinctive reactions but do possess the capacity to think creatively and systematically about techniques. Human beings can therefore the environment innovatively in a way that other species cannot. A monkey will occasionally use a stick to retrieve a banana from a tree, while on the other hand a human would turn the stick into a cutting tool to retrieve multiple bananas. Human beings have been technologists since ancient times due to the inherent nature of being toolmakers. The history of technology therefore encapsulates the evolution of humanity (Buchanan, 2025).

The technological progress, which shaped the perspectives on the nature and impact of technology, since the Renaissance, is based on six assumptions. First and foremost, artifacts undergoing changes will face a marked improvement due to technological innovations; secondly, advancement in technology contributes directly to improvements of our material, cultural and social lives, thus rapidly improving the society; third, progression in technology can be measured through quantitative measurements, such as power, speed, efficiency etc.; fourth, the influence of technological changes are under absolute human control; fifth, technology has conquered nature adapted it to achieve human goals; sixth, technology and the society has reached their highest forms in the Western industrialized nations. The proponents of the technological progress initially viewed the assumptions as the primary goal of technological progress; however, they found it increasingly difficult to present the improvement of human life or control of nature as the sole goals of technological advancement. These assumptions were based on traditional views and oversimplifies the progress of technology advancement. New inventions do not appear out of nowhere; they are built on previous inventions. For example, the airplane was not an entirely new concept – it evolved from innovations in engine technology and earlier glider experiments. Furthermore, technology does not progress solely because of necessity. Some innovations emerge from experiments or unexpected uses (Basalla, 1988).

Technological progress is encapsulated by social involvement. In a similar fashion to biological evolution, not all technological innovations survive. Some are accepted, while others are abandoned due to social, cultural or economic factors (Basalla, 1988). The relationship between society and technological progress involves three points: social resources, social needs and a supportive societal mindset. If a technological innovation fails to meet any of the three points, it is less likely to become successfully or adopted. People will not devote resources to an innovation if it lacks the sense of social need. A society must be rich with suitable resources to maintain technological innovation. A supportive social mindset implies an environment that embraces new ideas. These points highlight the importances technological progress and its relations to social factors (Buchanan, 2025).

The rapid growth of technology progression and its achievements in our society has led to increasing concerns in many sectors. Modern technology has become a dominant influence in our society which has resulted in a technological dilemma. The dilemma consists of two problems; first, technology is being overly used and is therefore considered a threat to the quality of life and can endanger our society in itself; second, excessive reliance on technology in advanced industrial countries (Buchanan, 2025). These problems are relevant in our current state of society, because of the popularization of artificial intelligence (AI) - specifically generative AI. For example, ChatGPT – a conversational chatbot created by OpenAI. ChatGPT uses natural language processing to generate responses to user inputs akin to human beings. The AI application has gained worldwide popularity and has proven to be capable of earning a university degree. Despite the success that ChatGPT has gained, it also raises concerns and challenges within the educational sector. Although the intended use of the AI application was designed to provide with information quickly, students also used it for cheating, because ChatGPT could be used to complete written assignments. The students would therefore not learn as effectively because they were overly dependent on ChatGPT (Lo, 2023).

Another example of AI is Deepfake technology. The technology consists of advanced generative artificial intelligence to create realistic synthetic media. The category Deepfake technologies refers to the manipulation of existing content or the creation of entirely new media. It is therefore possible to alter facial features, swap faces or synthesize scenarios where individuals appear to do or say something that they have not actually done (Babaei et al., 2025). As Deepfakes become more sophisticated, it becomes increasingly difficult to distinguish between authentic and manipulated media. Additionally, traditional detection methods struggles to keep up with the advancement of Deepfake technology (Shoaib et al., 2023). The proliferation of Deepfakes has led to ethical dilemmas and threatens societal trust (Babaei et al., 2025).

Although a technological stigma has occurred around AI and Deepfakes, it is also necessary important to acknowledge the positive impact and the role AI has in our society. This study emphasizes the importance of AI, particularly Deepfakes and its positive applications.

Problem analysis

Deepfake technology derived from advancement in computational techniques and AI. In 2014 to 2016 breakthroughs in neural networks and deep learning, with the purpose of enhancing the ability to edit videos and images, became the foundation of more sophisticated manipulations. The advancements gradually shifted the purpose of traditional photo editing to automated AI-driven methods (Babaei et al., 2025). Deepfake technology is a generative deep learning algorithm capable of creating or modifying facial features to such an extent that it becomes difficult to distinguish between fake and real ones. (Malik et al., 2022). The earlier stages of Deepfake technology primarily consisted of autoencoders, which consisted of compressing input data through an encoder and a decoder to reconstruct the data. These methods often included two autoencoders, one for the original source face and one for the targeted face. They shared a common encoder to coordinate between the face swapping. Although these methods presented the possibility to manipulate faces, they unfortunately had an artificial appearance, due to having smooth textures and lacking fine details (Babaei et al., 2025).

Generative adversarial network (GAN) was introduced in 2014, which revolutionized artificial media generation. GANs consist of two neural networks, a generator and a discriminator. The generator generates plausible data, and the discriminator evaluates the authenticity of the generator's data. The generator and discriminator are trained against each other. The generator aims to produce realistic data while the discriminator's objective is to identify fake data. Through continuous training, the generator becomes better at producing realistic data, making it difficult for the discriminator to distinguish fake from real data. The generator's main objective is to produce so realistic data that the discriminator is unable to differentiate it from real data (Google, 2025). Deepfake technology based on GANs modules, especially StyleGAN and StarGAN have enabled the possibility to produce high-resolution, photorealistic images. Furthermore, the integration of synthetic audio with visual content has enabled the production of highly convincing audio-visual Deepfakes. Despite improvements in image and audio quality, GAN-based Deepfakes can still be distinguished from real images and videos due to flaws such as inconsistencies in blinking, lighting variations, and mismatched lip movements, which detection systems can recognize. (Babaei et al., 2025).

The current state of Deepfake technology is based on diffusion models. Diffusion models can transform simple noises, such as Gaussian noise (signal noise) (Yadav, 2025), into complex data, replicates real-life videos or images. In contrast to GAN-based Deepfakes, diffusion models are capable of generating state-of-the-art results, albeit at the cost of higher GPU (graphics processing unit) utilization. Current advancements in optimization techniques and hardware efficiency aim to address the challenges of diffusion models-based Deepfakes (Babaei et al., 2025).

Given the prominence and widespread accessibility of Deepfake technology, its primary use tends to be for malicious purposes, such as spreading misinformation, revenge porn, disrupt government functioning and to tarnish the reputation of high-profile figures (Masood et al., 2023). The term “Deepfake” gained popularity in 2017 when face-swapping of celebrities appeared on Reddit. This incident became the cornerstone of creating misleading content through Deepfake technology. In 2018, a video of former U.S. President Barack Obama surfaced on the internet, demonstrating the capability of Deepfake technology to create realistic content. Moreover, user-friendly software such as DeepFaceLab and FakeApp were released, presenting an opportunity for users with minimal technology knowledge to create Deepfakes (Juefei-Xu et al., 2022). Even though the software allowed for creative applications, it also raised significant concerns for security and ethics, specifically in relation to privacy and malicious usage (Nguyen et al., 2022). Over the course of 2019 and 2020, the malicious use of Deepfake technology and its associated risks became more prominent, leading to a global response from organizations, governments, and the research community, support the development of detection tools designed to mitigate the harmful use of Deepfake technology (Jacobsen & and Simpson, 2024). Since 2021, the progression of Deepfake technology through innovative methods, such as diffusion models and vision transformers (Khormali & Yuan, 2022), has enabled the creation of highly realistic Deepfakes, thus pushing the boundaries of generative AI. This advancement emphasizes the creative potential of Deepfakes and the increasing challenge of differentiating authentic material from fabricated content (Cao & Gong, 2022).

Despite the stigma surrounding Deepfake technology, it also demonstrates a significant range of positive applications across multiple fields. For instance, in media and entertainment, the technology is used for realistic video dubbing. In education, Deepfakes enhances interactive learning by brining historical figures to life or allowing the students to explore different scenarios (Altuncu et al., 2024). In healthcare, Deepfake technology can create realistic scenarios for medical training, enabling students to practice procedures on hyper-realistic characters. Furthermore, the technology can generate personalized characters to aid the patients with therapeutic support (Mukta et al., 2023). In marketing, companies and brands utilize Deepfakes to generate personalized ads and products for the customers to visualize (Sun et al., 2021).

It is evident that positive applications of Deepfake technology exist; however, the term is unfortunately associated with harmful uses, which overshadows the potential benefits Deepfake technology may have. The development of guidelines of Deepfakes could potentially result in a healthier perception of them. The purpose of this thesis is to raise awareness about the potential of Deepfake technology, which leads to the following research question:

"How can ethical guidelines balance the benefits and risks of Deepfake technology to promote a healthier public perception?"

Theory

Don Ihde (14th of January 1934) was a philosopher of science and technology. He was the first North American pioneer within philosophy of technology (Stony Brook University, 2025). Don Ihde developed postphenomenology through the usage of methodological tools from classical phenomenology to analyse technology, and particularly how technologies mediate human experiences. He categorizes technologies as non-neutral mediators in the relationships between human beings and the world. Additionally, he presents the idea that technologies actively co-shape human perception and experience (FutureLearn, 2016).

Peter-Paul Verbeek (6th of December 1970) is a philosopher, who specializes in ethics of science and technology. Verbeek is well-known for his theory of technological mediation, which encapsulates how technologies affect human-world relations. Through his theory of technological mediation, that is based on the postphenomenological framework by Don Ihde, he states that morality is not exclusively a human characteristic but is co-shaped through the interactions between technologies and humans (ppverbeek, 2010).

Postphenomenology describes the work of a global and interdisciplinary collective of scholars who studies the relations between humans and technologies (Irwin & Rosenberger, 2014). Postphenomenology revolves around the interplay between technology as a cultural instrumentation and traditional culture, the implications of technoscience on multiculturalism and the application of postphenomenology through perceptual technologies and feminist philosophy of science. Technology as a cultural instrumentation is portrayed as a mediator which influences cultural practices and perceptions. Ihde explores the relations between technoscience and multiculturalism, proposing that technology is capable of unifying and divide cultural identities. The theoretical framework aims to explore how the advancement of technology shape human experiences and cultural narratives, while also highlights the importance to understand these relations in modern society. Furthermore Ihde suggests that our comprehension of reality is mediated through technologies, rather than considering experience as solely subjective or internal (Ihde, 1993). Postphenomenology provides a nuanced framework to study technology as an entity that is embedded in human-world relations. More specifically, Ihde established 4 different types of technology-human-world relations. First of all, embodiment relations, which describe how the focus is not on the technology itself, but on what it allows us to experience or see. Secondly, hermeneutic relations, where the technology represents elements of the world. Thirdly, alterity relations, where technology is portrayed as quasi-other - a term used to describe technology that appears as another entity (e.g. voice assistant or robots). Lastly, background relations, where the technology influences the environment silently (Ihde, 1993). Additionally, Verbeek established three different types of technology-human-world relations on the premise of Don Ihde's framework. First and of foremost, cyborg relations, which describe how technology merges with the human. Secondly, immersion relation, which refers to technology creating an interactive environment. Third and lastly, augmentation relations, where technology is portrayed as mediators and reshape the human-world experiences (Verbeek, 2015). Multistability is another key concept within postphenomenology, which refers to the multiple ways a single technology can function or be interpreted depending on context and user interaction (Ihde, 1993).

While postphenomenology offers a grounded approach to understand the complex interplay between technology and humans in modern society, the mediation can have significant political and ethical implications, thus affecting societal norms and human behaviour. It is therefore necessary to adapt a more socially situated understanding of postphenomenology, which would require a shift from an individualistic perspective to a collective understanding of technological mediation (Rosenberger et al., 2015). Additionally, a deeper understanding of technological mediation would help in critically evaluating the biases and values incorporated in technologies (Verbeek, 2016). As a result, the framework could potentially improve its approach on the socio-cultural and political contexts (Rosenberger et al., 2015). Technology mediation also raises concerns within political contexts. One example is social media interactions, which can reshape perceptions of universal issues (Verbeek, 2016), potentially leading to power imbalances and ethical dilemmas. It is essential to adopt a holistic and practical approach when evaluating how technology impact society - assessing the potential risks and benefits while also minimizing the harm during the design process of technologies (Robson & Tsou, 2023).

When positioning Deepfake technology within the framework of postphenomenology, it is important to consider how the technology reshape human-world relations. First and foremost, Multistability is a key concept that is applicable to Deepfake technology because it is used or interpreted in multiple ways depending on the context or user experiences (Ihde, 1993). As mentioned in the previous sections the reputation of Deepfake is mostly negative due to its malicious use to tarnish the reputations of well-known people and the political manipulation and can therefore be considered as an autonomous weapon. The positive applications implies that the technology can be considered as an aid, assistant or salesman within various sectors like education, healthcare, marketing and many more. Deepfake can also be considered as a chameleon or as quasi-other (another entity) – which fits the alterity relations (Verbeek, 2015). In regard to hermeneutic relations, cyborg and immersion relations, Deepfake can represents elements of the world, create interactive elements in various of fields and merge with humans through the generative AI models. Overall, the technology reflects on augmentations relations by reshaping humans-world experiences through its role as a mediator. Deepfake technology raises ethical concerns that the framework of postphenomenology alone cannot fully address. First of all, similar to other technologies, Deepfake allows for mass data collection through social media and facial recognition, which put users' privacy at risk. Secondly, when people with a following or power uses the technology for their own profit, the mass can be swayed which indicates a blind sense of trust. Lastly, because of the ability of being a quasi-other, Deepfake can impact work employment due to its efficiency (Robson & Tsou, 2023).

To address the ethical concerns of Deepfake technology, I use the 4 principles of Common Morality by Tom L. Beauchamp (2nd of December 1939), who was a prominent philosopher. He was best known for his influential work in bioethics and the philosophy ethics (Bioethics, 2025). Common morality suggests that a set of moral norms and principles exists and is shared by all people regardless of their background or culture. Common morality is well suited for Deepfake technology because it allows for pluralism in moral reasoning. This means that, although diverse societies hold different and sometimes conflicting moral beliefs, pluralism accepts multiple ways of understanding what is right and wrong. The conflict of morals can be resolved through shared ethical principles (Beauchamp, 2003). The 4 principles consist of respect for autonomy, beneficence, non-maleficence and justice (Burks, 2015). The principle of respect for autonomy considers autonomous people as capable of self-governance and to make their own decisions as long as it do no harm to others. Beneficence suggests that it is necessary to actively act for the benefit of others' wellbeing and to the good of society as a whole. Non-maleficence presents the idea of doing no harm to others intentionally, however some harm can be justified if it leads to greater good. Justice represents equality respects and fair distribution of benefits and burdens (Burks, 2015). The principles are prima facie, meaning they are binding as long as they are not in conflict with other prima facie principles. If the principles conflict, they must be balanced. For instance, beneficence could conflict with respect for autonomy (e.g. when a patient declines life-saving treatment) (Beauchamp, 2015). Deepfake technology conflicts with respect for autonomy, when the intentions are malicious, as previously mentioned when Deepfake is used for defamation, political manipulations or personal attacks it does cause harm to others. On the other hand, if Deepfake is used for educational, medical or as an assistant it aligns with respect for autonomy. If Deepfake is used as simulation training for medical students or to enhance customers' experiences, Deepfake aligns with the principle of beneficence. In relation to justice, Deepfake is in a grey zone. Deepfake is an open-source AI, which means everyone has access to it. The technology in itself does not require a high level of technical knowledge and is therefore available to everyone. However, in a different perspective, for the less developed countries, they unfortunately might not have access to the technology. Furthermore, the use of Deepfake does not align with the principles of justice and respect for autonomy, because it does not ask for consent, as long as the users provide with the necessary material of others. Deepfake technology also does not align with the principle of non-maleficence since it is not capable of doing unintended harm. The users have to intentionally use specific material for their intended purposes in order for the technology to produce content.

Methodology

ChatGPT[1] has been used throughout the entirety of the thesis in alignment of the official guidelines by Aalborg's University (AAU, 2023). In this specific case, ChatGPT has been used for grammatical purposes and stylistic edits, such as choice of words. The reasoning behind the choice was to ensure a pleasant reader experience with a proper and more coherent structure of the paper.

The empirical data was collected based on the stakeholders I deemed necessary for the project. Social media users and an informant within healthcare participated in the interviews. The informants signed a contract of consent. Additional data were collected through specific Instagram videos related to Deepfake – where the comments of the videos were used for the analysis. Furthermore, observations were made by exploring the webpage of the Danish Police Online Patrol, which led to their Discord[2] server. The server had text channels, providing with relevant information which will be elaborated on in the analysis.

In relation to the interviews, they were planned out in chronological order. The information the social media users provided with were necessary to create a foundation for future interviews with other stakeholders and work. All of the interviews were semi structured interviews with open ended questions through the lens of James Spradley's theory (Spradley, 1979). The overall framework of the interview guide was based on descriptive questions. They are open-ended questions that are designed to elicit the informants' experiences, cultural practices and activities. The types

[1] An AI chatbot developed by OpenAI. Uses prompts to generate human-like text (OpenAI, 2024)

[2] A free communication platform where people can communicate through calls, texts and videos (Discord, 2025)

of descriptive questions that were used for this specific thesis were grand tour questions, mini-tour questions, example questions and experience questions. The grand tour questions are broad questions that allows the informants to describe a general activity or event. The mini-tour questions are supplementary to the grand tour questions by introducing sub-questions related to the events or activities that the informants were describing. Example questions enable the informants to give concrete examples related to previously asked grand tour questions. Experience questions evoke the informants to share personal experience which often reveal values and norms in cultural behaviour (Spradley, 1979).

Interviews tend to be solemn due to the professional setting while also interacting with strangers. These scenarios may affect the informants to be very reserved – especially because they can be conscious their answers being recorded and used as data. Furthermore, it diminishes the chances of the informants talking as freely as possible thus not aligning with Spradley's concepts and an inductive research approach. Spradley advocates for developing rapport with the informants through a four-stage process: Apprehension, exploration, cooperation and participation. The interview guide has been designed to ensure that these four-stage processes are applied. The introduction, grand tour and some mini-tour questions were designed to address apprehension and exploration, in which would remove initial nervousness during the first meeting thus resulting in becoming more comfortable around each other. The example questions were tailored to align with cooperation, encouraging the informants to share in-depth information. The experience questions were designed based on participation, allowing the informants to become active participants and, at times, lead the conversation. Building rapport with the informants and using Spradley's concept of descriptive questions provided meaningful data to be used for the analysis. (Spradley, 1979).

The interviews were recorded and transcribed through Microsoft Teams[3]' built in transcription and recording feature with manual correction of the automatic transcription tool. The interviews that were conducted in Danish were translated to English, with efforts made to preserve the authenticity of the information provided by the informants. The process of translation will be further discussed in the discussion section.

[3] A communication platform developed by Microsoft (Microsoft, 2017)

Analysis

The analysis will be conducted through the lens of post-phenomenology and the principles of ethics within common morality. Excerpts from the interviews will be used to support the analysis. Additional data from the Instagram videos and Discord server of the Police Online Patrol (POP) will also be used in the analysis due to its relevance to the thesis.

The initial questions were intended to figure out where activity of Deepfake could potentially occur. While the informants mentioned well known social media platforms such as Facebook, Instagram, X (formerly known as Twitter) and TikTok, which are popular platforms were Deepfake occurs, there were also mentions of non-traditional platforms like Discord and YouTube. All of the informants had a general idea of what Deepfake technology was capable of. The SoMe (social media) informant 1 noted:

"I have seen deep fake media, I believe. I see them in Instagram videos and TikToks. For example, a video where there is a person in the background talking but the person doesn't really look real, and the voice is kind of robotic"

The SoMe informant 1 further explained:

"The ones I have seen so far I can differentiate it from real people because you can usually see it in the eyes, but for the untrained eyes I don't think it is noticeable."

Another example is from the SoMe informant 2:

"Deepfake is related to AI and imitates real life or fill in gaps in videos by adding new faces in pictures that were not there to begin with."

The informants were aware of the general use of Deepfake technology which aligns with Ihde's types of technology-human-world relations (Ihde, 1993) – specifically embodiment and alterity relations. Because how they described Deepfake technology, it implies that the focus is not on the technology in itself but how it is perceived. Furthermore, it intertwines with the ability to appear as another entity. Through the perception of the informants, Deepfake appears as another person or object in the videos they watched. Although Deepfake demonstrates creative usages, the informants had a negative perspective on the use of Deepfake based on their own experience, which reflects on the general stigma that surrounds the term Deepfake. The SoMe informant 1 noted, *"I have seen instances or read about it [Deepfake], where people got scammed."*

Another instance is from the healthcare informant:

> *"They are using people's faces on generated bodies without their consent."*

The SoMe informant 2 were well aware of the common use of Deepfake:

> *"Considering there was drama or a scandal with some influencers being used in Deepfake for explicit pictures or videos."*

While Deepfake can be categorized as a multistability technology, since it has multiple ways of being used, it is unfortunately not always perceived as positive. Unfortunately, with the malicious use of Deepfake as seen in the excerpts above – they have 1 thing in common. They violate 3 principles of ethics (Beauchamp, 2015). People are free to produce Deepfake content as long as it does no harm to others; however, in these scenarios, Deepfake is used with ulterior motives causing harm to others – which disregard respect for autonomy. Deepfake content also disregard non-maleficence, because the technology cannot unintentionally cause harm to others. Deepfake requires user inputs to produce content, meaning that users must deliberately prompt the technology to harm others. Additionally, if Deepfake is used for malicious purposes it undermines the principle of beneficence, since it does not promote the well-being of others. This is also stated by the SoMe informant 1:

> *"[Deepfake] can impersonate other people, and if you keep that in mind, one could assume for the worst - people could ruin other people's images."*

Despite the negative perspectives of Deepfake technology, it also offers positive applications in various fields. Deepfake technology aligns with immersion relations, because it can create an interactive environment as mentioned by the healthcare informant:

> *"In terms of my profession, something related to learning material or video material, where it shows different procedures and how to do them"*.

Creative uses of Deepfake were also mentioned – which goes a following by the SoMe informant 1:

“You have the option for anonymity. If you want to create media but don’t want to show your face or use your own voice.”

And the SoMe informant 2 suggests:

“I have heard of it being used to make messages from relatives who passed away or using past videos of them. But I am a bit concerned about the moral and ethical aspects, because how do you get these people’s consent?”.

While these applications align with the cyborg relations, since it merges humans with the technology, it does have ethical implications. If the users prompt Deepfake technology with the data of others without their consent, it would violate the principle of respect for autonomy. Additionally, the implementation of Deepfake must be carefully considered – especially in fields where personal sensitive information is used on daily basis, such as the healthcare department. The healthcare informant noted:

“It is difficult to imagine it being used for something positive without it easily becoming misused”.

Which was elaborated on:

“If you give it [Deepfake] too much data and it ends up in the wrong hands, you will get in serious trouble. I think it is too risky on many aspects.”

The excerpts present political and additional ethical concerns. First and foremost, with open access to Deepfake technology, it can easily become overly prominent, thus resulting in the risks and concerns. It is therefore necessary to actively regulate and restrict the use of the technology to ensure that is only available for its intended use. Secondly, the regulation and restrictions are hard to define, it requires several testing phases in order to gauge where the restrictions should exist. If the technology is too restrictive, it does not provide with creative solutions. On the other hand, if Deepfake has almost no restrictions in place, it may increase the chances of being used for malicious purposes. The solution could possibly be to design guidelines for the use of Deepfake technology through the lens of the principles of ethics. The framework should be based on principles being prima facie, meaning they are binding as long as they are not in conflict with other prima facie principles. The technology in its current state could align with the principles of beneficence in the healthcare department, if it is used for to the betterment of the healthcare employers and patients, however without any guidelines, the technology conflicts with respect for autonomy, because it could potentially harm the patients. Third and lastly, the process of designing the guidelines must involve the state and municipality. Although the individuals can take precautions against Deepfakes, it is necessary to highlight the pros and cons of Deepfakes. Creating guidelines may potentially improve technical mediation and the relations between the technology and human-world experiences, thus creating awareness of Deepfakes. This is also suggested by the SoMe informant 1:

"I think they [the state] should introduce some regulations. Build some frameworks."

This is also supported by the healthcare informant:

"It is necessary to have clear guidelines in relation to which face you use, the original owner of the picture you use and who uploads the content."

Additional ethical implications were identified in relation to social media experiences. Being conscious of Deepfake manipulation while using social media, creates fear and concerns for the users, since they could potentially become a victim. The SoMe informant 2 stated:

"I feel very insulted and somewhat violated. I do not have any control over it [Deepfakes]. There are no current legalisations as far as I am aware, that can protect me from it, since it is very new. Even the awareness is not as widely spread among people who don't use social media."

Which further validify the need for guidelines and restrictions to exist in order to combat the malicious use of Deepfake. It would be a solution to ensure that Deepfake technology does not conflict with the principles of respect for autonomy and non-maleficence. From the perspective of embodiment relations, Deepfake is perceived as unethical due to its continued use with malicious intentions. Furthermore, from the perspective of hermeneutic relations, Deepfake may be regarded as a technological weapon – which the healthcare informant 2 stated:

"I associate the technology with revenge porn or satirical portrays of famous people, where they are mocked."

When Deepfake technology is used as a weapon and the perpetrators face no legal actions, it becomes a societal problem. If people of status use the technology for their own profit it could potentially lead to power imbalances. In this case Deepfake technology contradicts with the principle of justice, since all person would not be treated with equally. Additionally, proving that Deepfake technology functions as a mediator and reshape the human-world experiences, which is known as augmentation relations. To address these concerns, I observed the Discord server of POP (Police Online Patrol) (Appendix 1). The server consists of text channels, including information regarding safety, awareness of cybercrimes and contacts on different social media platforms, which could spread more awareness surrounding Deepfake technology. By branching out to multiple platforms, POP gains a deeper understanding of where crimes could occur, resulting in potential frameworks for future guidelines that align with the principles of common morality.

Since Deepfake technology is continuously evolving, 2 Instagram videos were randomly encountered, but I consider them worthwhile to include in the analysis. They illustrate the potential utopian and dystopian scenarios of Deepfake technology. The first video represents the dystopian scenario, which includes Deepfakes of well-known celebrities that are so realistic, it is almost impossible to confirm the authenticity of the video (leswishtv, 2025). Comments from the video mostly indicate the fear among the viewers, since it could lead to harmful and unethical use (Appendix 2). The other video represents the utopian scenario and includes how Deepfake is used to visualize the children's dream profession (iamdomfarnan, 2025). The video is well received by the viewers, where many of them expresses the support, due to the educational purpose Deepfake is capable of (Appendix 3).

This further proves that that proper regulations and guidelines for Deepfake technology can facilitate its implementation across multiple fields. It requires awareness through a deeper understanding of the technology in a human-world relation, guided by postphenomenology and ethical principles.

Discussion

First and foremost, as mentioned in the methodology section, the interviews conducted in Danish had to be translated into English so that excerpts could be presented in the analysis. Despite efforts to translate with as little bias as possible, I must acknowledge that some of the essence or values of the data may have been lost in translation, as the process was done manually. Moreover, an inductive approach may have given informants too much freedom, leading to off-topic discussions. Furthermore, since Spradley's framework was used for the interview guides, it resulted in dynamic data—meaning that recreating this thesis with different informants might lead to different results.

Secondly, while empirical data of Deepfake technology does exist, it does not warrant a current perspective on Deepfake. Especially in the spirit of a techno-anthropologist, it is necessary to interact with stakeholders to gain a deeper understanding of potential implications of the technology. This is also why postphenomenology was used in combination with the principles of ethics. They complement each other and are relevant for greater insights of the relations between technology and human-world experiences. Unfortunately, the number of informants were not as many as I would have wanted, which could diminish the validity of the thesis. To make the thesis feel more complete, I would have liked to present the empirical data to other relevant stakeholders (e.g. the municipality, politicians, and the Police Online Patrol). Some of them were contacted, however despite the stigma surrounding Deepfake technology, it is still a loaded term for people, that do not use social media on a regular basis. The police department works with personal sensitive information, which is why it was difficult to answer the interview questions I originally designed. Other stakeholders simply did not reply. To ensure dialogue in the future, I could simplify the explanation of Deepfake technology and how it works.

Third and lastly, because Deepfake technology aligns with being a multistability technology, it becomes increasingly difficult to define what it actually does. However, the value and potential of being a multistability technology diminish if it is continuously used with harmful intentions, as people only associate the technology with something negative. Furthermore, as presented in the dystopian and utopian scenarios, the technology's role and its relations to human experiences is dynamic, depending on how Deepfake is used. If no regulations or restrictions are established, the use of Deepfake could potentially lead to a virtual warfare. If we practice the principle of justice—meaning that everyone should have access to Deepfake and use it as they wish—it could lead to societal issues. The other principles would also be in conflict, since people might use it to harm others. Although the utopian scenario presented a healthier perspective on Deepfake technology, it also raises concerns. Additional conflicts with the principle of justice may arise if the technology is more efficient than humans and serves as a cheaper resource, potentially leading to redundancy in various fields. In such scenarios, some individuals may gain an unfair advantage. The topic of redundancy was unfortunately not within the scope of the thesis but is important to highlight as well. It is therefore crucial to digest Deepfake technology gradually, while carefully considering all of the ethical and political implications that may be associated with the technology. It is also necessary to emphasize that Deepfake technology itself is not the perpetrator; the responsibility lies with the users.

Conclusion

The thesis aims to address how ethical guidelines can balance the benefits and risks of Deepfake technology to promote a healthier public perception. The analysis of the empirical data suggests that guidelines are necessary – especially since the prominent use of the technology is unethical. The lack of consent when using others' identities for Deepfake content is not considered as something serious, since no one is being held accountable. Even prominent figures use deepfake technology for personal gain and to defame others, yet they face no legal consequences, resulting in power imbalances. Furthermore, if the technology is not regulated or has restrictions, it makes social media an unsafe place for users, as they could become victims of Deepfakes. There need to be ethical guidelines to reduce the risks of deepfake technology if it is implemented across different sectors.

The findings highlight the importance of guidelines, as the current use of Deepfakes is often associated with harmful intentions. This is also reflected in the empirical data that was collected, as the informants' initial opinions regarding Deepfakes were negative. Deepfake is a multistability technology that reveals political and ethical implications which must be addressed. It is therefore important to create guidelines through the lens of postphenomenology that align with ethical principles. This is relevant not only for techno-anthropologists but also for other professions, as deeper insights into Deepfake enable improved technical mediation, thereby benefiting everyone. In a technology-driven world, it is not possible to remove Deepfake technology as a whole. Instead, we should learn to coexist with the technology through a healthier perception, since Deepfake only becomes what we perceive it to be.

Future perspectives

As an extension to the discussion section, it is important to include other stakeholders in considering how Deepfake technology should be regulated and restricted. Presenting the findings to the police department with carefully tailored questions, could hopefully result in insights related to how they track the perpetrators behind the harmful use of Deepfakes. On the other hand, it could also be interesting to observe what sort of impact Deepfake technology may cause, if it were to be implemented in the police department. Would it coexist within the environment, or would it render forensic artists redundant?

The findings could also be presented to the municipality to help raise awareness about Deepfake technology. Moreover, their input could also aid in the creation of necessary guidelines. If Deepfake technology were to be implemented in fields such as education, a co-design framework involving techno-anthropologists, the municipality, and the educational sector could foster a positive perception of Deepfake.

To address the fear of redundancy in various professions, it makes sense to explore fields that already employ positive applications of generative AI. The insights would be compared to the same fields that does not use generative AI. This is important to highlight, since some jobs, such as digital artists, have been replaced by generative AI (didoco, 2025).

The comments on the 2 Instagram videos provided with interesting insights into how Deepfake technology is perceived. Unfortunately, the API of Instagram makes it difficult to extract data from their webpage. More time would be required to research for methods to harvest data from Instagram. If harvesting data from Instagram is feasible, it would create opportunities to conduct online ethnography through the method of controversy mapping.

Acknowledgement

First and foremost, I would like to thank the social media users who participated in the initial interviews. Their opinions were used as a foundation for future interviews. Secondly, I would like to thank the healthcare employer, Elvira, since she provided with data, that was used for possibly applications of Deepfake. Last but not least, I would like to thank my supervisor, Mette Ebbesen, who has been a huge part of this thesis since the beginning. I could not have done it without you.

Bibliography

AAU. (2023). *Rules for the use of generative AI*. Aalborg University. https://www.students.aau.dk/rules/rules-for-the-use-of-generative-ai

Altuncu, E., Franqueira, V. N. L., & Li, S. (2024). Deepfake: Definitions, performance metrics and standards, datasets, and a meta-review. *Frontiers in Big Data*, *7*. https://doi.org/10.3389/fdata.2024.1400024

Babaei, R., Cheng, S., Duan, R., & Zhao, S. (2025). Generative Artificial Intelligence and the Evolving Challenge of Deepfake Detection: A Systematic Analysis. *Journal of Sensor and Actuator Networks*, *14*(1), Article 1. https://doi.org/10.3390/jsan14010017

Basalla, G. (1988). *The Evolution of Technology*. Cambridge University Press.

Beauchamp, T. L. (2003). A Defense of the Common Morality. *Johns Hopkins University Press*, *September 2003*(3). https://doi.org/10.1353/ken.2003.0019

Beauchamp, T. L. (2015). The Principles of Biomedical Ethics as Universal Principles. In *Islamic Perspectives on the Principles of Biomedical Ethics: Vol. Volume 1* (pp. 91–119). WORLD SCIENTIFIC (EUROPE)/IMPERIAL COLLEGE PRESS. https://doi.org/10.1142/9781786340481_0004

Bioethics. (2025). Tom Beauchamp, PhD. *Johns Hopkins Berman Institute of Bioethics*. https://bioethics.jhu.edu/people/profile/tom-beauchamp/

Buchanan, R. A. (2025, February 12). *History of technology | Evolution, Ages, & Facts | Britannica*. https://www.britannica.com/technology/history-of-technology

Burks, D. J. (2015). *Beauchamp and Childress The Four Principles*.

Cao, X., & Gong, N. Z. (2022). Understanding the Security of Deepfake Detection. In P. Gladyshev, S. Goel, J. James, G. Markowsky, & D. Johnson (Eds.), *Digital Forensics and Cyber Crime* (pp. 360–378). Springer International Publishing. https://doi.org/10.1007/978-3-031-06365-7_22

didoco (Director). (2025, April 27). *These Art Jobs Are DOOMED by AI* [Video recording]. https://www.youtube.com/watch?v=N21kxcYZI_U

Discord. (2025, May 13). *Discord—Group Chat That's All Fun & Games*. https://discord.com

Evans, L. (2023, April). *Top 3 Animals With the Best Construction Skills*. Construction & Facilities Management from Hawthorn Estates. https://www.hawthorn-estates.co.uk/animals-with-construction-skills/

FutureLearn. (2016). What can we learn from Don Ihde? *FutureLearn*. https://www.futurelearn.com/info/blog

Google. (2025, February 26). *Overview of GAN Structure | Machine Learning | Google for Developers*. Google.Com. https://developers.google.com/machine-learning/gan/gan_structure

iamdomfarnan. (2025, May 23). *Dom Farnan på Instagram: 'Okay but… this might be one of the coolest uses of AI I've seen! And I don't know for you, but I would have absolutely loved this as a kid 💯'*. Instagram. https://www.instagram.com/iamdomfarnan/reel/DJ_3kDsM1JI/

Ihde, D. (with Internet Archive). (1993). *Postphenomenology: Essays in the postmodern context*. Evanston, Ill. : Northwestern University Press. http://archive.org/details/postphenomenolog0000ihde

Irwin, S. O., & Rosenberger, R. (2014, October 27). Postphenomenology. *Postphenomenology*. https://postphenomenology.org/about-this-site/

Jacobsen, B. N., & and Simpson, J. (2024). The tensions of deepfakes. *Information, Communication & Society*, *27*(6), 1095–1109. https://doi.org/10.1080/1369118X.2023.2234980

Juefei-Xu, F., Wang, R., Huang, Y., Guo, Q., Ma, L., & Liu, Y. (2022). Countering Malicious DeepFakes: Survey, Battleground, and Horizon. *International Journal of Computer Vision*, *130*(7), 1678–1734. https://doi.org/10.1007/s11263-022-01606-8

Kant, I. (1894). *Dissertation on the Form and Principles of the Sensible and the Intelligible World*. Https://En.Wikisource.Org/. https://en.wikisource.org/wiki/Kant%27s_Inaugural_Dissertation_of_1770

Khormali, A., & Yuan, J.-S. (2022). DFDT: An End-to-End DeepFake Detection Framework Using Vision Transformer. *Applied Sciences*, *12*(6), Article 6. https://doi.org/10.3390/app12062953

leswishtv. (2025, May 2). *LeSwishTV på Instagram: 'C'est le début de la fin. 📷Via @r4f43lo (follow him for more info)'*. Instagram. https://www.instagram.com/leswishtv/reel/DJB9QKnoTOS/

Lo, C. K. (2023). What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. *Education Sciences*, *13*(4), Article 4. https://doi.org/10.3390/educsci13040410

Malik, A., Kuribayashi, M., Abdullahi, S. M., & Khan, A. N. (2022). DeepFake Detection for Human Face Images and Videos: A Survey. *IEEE Access*, *10*, 18757–18775. https://doi.org/10.1109/ACCESS.2022.3151186

Masood, M., Nawaz, M., Malik, K. M., Javed, A., Irtaza, A., & Malik, H. (2023). Deepfakes generation and detection: State-of-the-art, open challenges, countermeasures, and way forward. *Applied Intelligence*, *53*(4), 3974–4026. https://doi.org/10.1007/s10489-022-03766-z

Microsoft. (2017). *Video Conferencing, Meetings, Calling | Microsoft Teams*. https://www.microsoft.com/en-us/microsoft-teams/group-chat-software

Mukta, M. S. H., Ahmad, J., Raiaan, M. A. K., Islam, S., Azam, S., Ali, M. E., & Jonkman, M. (2023). An Investigation of the Effectiveness of Deepfake Models and Tools. *Journal of Sensor and Actuator Networks*, *12*(4), Article 4. https://doi.org/10.3390/jsan12040061

Nguyen, T. T., Nguyen, Q. V. H., Nguyen, D. T., Nguyen, D. T., Huynh-The, T., Nahavandi, S., Nguyen, T. T., Pham, Q.-V., & Nguyen, C. M. (2022). Deep learning for deepfakes creation and detection: A survey. *Computer Vision and Image Understanding*, *223*, 103525. https://doi.org/10.1016/j.cviu.2022.103525

OpenAI. (2024, March 13). *Introducing ChatGPT*. https://openai.com/index/chatgpt/

ppverbeek. (2010). *Peter-Paul Verbeek*. Peter-Paul Verbeek. https://ppverbeek.org/

Robson, G. J., & Tsou, J. Y. (Eds.). (2023). *Technology Ethics: A Philosophical Introduction and Readings*. Routledge. https://doi.org/10.4324/9781003189466

Rosenberger, Kiran, A., Verbeek, P.-P., Ihde, D., Langsdorf, L., Besmer, K. M., Hoel, A. S., Carusi, A., Nizzi, M.-C., & Secomandi, F. (2015). *Postphenomenological Investigations: Essays on Human-Technology Relations*. Lexington Books/Fortress Academic. http://ebookcentral.proquest.com/lib/aalborguniv-ebooks/detail.action?docID=2055678

Shoaib, M. R., Wang, Z., Ahvanooey, M. T., & Zhao, J. (2023). Deepfakes, Misinformation, and Disinformation in the Era of Frontier AI, Generative AI, and Large AI Models. *2023*

International Conference on Computer and Applications (ICCA), 1–7. https://doi.org/10.1109/ICCA59364.2023.10401723

Spradley, J. (1979). *ASKING DESCRIPTIVE QUESTIONS*.

Stony Brook University. (2025). *Don Ihde*. https://www.stonybrook.edu/commcms/philosophy/people/_faculty/ihde.php

Sun, F., Zhang, N., & Song, Z. (2021, November 24). *Deepfake Detection Method Based on Cross-Domain Fusion—Sun—2021—Security and Communication Networks—Wiley Online Library*. Https://Onlinelibrary.Wiley.Com/. https://onlinelibrary.wiley.com/doi/full/10.1155/2021/2482942

Verbeek, P.-P. (2015). Beyond interaction: A short introduction to mediation theory. *Interactions*, *22*(3), 26–31. https://doi.org/10.1145/2751314

Verbeek, P.-P. (2016). *Toward a Theory of Technological Mediation: A Program for Postphenomenological Research*. Lexington books.

Yadav, A. (2025, March 6). What is Gaussian Noise and Why It’s Useful? *Medium*. https://medium.com/@amit25173/what-is-gaussian-noise-and-why-its-useful-b3c50dd14628