Deepfakes in Use

Rethinking the Infopocalypse through Postphenomenology and Wittgenstein

Luuk Stellinga

04-02-2022

Supervisor: Dr. Julia Hermann

Second reader: Prof.dr.ir. Peter-Paul Verbeek

Master thesis for the completion of MSc Philosophy of Science Technology and Society

Faculty of Behavioural, Management and Social Sciences

University of Twente

Enschede, the Netherlands

TABLE OF CONTENTS

INTRODUCTION

This thesis is an exploration of the societal impact of deepfake technology, with a focus on the so-called *Infopocalypse*. Deepfakes are convincing fake versions of various types of digital media content, achieved through the application of Machine Learning systems (De Vries, 2020). There are numerous concerns that can be raised with respect to the rising prominence of deepfakes in the digital landscape, but perhaps the most alarming concern refers to the belief that deepfake technology will irreversibly blur the lines between what can be considered 'real' and what can be considered 'fake' (Ovadya, 2018). This concern is captured in popular discourse by the notion of the Infopocalypse. While I believe this notion is fueled by legitimate concerns, I suggest that it is lacking with regard to its philosophical underpinnings.

The existing philosophical discourse regarding the societal impact of deepfake technology can be characterized as having an epistemological focus, whereby the infopocalypse is largely understood as a problem of acquiring information from digital media (see, e.g., Rini, 2019, Fallis, 2020). While this focus addresses an important aspect of human-media interaction, it fails to acknowledge the nuances of human experience, especially in relation to technologies. A proper analysis of the impact of deepfake technology, then, is based on an understanding of technologies as technologies *in use*. This can be achieved through the application of notions from the field of Postphenomenology, and insights from a Wittgensteinian understanding of technology use. Combined, these perspectives lead to an appreciation of deepfakes as mediating human experience of digital media, and as necessarily embedded in social and cultural contexts. This understanding overcomes both the shortcomings of the popular understanding of the infopocalypse, as well as the current state of the philosophical discourse. The societal impact of deepfake technology is not to be understood as information-centric, described in abstract terms, and depicted as an apocalyptic scenario, but rather as complex, multifarious, and ongoing.

The thesis is divided into four chapters. Chapter one, titled 'Infopocalypse,' introduces the popular concern regarding deepfake technology, captured by the notion of the *Infopocalypse*, and as such sets the stage for the following chapters. Chapters two and three, titled 'Human-deepfake-world' and 'Deepfake-games,' respectively, both offer a philosophical analysis of the infopocalypse. The former does so from a postphenomenological perspective, and the latter from a Wittgensteinian perspective. Finally, chapter four, titled 'Deepfakes in Use' reflects on the analyses of the previous chapters by considering the resulting understanding of the infopocalypse, as well as the relation between the two perspectives. I conclude that the societal impact of deepfake technology must be understood as an ongoing dynamic between individual human-technology relations, and the grammars of social practices and cultural meanings.

CHAPTER 1   INFOPOCALYPSE

This chapter introduces the notion of the *Infopocalypse*, which I take to refer to the general concern surrounding the societal impact of deepfakes. I take that the Infopocalypse goes beyond mere cases of bad-use, and rather suggests a type of societal transformation, or disruption, akin to that of early cinema. Finally, I argue that the popular understanding of the infopocalypse is philosophically lacking, and that a proper understanding of technologies *in use* is required, both to appreciate the interaction between human beings and deepfakes, as well as the social and cultural embeddedness of this interaction.

Part 1 - What are Deepfakes?

On the 17th of April 2018, BuzzFeed published a video to the internet with the title 'You Won't Believe What Obama Says In This Video! 😉'. The upload, which has since attained a combined 18 million views on Youtube, Facebook and Twitter, depicts former U.S. president Barack Obama giving a public service announcement on the topic of deepfake technology (Facebook, 2018, Twitter, 2018, YouTube, 2018). Initially, the content of the video is convincing. The former president starts out by addressing his audience in a familiar tone and with his typical facial expressions. After about fifteen seconds, however, the video takes an unexpected turn, as the audience hears and sees the former president utter uncharacteristic sentences. Finally, it is revealed that the person talking is not actually Barack Obama, but actor, director, and impersonator Jordan Peele. An audio recording of Peele was matched with a computer-generated video of Obama 'lip-syncing' to the words, to make it appear as though the latter was making a public service announcement that never actually happened.

      The video described above both offers an example of a so-called deepfake, and simultaneously functions as a warning of the unwanted consequences of deepfake technology. The word "deepfake" is a portmanteau of the terms *deep learning* and *fake*, to refer to the use of machine learning systems in creating convincing but fabricated video footage, photographs, audio recordings, and more (De Vries, 2020). Since the incident in April of 2018, deepfake technology has not only improved immensely, but has also become much easier for ordinary users to get their hands on. As of 2022, dozens of free mobile phone applications allow users to create deepfakes of similar quality to the Obama video, with some having tens of millions of downloads (e.g., FakeApp, ReFace). For the more dedicated deepfake enthusiast there exists publicly available software to create Hollywood-grade fake video and audio that anyone with an above average computer can run (e.g., faceswap, Deepfacelab).

As a result of these developments, deepfakes are no longer the anomaly that they were in 2018 but have become a common feature of digital media on the internet. Deepfakes appear regularly on timelines and have even become the subject of viral trends (see, e.g., Victor, 2021). This increase in social presence provides all the more reason to be critical of how deepfake technology can be used and abused, as well as how it may affect the ways people experience and interact with the internet and each other. In the remainder of this section, I will briefly describe the underlying systems that are used to create deepfakes, and consider some of the more obvious concerns that people have articulated with regard to deepfakes.

Machine Learning and Generative Adversarial Nets

The field of Machine Learning (ML) is predicated on a commitment to indirect rather than explicit programming. Whereas in more traditional AI methods a computer system is provided with an explicit model by its designer, ML systems are programmed to create models themselves from sets of examples (de Vries, 2020, p. 2111). As such, the computer is not told what to do directly, but *learns* what to do on its own (hence leading to the name Machine *Learning*). Within the field of Machine Learning, a further distinction can be made between supervised and unsupervised learning. These two types of learning differ with respect to the type of input that is used, as well as with respect to the output that is produced (de Vries, 2020, p. 2112).

Supervised ML systems work with *labeled* examples. This is to say that if you wanted a system to learn the differences between, e.g., different dog breeds, you would train it on a dataset of pictures of dogs with a label stating their breed. As such, the system would learn how to recognize particular breeds and classify them. Unsupervised ML systems, on the other hand, work with *unlabeled* examples. To continue the same analogy, this system would be trained on pictures of dogs *without* the label. In turn, the system does not learn (in the first place) how to classify dog breeds. Rather, it discovers common patterns in the pictures, ultimately leading it to a general understanding of what dogs look like. With respect to its output, this gives unsupervised ML systems the capacity to describe data structures, and, largely due to the recent development of Generative Adversarial Nets (GANs), generate output that resembles their training-data.

Generative Adversarial Nets are built on *neural networks*, which are in turn characterized by a layered approach to analyzing its given input (de Vries, 2020, p. 2113). Here, the analysis begins with very basic features to increasingly more high-level ones, each layer compressing the input more. In GANs, two neural networks are pitted against each other in an adversarial relationship (Goodfellow et al., 2014, p. 1) (see figure 1). One of those networks takes on the role of the *Generator* and the other the role of the *Discriminator*. These two neural networks then engage in a game where the Generator attempts to trick the Discriminator into thinking a piece of generated output is real, and the Discriminator tries to correctly

distinguish between real input and generated output. The neural networks improve through back propagation, learning from the accuracy of the Discriminator's predictions. When the Discriminator can no longer improve its performance, the game reaches a zero-sum state, and the discriminator is removed from the system. At this point, the generator has successfully learned to produce output resembling its training data.
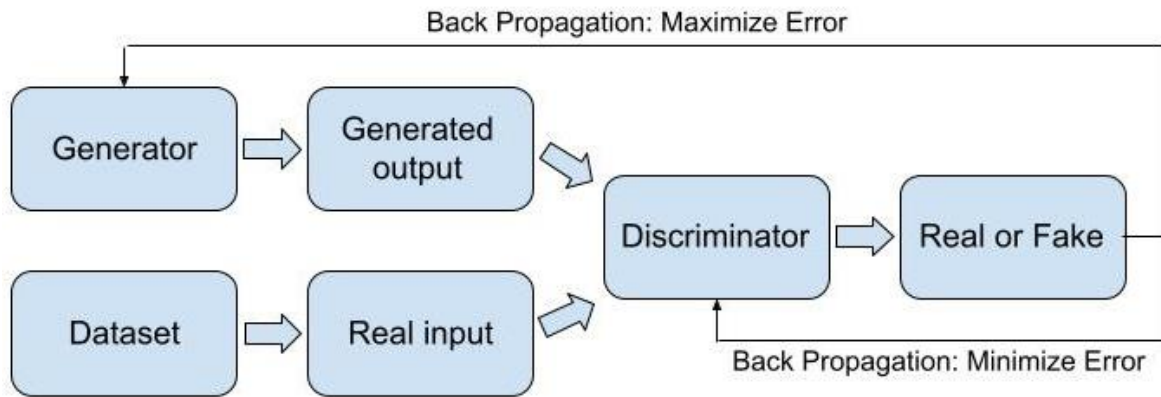
One application of GANs is to create fake versions of digital media content, or, *deepfakes*. It is important to acknowledge, however, that GANs are used for various applications besides creating deepfakes. Especially in the fields of image processing and computer vision, GANs have been successfully used for upscaling, image synthesis and manipulation, texture synthesis, and more (Giu et al., 2020). Those applications certainly raise interesting philosophical questions on their own, but they are beyond the scope of this thesis. In this thesis, I focus solely on deepfakes, whereby ML systems are used to create fabricated versions of digital media content. As deepfakes are increasingly saturating digital platforms, it is important to consider how they may come to affect society. The next part of this chapter will consider concerns surrounding deepfake technology, starting with concerns regarding malicious use cases before turning to the perceived long-term societal impact of deepfake technology referred to as the 'Infopocalypse.'

Part 2 - Bad Use and the Infopocalypse

The concerns surrounding deepfake technology can broadly be categorized into two camps. On the one hand, there exists concern about how the technology may be used by people with bad intentions. I will refer to such concerns as *bad use* concerns. Although it is important to explore and anticipate bad use, it will here only be discussed as a precursor to the main topic of this thesis project, which concerns deepfake

technology's broader societal impact. Such concerns regard the impact that deepfakes may have on society as a whole, regardless of the intentions of individual actors. I appeal to the notion 'Infopocalypse' to present concerns of this kind.

2.1 - Bad use

One immediate concern that should be raised is the harm that deepfakes can do to its subjects. Deepfake technology grants its users the ability to make others appear as though they are doing specific tasks and/or saying specific things. As such, deepfake technology can allow someone with bad intentions to portray others in ways harmful to them. In March of 2021, for example, a mother from Pennsylvania reportedly shared deepfaked videos of her daughter's cheerleading rivals, in the hope of getting them removed from their cheerleading program (BBC News, 2021). The videos, which depicted the teenage girls naked, drinking and smoking, were shared with their coach. Although in this instance the videos were recognized as fake, it shows the potentially harmful effects that deepfakes can have on individuals.

Another clear example of bad use regards the use of deepfake technology for committing fraud. An example of this occured in 2019, when it was reported that a fraudulent transfer of €220,000 resulted from a deepfaked audio recording (Stupp, 2019). Criminals had used deepfake technology to impersonate the voice of the chief executive of an energy firm. This faked recording was used to direct the transfer of the aforementioned sum into a bank account that later appeared to belong to scammers. The attack was made possible by a convincing deepfake of the CEO's voice, which granted the fraudster immediate trustworthiness to the person picking up the phone. Without a good way of detecting such fabricated phone calls, there may be many cases of deepfake fraud in the future.

Finally, one may consider the use of deepfake pornography as bad use. It is important to note that this type of use is not an outlier, as fake non-consensual adult content makes up the vast majority of deepfakes that are currently online (Patrini, 2019). Here, deepfake technology is used to superimpose the face of one person on the body of someone else, often with the intention to be used as blackmail or so-called 'revenge porn' (Henry et al., 2018). Although this certainly brings harm to the subject of the video, the impact of such cases reach far wider. In the third chapter of this thesis, I will revisit deepfake pornography. I will suggest that the Wittgensteinian notions of language-games and forms of life can be used to address how deepfakes are embedded in social and cultural contexts of gender meanings.

2.2 - The Infopocalypse

Beyond what had been mentioned above, there exists a set of concerns which are not characterized by use, but rather regard the transformative effect that deepfakes may have on society. This sentiment is expressed, for example, by deepfake researcher Aviv Ovadya, who coined the term *Infocalypse* or *Infopocalypse* to describe a future irreversibly impacted by deepfake technology (Ovadya, 2018). Specifically, Ovadya claims that deepfakes constitute a threat to democracy, stating that "[advances] in communications technology and artificial intelligence are making it easy to create audio or video content with potentially dangerous consequences, from making it appear that a world leader is ordering a nuclear strike to simulating your spouse's voice on the phone asking for a bank password" (Ovadya, 2018, p. 43). Although it may seem like Ovadya is simply conjuring up more bad use cases, it is made clear that his caution extends further when he describes deepfake technology as "the distortion of reality itself" (Ovadya, 2018, p. 43).

Warnings of the Infopocalypse have since featured regularly in popular media, with correspondingly strong rhetoric (Schwartz, 2018, Toews, 2020). A common characteristic of the discussions in these pieces is the fact that they are not so much concerned with harm done to individuals, but rather focus on how deepfakes may affect the way people interact with digital media generally. Along these lines, the Guardian published about deepfake technology that "the greatest threat isn't that people will be deceived, but that they will come to regard everything as deception" (Schwartz, 2018). Similarly, Forbes warns that "the mere possibility that a video could be a deepfake can stir confusion and facilitate political deception regardless of whether deepfake technology has actually been used." (Toews, 2020). A recent study on the deceptiveness of deepfakes, which shows that over 30 percent of people confronted with a deepfake report being uncertain about whether its content is real or fake, seemingly repeats this conclusion (Vaccari & Chadwick, 2020, p. 7). The authors of the study conclude that "[deepfakes] may sow uncertainty which may, in turn, reduce trust in news on social media" (Vaccari & Chadwick, 2020, p. 9).

The Infopocalypse, based on the above mentioned discussions, is to be understood as referring to the belief that deepfake technology will irreversibly blur the lines between what is considered 'real' and 'fake.' As discussed, warnings of the infopocalypse are often not so much concerned with deepfake technology's direct and measurable impact, but instead posit deepfakes in relation to everyday practices and even abstract concepts such as "truth" and "reality" (Schwartz, 2018, Ovadya, 2018). As such, the type of technological impact that the Infopocalypse refers to is perhaps best captured by terms such as *social transformation* or *social disruption*. Recently, such terms have received increasing attention in the Philosophy of Technology, where concerted efforts have been made to address not just the benefits and harms that new technologies cause to individuals, but to address moreover how they may cause widespread change at the level of shared norms and behaviors (see, e.g., Nickel et al., 2020, Hopster, 2021).

To account for the "broader dynamics of social transformation engendered by technologies," Jeroen Hopster introduces the notion of *Technosocial Disruption* (Hopster, 2021, p. 1). This notion can be

characterized by two main characteristics. First, it emphasizes a close-knit relationship between technology and human life. Technological artifacts do not only impact human beings by allowing them to use the technology in various ways, but moreover mediate their perceptions and actions, change their moral norms and values, and even challenge fundamental categories of thought (Hopster, 2021, p. 6). As such, disruptive technologies may affect human life at various levels. Secondly, technosocial disruption is characterized by causing uncertainty. Hopster argues that "the notion of disruption foregrounds the reactive mode in the face of change" (Hopster, 2021, p. 8). This means that for a new technology to be disruptive, its impact has to be unpredictable. In contrast with other forms of social transformation, disruptions are likely to be rapid, unanticipated, and causing uncertainty.

2,3 - L'Arrivée

To clarify how deepfakes may be understood as challenging concepts as as truth and reality, it may be useful to consider it as analogous to the historical case of cinema. Consider the French film *L'Arrivée d'un train en gare de La Ciotat* (translated from French to English as *The Arrival of a Train at La Ciotat Station*). According to a longstanding myth in the history of cinema, this film had a profound impact on its audience when it premiered in January of 1896 (Loiperdinger & Elzer, 2004). Even though the film was a mere 50 seconds long, shot in black and white, and entirely silent, it reportedly caused great terror to those viewing it in theaters. Audience members, who had never seen film before, hurried to the back of the room, as they were unable to comprehend that the train they witnessed moving towards them was simply a rapid sequence of photographs, and not an actual train hurdling into the theater.

The introduction of cinema, according to this story, created an experience so unlike anything people had ever had before that it challenged what they knew to be real. Although the famous tale of L'Arrivée has now been identified as most likely exaggerated, it nonetheless illustrates how the human perception of reality can be shaped by technological artifacts (Loiperdinger & Elzer, 2004). By allowing people to record events and play those events at later times and in different locations, film creates new ways of experiencing the world. As, at that time, people had no proper frame of reference to make sense of these new ways of experiencing the world, it may very well have challenged some of the deeply rooted beliefs people held at the time, such as illustrated by the story above.

I suggest that the deepfake technology of today can be considered analogous to cinema in the 1890's. As *L'Arrivée* challenged the experiences and beliefs of its 19th century audience, so does the deepfaked Obama public service announcement challenge the beliefs of contemporary society. In both cases, new technological developments reveal an unprecedented way of experiencing and relating to the world that challenges how human beings unwittingly interact with their environment. Of course, the

similarities between cinema and deepfake technology are not endless, and the history of cinema is most likely not a reliable predictor for the future of digital media. Nonetheless, it demonstrates fertile ground for posing the question as to how deepfakes may challenge human beings' conceptions of reality. After all, if deepfakes will really create a future where the line between real and fake is blurred irreversible, as hypothesized by the aforementioned popular literature, it will surely impact how human beings make sense of the world at a deep and conceptual level.

Part 3 - Requirements for a Philosophical Understanding of the Infopocalypse

In the previous part of this chapter, I introduced the notion of the Infopocalypse. This notion, often used in discussions about deepfake technology, typically refers to the belief that deepfake technology will irreversibly blur the lines between what is considered 'real' and what is considered 'fake.' Although the fears expressed in discussions about the Infopocalypse might be valid, its predictions appear vague and imprecise. I suggest that there are two crucial things missing in this discussion that can be provided by the Philosophy of Technology. First, to understand the impact of deepfake technology, a proper understanding of human-deepfake interaction is required. This is to say, it is important to describe and evaluate how people experience and interact with deepfakes in their everyday lives, so that predictions about their future impact are grounded in actual use-cases. Second, it is crucial to understand how such use-cases are embedded in particular social and cultural contexts, to consequently recognize how they may disrupt those contexts. I suggest that those two elements can be provided by Postphenomenology and a Wittgensteinian perspective on technology. Supplementing the discussion about the societal impact of deepfake technologies with this philosophical understanding of technologies in use allows one to rethink the Infopocalypse.

3.1 - Human-deepfake-world

Understanding the impact that technologies have on human beings is perhaps the main purpose of the philosophy of technology. As already mentioned in this chapter, technologies can impact people in ways that do not fit with the instrumentalist schema of technology-as-tool. Along these lines, understanding the impact of deepfake technology, too, requires a nuanced account of human-technology relations. One especially fruitful approach to understanding the shaping character of technologies is *Postphenomenology*, which is not only strongly rooted in the traditions of 20th century phenomenology and American pragmatism, but also provides a rich vocabulary to explore human-technology relations. According to the postphenomenological perspective, technologies are not mere tools, but should always be understood as

mediators in the co-constitutive relation between object and subject, or human being and world (Verbeek, 2001).

As such, understanding the impact and meaning of a technology requires understanding how it is used in an everyday context, and how it shapes a human being's relation to their environment (Ihde, 1990, p. 70). This insight has led postphenomenologists to articulate a vocabulary of *mediated human-technology-world relations*, through which the relationship between a human being and a technological artifact can be explored. Applying this vocabulary to the case of deepfake technology allows for an exploration as to how human beings experience deepfake, as well as how they experience their environment as mediated by deepfakes.

The second chapter of this thesis project attempts to meet the challenge of exploring human-deepfake-world relations. To do so, it will first survey the existing appraisals of deepfake technology in philosophical literature (e.g., Floridi, 2018, Rini, 2019, Fallis, 2020), and argue that the existing literature is limited in its mostly epistemological focus, whereby it considers the impact of deepfake technology as posing a question of knowledge and information. The subsequently presented postphenomenological analysis is able to overcome this limitation by placing its focus on experience. The experience of a deepfake includes obtaining knowledge, but is moreover embodied, emotional, and has lasting impact even when the deepfake is not actively engaged with. As a result, Postphenomenology offers a more nuanced understanding of human-deepfake relations.

3.2 - Deepfake-games

The postphenomenological approach is conducive to an insightful understanding of deepfakes in use. As much as the postphenomenological artifact-centered focus enables, however, it is also limiting. By focusing on relations between individual artifacts and individual users, Postphenomenology is perceived as less capable of addressing social and cultural dimensions of use (Rosenberger, 2014, p. 372, Coeckelbergh, 2018a, p. 5). I suggest that this shortcoming can be overcome by accompanying Postphenomenology with a Wittgensteinian perspective, that has recently been developed in the Philosophy of Technology by, most prominently, Mark Coeckelbergh (Coeckelbergh, 2018b). Here, the notions of *language-games* and *forms of life*, which Wittgenstein originally introduced to elucidate the nature of the use of language, are applied to the use of technologies.

The Wittgensteinian perspective is particularly useful in addressing the social and cultural embeddedness of technology use. Artifacts are understood as existing within so-called "technology-games," which are themselves embedded in larger forms of life (Coeckelbergh, 2018b). As such, individual instances of the use of technological artifacts can never be understood in isolation, but must rather be

regarded as constituted by its social and cultural contexts. At the same time, those cultural and social contexts are themselves constituted by individual instances, leading to a holistic relationship between use and culture. Understanding the social and cultural embeddedness of technology use, then, consequently helps understand the social and cultural impact of technology.

Chapter three applies this Wittgensteinian perspective to the case of deepfakes and the Infopocalypse. It considers three cases, presented in a recent STOA report about the impact of deepfakes, to illustrate how deepfakes may be understood as given meaning by deepfake-games and forms of life. The cases of manipulated court evidence, deepfake pornography, and a false political statement all show how the impact of deepfake is in part determined by the rules of particular language-games and by forms of life, while at the same time strengthening or challenging those rules. In doing so, it supplements the Postphenomenological perspective by addressing the social and cultural embeddedness of technologies in use.

Conclusion

This chapter introduced the notion of the Infopocalypse, referring to the perceived disruptive societal impact of deepfake technology. I suggest that discussions surrounding this notion are philosophically lacking, and that a proper understanding would incorporate an appreciation of human-deepfake interactions and the relationship between those interactions and the social and cultural contexts that they appear in. This can be done, I will argue in the following two chapters, through a Postphenomenological and Wittgensteinian analysis.

CHAPTER 2   HUMAN-DEEPFAKE-WORLD

The previous chapter introduced the notion of the 'Infopocalpyse,' referring to the popular concern surrounding the societal impact of deepfake technologies. Those putting forward the notion suggest that the advance of deepfake technology will have widespread and irreversible consequences for society, leading to things such as "the distortion of reality itself" (Ovadya, 2018). I argued that, although the infopocalypse might be fueled by legitimate concerns, it is lacking with respect to a philosophical understanding of technologies in use. This chapter attempts to provide part of that philosophical understanding. I briefly discuss three existing philosophical appraisals of deepfakes, and suggest that they do not adequately address the role of human experience. This can be accounted for through a postphenomenological perspective, which understands technological artifacts as mediating human-world relations. After providing a *postphenomenological* analysis of deepfakes, is it argued that both the popular conception as well as existing philosophical literature overemphasize the importance of information, effectively ignoring other aspects of how human beings experience deepfakes.

Part 1 – Information-centered Accounts[1]

Before offering my own analysis, I first review some of the existing philosophical appraisals of deepfake technology. I introduce Luciano Floridi's conceptualization of deepfakes as ectypes, which I believe to be the first philosophical analysis of deepfakes. I then discuss two later publications by Regina Rini and Don Fallis who both argue that deepfake technology challenges our capacity for acquiring knowledge, albeit in different ways. After discussing these works, I point out that they are guided by a focus on knowledge. Although I concede that knowledge plays an important role in how human beings experience deepfakes, it does not paint the full picture. Human beings do not only experience deepfakes as containing information. Moreover, deepfakes present themselves as others and can become part of a person's perceptual background.

1.1 - Ectypes, backstop crises, and the epistemic status of video

---

[1] The works discussed in this section are not intended to cover all existing literature. It does not mention, e.g., the work of Kerner & Risse or Chesney & Citron, which are undoubtedly valuable for the discussion (Kerner & Risse, 2021, Chesney & Citron, 2019). I have decided to not discuss these works in favor of others that are explicitly philosophical. do believe, however, that my treatment of the selected works gives a decent indication of the contents of the general academic discourse about deepfakes

The first acknowledgement of deepfake technology in the philosophical literature consists of a short article written by prominent Oxford philosopher of technology Luciano Floridi, published less than four months after the release of the now infamous Obama deepfake discussed in the previous chapter. This article, titled 'Artificial Intelligence, Deefakes and a Future of Ectypes,' is primarily concerned with the question as to what kind of artifact a deepfake is (Floridi, 2018). Approaching this question from the context of art, and in particular art replicas, Floridi suggests that a deepfake may best be understood as a kind of ectype; "a copy that has a special relation with its source (the origin of its creation), the archetype" (Floridi, 2018, p. 319). By understanding deepfakes as ectypes, Floridi attempts to shed light on how deepfakes may be both authentic while unoriginal, or original while inauthentic.

The article presents 'The Next Rembrandt,' a project where artificial intelligence was used to analyze the known works of famous painter Rembrandt van Rijn, to generate a painting that could easily pass as one of Rembrandt's works. This fake Rembrandt, which is essentially a sophisticated deepfake, stands in a special relation to its source by closely, and arguably authentically, replicating the style and content of Rembrandt's paintings. At the same time, the fake Rembrandt lacks the originality of real Rembrandts. Conversely, deepfakes may maintain originality while lacking authenticity, as demonstrated by a deepfaked recording of John F. Kennedy's speech of 22 November 1963, which never actually occurred. AI was used to analyze recordings of Kennedy's voice and give a voice to his 'final' speech. In this case, the ectype is original, in that the text came directly from the source, but inauthentic, in that it involved a mimic of JFK's voice (Floridi, 2018, pp. 319-320).

While Floridi's article extends an intriguing invitation to thinking about the philosophical puzzlement surrounding deepfakes, it cannot properly be classified as a work about the infopocalypse. As mentioned, Floridi seeks to investigate the unique nature of deepfakes, and is not primarily concerned with their societal impact. In discussing concerns of deepfake technology, the article limits itself to individual use-cases, suggesting that problems arise when those creating deepfakes "cheat," and try to illegitimately pass their creations off as authentic or original (Floridi, 2018, p. 320). Thus, while the article convincingly demonstrates an interesting relationship between deepfakes on the one hand and concepts such as authenticity and originality on the other, it does not consider whether deepfakes may have a long-term or wide-spread transformative effect due to this relationship.

Since Floridi's article in 2018, more philosophers have turned their attention toward deepfake technology. Hereunder I discuss two publications, namely Regina Rini's 'Deepfakes and the Epistemic Backstop' and Don Fallis' 'The Epistemic Threat of Deepfakes' (Rini, 2019, Fallis, 2020). The works of Rini and Fallis go beyond that of Floridi in that they explicitly aim at understanding the societal impact of deepfake technology, and thus the infopocalypse. Both Rini and Fallis' work discuss the impact of

deepfakes as an epistemological question, arguing that deepfakes affect societies by interfering with and harming our abilities to acquire knowledge, albeit in different ways.

Rini discusses the epistemic consequences of deepfake technology by arguing that it may cause a so-called "backstop crisis" (Rini, 2019). This argument is based on the idea that audio and video conventionally function as backstops in public discourse, such that disagreements can be settled by referring to evidence in the form of a recording. Video and audio recordings, it is suggested, regulate discourse by functioning as testimony and motivating people to testify truthfully in the knowledge that there may exist recordings. The danger of deepfakes, then, is that "they will gradually eliminate the epistemic credentials of all recordings, to an extent that video and audio no longer serve their passive regulative function in testimonial practice" (Rini, 2019, p. 8). As a result, Rini argues that deepfakes challenge existing testimonial norms, leaving an uncertain future that may significantly damage social and political systems (Rini, 2019, p. 14).

Fallis similarly argues that deepfake technology challenges our ability to acquire knowledge, but places his focus on the epistemic value of video itself, rather than testimony. According to Fallis, the existence of deepfake technology causes all videos to carry less information, as the possibility of any video being a deepfake increases (Fallis, 2020, p. 10). This argument is based on the idea that the information carried by a particular video depends on the probability of that video depicting something false. That is to say, a video depicting a politician taking a bribe will carry more information the more likely it is that the event depicted actually happened. However, with the advance of deepfake technology, combined with the inability of audiences to recognize whether videos are real or fake, the chance of the video being faked becomes more likely and the video will therefore carry less information. Based on this argument Fallis concludes that deepfakes challenge our ability to acquire knowledge from video.

1.2 - Addressing the role of human experience

Reviewing the existing philosophical literature about deepfakes reveals that the question of understanding the infopocalypse is generally understood as an epistemological question. This is to say, deepfakes are considered to be philosophically interesting insofar as they may interfere with our capacity for obtaining knowledge from digital media. Although this focus certainly addresses an important aspect of the infopocalypse, it seems to be based on an overly simplistic view of how human beings experience and interact with deepfakes, digital media and technology more generally. Human beings do not merely perceive digital media content as presenting true or false information about their environment, but moreover experience it as actively shaping their perception in various ways. While the analyses of Floridi, Rini and Fallis may at times hint at more complex human-technology interactions, those remain in the background.

I suggest that a proper philosophical appreciation of the societal impact of deepfake technology should bring the relation between technological artifacts and human experience to the forefront.

The remainder of this chapter will attempt to provide such an appreciation by introducing and applying central concepts of *Postphenomenology*. The postphenomenological perspective, which understands technological artifacts as mediating perceptions and actions, not only does justice to the complexities of human-technology interaction, but also provides useful vocabulary to study particular cases. By applying this vocabulary to the case of deepfakes, I make sense of the various ways in which human perception is shaped by digital media, and the role that deepfakes play in this.

Part 2 - Postphenomenology

Postphenomenology was coined and founded by American philosopher of technology Don Ihde, whose seminal book *Technology and the Lifeworld* is widely considered to have launched the tradition of postphenomenological thought (Ihde, 1990). Since then, the approach has gained popularity and is quickly maturing into one of the most exhaustive traditions in contemporary Philosophy of Technology. This section will introduce Postphenomenology as characterized by a commitment to a Phenomenological understanding of how technology shapes human experience, while maintaining a pragmatic commitment to a situated ontology.

2.1 - The Phenomenology of Postphenomenology

The phenomenological tradition, which started in the first half of the 20th century as a response to logical positivism, can be defined by its commitment to the idea that there exists an intimate link between the human subject and their experienced world. Whereas logical positivists of the time sought to apply the discourses of the physical sciences to all forms of human enquiry, a young phenomenological tradition, led by Edmund Husserl, argued that attention should be paid to the role that humans play in the construction of the experienced world (Husserl, 197). Husserl was soon accompanied by such figures as Maurice Merleau-Ponty, Martin Heidegger and more, whose combined oeuvre is now regarded as the phenomenological cannon.[2] Although there exist significant differences between the works of these authors, extending from their preferred methods to philosophical views, there is a shared focus on how human beings experience and interact with their perceived environment.

---

[2] Maurice Merleau-Ponty's *Phenomenology of Perception* and Martin Heidegger's *Being and Time* are considered classics in the phenomenological canon (Merleau-Ponty, 1996, Heidegger, 1962).

The core ideas of phenomenology can be explained by appealing to the notions of *intentionality* and *constitution* (Ströker, 1997). Experiencing is an intentional act, meaning that one cannot experience without having an experience *of something*. As such, understanding someone's experience requires understanding what they are experiencing. At the same time, the external world cannot be known without being experienced; it is constituted by human experience and is thus necessarily a world *for someone*. Consequently, understanding the world requires understanding how it is experienced. Considering both intentionality and constitution, it becomes clear that a human being and their environment cannot be thought of as separate from one another, but must be understood as mutually constituting each other.

With respect to understanding technology, Phenomenology allows for an alternative to the traditional instrumentalist view of technology as mere tools. Instead, technology can be understood as directing both intentionality and constitution, effectively shaping both a human being's experience and their experienced environment. As a result, the human being and the experienced environment are both shaped through technology. Technology presents the world to human beings in a specific way, and allows for a specific way for human beings to experience it. Martin Heidegger, perhaps the most influential philosopher to develop a Phenomenology of technology, captures this thought by stating that "technology is a mode of revealing," where (modern) technology reveals the world to human beings in a specific way (as standing-reserve) (Heidegger, 1954, p. 13).

2.2 - The Pragmatism of Postphenomenology

American scholar Don Ihde appropriates this Heideggerian perspective to develop a more empirically oriented philosophy of technology known as Postphenomenology (Ihde, 1990). Postphenomenology combines the core ideas of Phenomenology with the ontological commitments of the American Pragmatist tradition, posing all claims from an embodied and situated perspective, referring to practical problems, and maintaining an empirical orientation (Rosenberger & Verbeek, 2015, p. 1). Ihde's move towards this pragmatic outlook is part of what Hans Achterhuis has called *the empirical turn* (see Achterhuis, 2001), wherein philosophers of technology started to focus on concrete technologies and issues to move away from abstract and deterministic theories that had been popular for the better part of the twentieth century (Brey, 2010, p. 39).

For Postphenomenology, the empirical turn has generally been understood as a *turn towards things*, indicating that understanding the shaping character of technology should happen at the level of the individual artifact (Verbeek, 2001, p. 76). Besides this, the empirical turn may also be understood as a *turn towards use*. By adopting the pragmatic vocabulary of *practice*, rather than theory, Postphenomenology turns its attention to the significance and complexities of the use of technology (Ihde, 2009, pp. 9-11). Ihde

writes: "it is the use of an artifact that makes it what it is" (Ihde, 1990, p. 70). As a result, Postphenomenology must be understood as an understanding of *technologies in use*; it asks not what is revealed in our relationship to Technology as an abstract and monolithic concept, but rather how individual technological artifacts shape human experiences and actions through use.

2.3 - Postphenomenology

The resulting view combines the strengths of Phenomenology and American Pragmatism to offer a philosophy of technology that understands artifacts as shaping human experiences and actions. Besides its philosophical roots, the success of Postphenomenology may be attributed to its production of a rich vocabulary to describe human-technology relations. Notions like mediation (see Verbeek, 2015), and multistability (see Rosenberger, 2014) provide useful frameworks to generate insights into cases of various human-artifact relations, ranging from reading ultrasound scans to driving over speed bumps and beyond. The applicability of Postphenomenology is moreover demonstrated by the impressive growth of literature, rendering it one of the leading traditions in contemporary philosophy of technology.

Regarding its role in this thesis specifically, Postphenomenology is attractive for a few reasons. First, it can contribute to the popular discourse on the Infopocalypse by providing a (currently lacking) philosophically informed understanding of how human beings interact with deepfakes. Second, it can provide an experience-focused perspective, which presents a welcome alternative to the knowledge-focused perspectives that dominate contemporary literature. As such, it does justice to the complexities of human experience and the role that technological artifacts play in constituting this experience..

Human-technology-world relations

Postphenomenology offers an assortment of insightful vocabulary to guide the exploration of human-deepfake interactions. As a result of these core-commitments, mediation theorists have considered the role of technologies in terms of human-technology relations. Most notably, Don Ihde identified four basic types of human-technology relations (Verbeek, 2001, pp. 127-133). First, there exist two sorts of relations wherein human beings relate to their environment via a technological artifact: in embodiment relations a technological artifact acts as an extension of the human body, as such allowing the human being to perceive their environment through the artifact, and in hermeneutic relations a human being relates to a technological representation of their environment. Furthermore, Ihde identified alterity relations, whereby a human being relates directly to the technological artifact, and background relations, whereby a human being relates to their environment under technological conditions. Although Ihde's precise categorizations can be criticized,

the vocabulary of human-technology relations does allow for a systematic way to assess the mediating character of technology.

Part 3 – A Postphenomenology of Deepfakes

As concluded in the previous section, Postphenomenology offers a fruitful perspective for understanding the infopocalypse. This section will therefore constitute a postphenomenological analysis of the infopocalypse, wherein it is specifically explored how human beings experience and interact with deepfakes.

3.1 – The frame and the window

The first thing to note when considering the experience of a deepfake is that a deepfake cannot be isolated from the digital technologies that surround and facilitate it. A deepfake would not exist without the screen that displays it, the website that hosts it, or any other technology that is necessary for its existence. As such, I suggest that the perception of a deepfake is best understood as incorporating the simultaneous perception of various digital technologies. Understanding these experiences is crucial to understand how human beings experience deepfakes. An early observation is that experiencing a deepfake involves interacting with a computer interface, whether that be a desktop computer, a smartphone, a tablet, or some other digital device; seeing a deepfake involves 'staring at a screen.'

It may seem counterintuitive to say that the experience of a deepfake involves staring at a screen, as the screen is generally not the subject of one's focus when one is interacting with a digital device. This phenomenon is described by postphenomenologist Robert Rosenberger, who writes that "[many] aspects of computer use are experienced with a high degree of transparency. (...) Focus is placed on the content displayed on the screen, rather than the screen itself or its borders. While engrossed in what is being done, the user may be scarcely consciously aware of her or his situation of sitting in front of a computer at all" (Rosenberger, 2009, p. 178). This notion of transparency can be understood through the embodiment relation, where a high level of embodiment leads itself to a high degree of transparency. When a user is familiar with an artifact like they are with their own body, the artifact becomes transparent to their conscious perception. Of course, screens are not always perceived as transparent. Consider the experience one has when one buys a new phone and sees it for the first time, or, perhaps more evidently, when the screen suddenly breaks.

From this analysis, one may understand the screen as a kind of *frame*, framing the deepfake and other pieces of media content. Much like the frame of a window, attention is not generally paid to the frame itself, but rather to what it frames. As such, the screen may be understood as a *window*; a way for human beings to experience their environment through digital representations (see Introna & Ilharco, 2004, Wellner, 2011). From this description, a *hermeneutic* relation between human beings and digital devices becomes visible, whereby a human being interprets (or *reads*) the digital representations displayed to learn about and experience the world around them. In the same way as looking through an ordinary window, looking at a computer or smartphone makes certain aspects of the world accessible to the person looking. However, similarly to how a window may be tinted or curved, so too digital representations may distort one's perception of their environment.

3.2 – The other and the background

So far, I have discussed the perception of a deepfake in terms of embodiment and hermeneutic relations, as such highlighting the experience of the digital technologies that frame and support its existence. These two relations lend itself to the following picture: the experience of a deepfake involves an embodiment relation to a screen, whereby the screen becomes transparent, and the user's attention shifts to what it displays. This, then, facilitates a hermeneutic relation to the digital media content, which acts as a window to one's environment. From this picture, a deepfake can be understood as distorting the windowpane. This understanding of human-media interaction supports the analyses of figures like Rini and Fallis, who consider how deepfakes may intervene with peoples' capacity for obtaining knowledge from digital media. Although this picture certainly captures a great extent of the experience of a deepfake, it is not complete, and consequently not sufficient in producing an understanding of the infopocalypse.

One thing that is not captured by the analysis so far is the direct relation one can have to a deepfake, where it does not act as a representation of the surrounding environment but is the direct focus of the user's awareness. In this type of *alterity* relation, the deepfake is presented to the computer user as though it represented another human being, requiring direct attention (See Wellner, 2014). This sort of relation between human beings and deepfakes is made explicit in demonstrations of deepfake technology, such as on the webpage www.thispersondoesnotexist.com, where upon visiting the computer user is shown an image that depicts a person, and consequently made to reflect on the fact that this person does not actually exist (as captured by the name of the page). The uneasiness in making sense of this image, shows the curious character of the alterity relation between human beings and deepfakes. When faced directly with deepfake technology, the computer user does not merely read the deepfake to learn about their environment, but moreover feels an emotional reaction. Going back to the notion of the infopocalypse, this emotional reaction

appears to be part of what sets misinformation spread through deepfake technology apart from misinformation spread through other means.

Moreover, the existence of deepfake technologies affects human experiences even when they are not directly present. This is to say that as deepfakes increasingly feature in daily experiences, they will also increasingly become part of the experiential *background* of human beings. The fact that there exist deepfakes will alter the experiences and practices of human beings, whether or not they are present in the current moment. This description lines up with Stacey O'Neal Irwin's analysis of digital media as a technological *texture*. Irwin argues that digital media technologies are so ubiquitous in the experience of contemporary human experience, that they are best understood as texturing their perception (Irwin, 2016, pp. 3-8). This background relation seems to be captured in warnings of the infopocalypse that claim that deepfakes will lead to the "distortion of reality," as deepfakes will have a lasting impact not only on singular experiences, but also on what it means to perceive one's surroundings (Ovadya, 2018).

3.3 - Shortcomings

Undoubtedly, Postphenomenology is conducive to an insightful understanding of technologies in use. At the same time, the postphenomenological focus on the relation between individual users and individual technologies has been perceived as limiting Postphenomenology's capacity for addressing the importance of social and cultural dimensions of use (Coeckelbergh, 2018a, p. 5). Postphenomenologist Robert Rosenberger writes: "Where postphenomenology is good at describing the nuances of the relationships developed between an individual user and a technology, [...] it is not as readily prepared to talk about the effects of these on the larger world" (Rosenberger, 2014, p. 372). As such, the analysis of this chapter appears insufficient in providing a proper understanding of the Infopocalypse, where the broader social and cultural effects of human-deepfake relations are at stake.

In recent years, postphenomenologists have started turning their attention to this apparent shortcoming, and have focused explicitly on human-technology relations in regard to social and cultural environments. Peter-Paul Verbeek, for example, draws upon Ihde's notion of *material hermeneutics*, whereby material artifacts are understood as organizing interpretational frameworks, to develop a "political hermeneutics of technology" (Verbeek, 2020). Building upon this notion, Verbeek suggests that technologies shape power relations and political interactions, as well as the character of political issues. Another recent development in Postphenomenology can be found in the work of Olya Kudina, who has attempted to connect human-technology relations and the 'sociocultural world' through the development of a "leminscate principle" (Kudina, 2021). According to this principle, human beings "necessarily engage

their moral histories to comprehend new technologies and fit them in daily practices" (Kudina, 2021, p. 233).

The following two chapters of this thesis may be considered as existing within the same research program as the mentioned works, insofar as it likewise attempts to address the connection between human-technology relations and their social and cultural contexts. However, while the mentioned works base their analyses on material hermeneutics, I introduce and draw upon a Wittgensteinian perspective on technology use. This perspective, I suggest, is particularly capable of describing the dynamic interaction between the use of an artifact and the 'grammar' of its social and cultural context. As such, it is not only effective in accompanying the postphenomenological analysis of deepfakes presented in this chapter, but may also help to overcome Postphenomenology's apparent shortcoming. While the insights produced in this thesis are most certainly related to the works of Verbeek and Kudina, exploring that relation is beyond the scope of this thesis.

Conclusion

This chapter introduced a Postphenomenological analysis of deepfakes. I have demonstrated that the current state of academic discourse regarding the societal impact of deepfake technology is largely focused on knowledge and information. I argued that this focus is misleading, as the way in which human beings relate to deepfakes is more complex than just interpreting information. Deepfakes present themselves as others and can become part of a person's perceptual background.

CHAPTER 3   DEEPFAKE-GAMES

In the previous chapter, I offered a postphenomenological analysis of the infopocalypse. I concluded that this framework generates the important insights that the experience of a deepfake goes beyond merely obtaining information, but appears to be lacking in its limited ability to address the importance of social and cultural contexts. This chapter will attempt to contribute to the postphenomenological analysis by introducing a Wittgensteinian perspective, and in particular the notions of *language-games* and *forms of life*. The main strength of the Wittgensteinian understanding of technology use presented in this chapter lies in its conceptualization of use as embedded in social and cultural contexts, allowing it to make sense of deepfake technology's cascading impact. To study such cascading impact, the chapter takes three possible scenarios involving deepfakes from a recent STOA report and uses Wittgensteinian notions to make sense of them.

Part 1 - Wittgenstein and Technology

Ludwig Wittgenstein is widely regarded as one of the most influential philosophers of the 20th century, having affected virtually every discipline of contemporary philosophical scholarship. In recent years, Wittgenstein's views have gained modest attention in the philosophy of technology to produce a Wittgensteinian understanding of technologies in use (see Coeckelbergh, 2018b, Coeckelbergh & Funk, 2018). This section will introduce that understanding, after first presenting key concepts from Wittgenstein's later philosophy.

1.1 - Wittgenstein

Wittgenstein's philosophical beliefs did not stay the same throughout his entire career, but shifted significantly, causing scholars to now speak of an 'early' and 'later' Wittgenstein. To properly understand the philosophical perspective of the later Wittgenstein, it is important to place it in the context of his earlier writing. During the early parts of his career, Wittgenstein concerned himself mainly with questions of logic and the nature of representation. The ideas conceived in this period culminated in the book Tractatus Logico-Philosophicus, wherein it is argued that thoughts and propositions should be understood as pictures of facts (Biletzki & Mater, 2002). As such, to understand the meaning of any given word or sentence, one should consider the facts represented by it, and the logic used to do so.

Although Wittgenstein's writings of this time were, and still are, incredibly influential, they have come under immense criticism, most notably from Wittgenstein himself (Kienzler, 2017). After an extended hiatus from philosophy, Wittgenstein returned with a perspective radically divorced from his earlier writing. Rather than understanding meaning in terms of representation, Wittgenstein shifted his attention to use. He writes: "For a large class of cases of the employment of the word 'meaning'—though not for all—this word can be explained in this way: the meaning of a word is its use in the language" (Wittgenstein, 1953, § 43). This entails that the meaning of a word or sentence is determined not by the facts it represents, but primarily by how it is used in a given context.

Language-Games and Forms of Life

From his interest in the use of language, Wittgenstein introduces the notion of language-games, to bring to light the different functions words can have. Although the idea of language-games is central in most of his later philosophy, Wittgenstein never provides his audience with a proper definition, instead opting to present the notion mainly through examples. These definitions range from quite primitive acts such as giving orders, describing an object and reporting an event, to more complex behaviors like forming a hypothesis, acting in a play and cracking a joke (Wittgenstein, 1953, § 23). What these examples make clear is that, for Wittgenstein, the notion of language-games has a heuristic purpose. The main idea that it is intended to convey is that everyday language involves practices of employing expressions in specific ways, as though governed by the rules of a (language-)game (Whiting, 2012, p. 422).

The examples also show that language-games involve more than just language, and that linguistic expressions are interrelated with other non-linguistic activities in human life. Acting in a play, to concentrate on one of the examples, consists of more than just speaking. It involves body-language, stage presence and a whole set of activities that go into executing a performance. The use (and by extension the meaning) of any word or sentence, then, depends on its role in a context of linguistic expressions and non-linguistic activities alike. The notion of language-games indicates that language can never be isolated from its context, and that understanding language requires understanding individual linguistic expressions in their respective contexts.

Another notion that is introduced in Wittgenstein's later thinking is that of *forms of life*, which, despite being less prominently addressed than language-games, plays an important role in Wittgenstein's understanding of language. As the meanings of individual words are to be understood as embedded in language-games, such language-games are themselves embedded in even larger wholes that Wittgenstein refers to as forms of life. Wittgenstein writes that forms of life are not themselves part of spoken language, but rather that "the speaking of language is part of an activity, or a form of life" (Wittgenstein, 1953, § 23).

As such, forms of life consist of shared ways of acting that together form a foundation of meaning (Moyal-Sharrock, 2015, p. 25).

An important consequence of understanding use of language as embedded in forms of life is that it allows for a cultural appreciation of meaning. Insofar as meanings depend on use, language-games, and ultimately forms of life, meaning is necessarily cultural. This consequence of Wittgenstein's writing is what causes prominent interpreter Peter Hacker to speak of Wittgenstein's perspective as an anthropological or ethnological conception of language, writing that language is "an open-ended series of interlocking language-games constituting a form of life or way of living (a culture)" (Hacker, 2015, p. 1). The use of language is, thus, necessarily cultural, insofar as it exists within, and is given meaning through, particular language-games and forms of life.[3]

Grammar and Certainty

A notion that connects language-games and forms of life is that of *grammar*. Simply put, grammar refers to the set of rules that govern everyday use of language. For Wittgenstein, grammar extends far beyond the traditional understanding of grammar as syntax, to include the social and cultural dimensions of the conditions of language use. To emphasize this point, Wittgenstein distinguishes between 'surface grammar' and 'depth grammar,' whereby the former roughly equates to syntax, and the latter refers to the network of rules which determine what expressions do or do not make sense (Wittgenstein, 1953, § 664).

It is through its depth grammar that use of language, language-games and forms of life are interconnected. The grammar of a language-game determines which individual uses of linguistic expressions are appropriate and meaningful within this language-game, and in turn, the grammar of a form of life determines which language-games are playable. Importantly, while grammar extends beyond the use of individual words and phrases, it should not be understood as existing outside of this use. Wittgenstein scholar Danielle Moyal-Sharrock writes: "Rules of grammar are simply expressions of the norms of sense that grow out of, and in tandem with, our natural ways of acting and our socio-cultural practices. Grammar does not generate language; nor does it exist independently of language or action; it is embedded and enacted in what we say and do" (Moyal-Sharrock, 2022, p. 13).

Wittgenstein's notion of grammar may also be understood in relation to the discussions in *On Certainty*, where Wittgenstein expands upon his view of language (Wittgenstein, 1969). Here, Wittgenstein looks specifically at basic beliefs that appear unquestionable in everyday language, such as "I know that I

---

[3] Peter Hacker's cultural interpretation of forms of life is controversial among Wittgenstein scholars. See (Moyal-Sharrock, 2015) for a discussion on various interpretations of Wittgenstein's use of the term 'forms of life.'

am a human being" or "my body has never disappeared" (Wittgenstein, 1969, § 4, 101). These kinds of beliefs, referred to as *certainties*, reside at the heart of language-games (Wittgenstein, 1969, § 403). Certainties are crucial in that they take on the function of *hinges*; they allow moves to be made within a language-game, while itself remaining unmoved, and may thus be understood as part of the grammar of the language-game (Wittgenstein, 1969, § 301, Moyal-Sharrock, 2005, p. 72). At the same time, certainties should not be understood as certain propositions, but rather as a kind of thoughtless, embodied acting. Wittgenstein writes: "Giving grounds, however, justifying the evidence, comes to and end; —but the end is not certain 'propositions' striking us immediately as true, i.e., it is not a kind of seeing on our part; it is our acting, which lies at the bottom of the language game" (Wittgenstein, 1969, § 204).[4]

Certainty, then, is best understood as a doxastic attitude, or blind trust, expressed in the ungrounded way of acting in all ordinary gestures and activities (Wittgenstein, 1969, § 110, Moyal-Sharrock, 2005, p. 78). Someone does not possess the certainty that their chair is made of a solid material through conscious thought, but rather expresses this certainty through their acting; the confidence with which they sit down on their chair every single day. This sort of everyday know-how is what forms the grammar of language-games and forms of life, and ultimately shapes the meanings of words and sentences.

A Holistic Picture

The view on language that results from the described notions is a holistic picture. It considers meaning as determined by use, which must in turn be understood as embedded in its particular context. This context consists of language games, certainties, and forms of life. Importantly, these notions do not refer to things that exist *outside* of language use; they are constituted by use while giving meaning to use. The various levels of meaning, then, can never be understood separately from one another, calling for a truly holistic picture (Coeckelbergh & Funk, 2018, p. 187). This picture may be illustrated as such:[5]

---

[4] It should be noted that some Wittgenstein scholars have interpreted certainties as propositions. I follow Moyal-Sharrock in rejecting this interpretation (Moyal-Sharrock, 2005).

[5] This figure is based on a similar figure from Mark Coeckelbergh's 'Technology Games' (2018b), which includes the layer of activities in between use and language-games. According to my interpretation activities are included at the level of language-games, which leads to the apparent alteration.
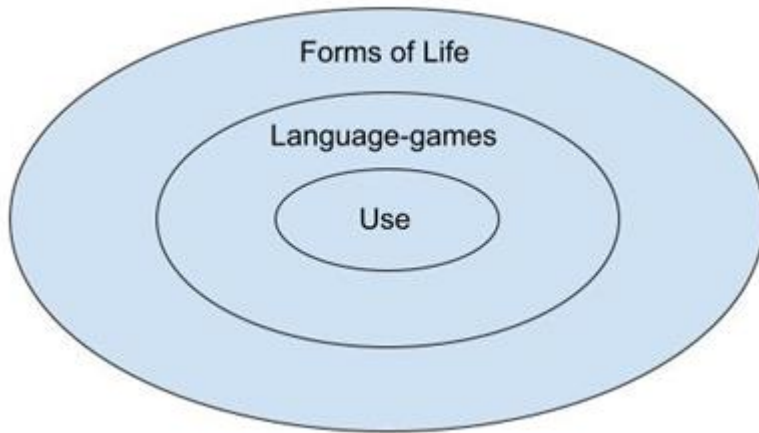
1.2 - Technologies in Use

Although Wittgenstein himself never placed his focus directly onto technology, his perspective on language may still be useful for thinking in this domain. Only a handful philosophers of technology, however, have expressed interest in applying Wittgensteinian thoughts to the Philosophy of Technology (cf. Winner, 1983, Coeckelbergh, 2018b, Coeckelbergh & Funk, 2018, Hermann, unpublished). One notable attempt is that of Mark Coeckelbergh, who coins the term Technology Games in applying Wittgenstein's holistic understanding of language use to technology use. I suggest that technology-games offer a useful vocabulary to apply the Wittgensteinian perspective to technology, and in particular to the project of understanding the societal impact of deepfake technology. Hereunder, I will briefly describe how Coeckelbergh appropriates Wittgenstein's perspective. In the following section, I will then apply this perspective to the case at hand, illustrating how interactions with deepfakes are embedded in language-games and forms of life.

Technology-games and Forms of Life

A Wittgensteinian understanding of technology is an understanding of technologies in use. Just as the meaning of language is determined by its use, so is the meaning of technology determined by how technologies are used in everyday life. Like in Postphenomenology, then, achieving a philosophical understanding of technology must be grounded in an analysis of individual human-technology relations.

Such human-technology relations, however, do not exist in isolation.[6] As the use of language is embedded in a pre-existing social and cultural context, so is the use of technology. Coeckelbergh writes: "How we use technologies is shaped by the games and forms of life that are already in place "before" we use them" (Coeckelbergh, 2018b, p. 1511). What are these games and forms of life then?[7] And how can we identify them?

How a person uses and interacts with a given technology is not predetermined. As acknowledged by the postphenomenological concept of *multistability*, "a single technology can be understood in multiple ways, taken up in many contexts and employed for various purposes" (Rosenberger, 2014, p. 376). What determines the way a technological artifact is used then, depends not only on its material features, but moreover on the context that it is used in. This context determines which uses are appropriate and which uses are not. This context, then, may be understood as a *technology-game*, which provides a grammar that constitutes the basis for meaningful use (Coeckelbergh, 2018b, p. 1512).[8] Human beings do not approach technologies without preconceived notions of what is appropriate use, and these conceptions are captured by the notion of technology-games.

Like all language-games, technology-games are connected and interlocked with other language-games and certainties, and are ultimately embedded in forms of life. These forms of life, which can be broadly understood as cultures, not only give meaning to technology games, but moreover provide those games with a background of certainties and know-how. On the other hand, it is important to acknowledge again that forms of life do not exist prior to language-games and certainties, and are in fact constituted by language-games and certainties as much as they constitute them. This also means that when technology use becomes part of shared activities, it inevitably becomes part of a form of life as well. While forms of life provide a structure for technology-games, technologies may also structure human activities, ultimately shaping our forms of life. Philosopher of technology Langdon Winner states that "as [technologies] become woven into the texture of everyday existence, the devices, techniques, and systems we adopt shed their tool-like qualities to become part of our very humanity" (Winner, 1983, p. 12).

Part 2 – Three Scenarios

---

[6] It would be deceptive to say that postphenomenologists assume human-technology relations to exist in isolation. Chapter 4 section 2.2 will discuss how Postphenomenology makes sense of social and cultural embeddedness, and how this may be related to Wittgensteinian concepts.

[7] I am not using the word "games" in a literal sense, but as a shorthand for language-games that the use of artifacts is embedded in.

[8] Technology-games are simply language-games that involve technologies (Coeckelbergh, 2018b, p. 1513). This means that technology-games do not refer to something substantially different from any other language-game, but that the term may still be heuristically useful in directing attention towards specific artifacts.

A recent study on the prospect of regulating deepfake technology, commissioned by The Panel for the Future of Science and Technology (STOA), discusses risks, harms, and impacts of deepfakes (Van Huijstee et al., 2021). One of the main claims in this study is that the impact of a deepfake is "not limited to a single type or category of risk, but rather to a combination of cascading impacts at different levels" (Van Huijstee et al., 2021, p. 34). What this means is that the impact of deepfakes extends beyond the individual level, to potentially harming groups and organizations, and finally affecting societies as a whole. Crucial for understanding long-term and wide-scale impact of deepfake technology, then, is understanding the interactions between the different levels of perceived impact.

Returning to the topic of this thesis, then, it becomes clear that understanding the Infopocalypse involves an appreciation of the interaction between individual cases of human-deepfake relations and their anticipated long-term societal outcome. To explore this interaction, I propose to use a Wittgensteinian understanding of technology, whereby the various levels of cascading impact can be understood through Wittgensteinian notions. The levels of the individual, the group, and the society as a whole can be roughly equated to the level of technology use, technology-game, and the technological form of life.[9] Understanding the holistic relationship between these levels allows for a philosophically grounded appreciation of the cascading impact, where human-technology relations are both embedded in, and constitute, their social and cultural contexts.

The seventh chapter of the Panel for the Future of Science and Technology report for the regulation of deepfake technology describes three scenarios to illustrate the cascading impact of deepfakes (Van Huijstee et al., 2021, pp. 48-57). These three scenarios involve: a piece of manipulated court evidence, a deepfake pornographic video, and a false political statement. In this section, I will briefly introduce each of these scenarios and apply the notions of technology-games and forms of life to elucidate how social and cultural contexts give meaning to human-deepfake relations.

Manipulated court evidence

The first scenario that I will discuss involves a synthetic audio tape of a telephone conversation, created using deepfake technology. In this conversation, a popular politician is heard participating in bribery. The deepfake is convincing to the public, and the politician ends up in a criminal trial for corruption. This scenario is not entirely fictional, as tampered recordings labeling individuals as guilty have figured in actual

---

[9] The levels of the individual, group, and society are most definitely not identical to that of use, language-game, and forms of life. I merely make the comparison as a starting point for the application of the Wittgensteinian perspective to particular cases of deepfake-use.

court cases.[10] Deepfake technology is not only developing to create more realistic looking deepfakes, but moreover becoming more readily available to large groups of people, making this scenario likely to occur at an increasing rate in the future.

In the courtroom, the deepfake in question is introduced to a new context, where it begets significance for various actors at various levels. At the individual level, the deepfake has clear implications for the politician under trial, potentially leading to them ending up in prison. This impact is due to how the deepfake was used (to misrepresent the actions of the politician) and for what purpose (to make sure the politician is removed from their position of power). At the same time, however, the implications of this deepfake are also determined by the rules of the courtroom. The specific ways in which digital materials get authorized, for example, are crucial for the deepfake to achieve its intended purpose. From a Wittgensteinian perspective, then, it can be argued that the meaning of this deepfake is embedded in a courtroom-language-game, with rules and ways of acting that have been in place for a long time.

There are two important insights that can be drawn from this Wittgensteinian perspective. First, it shows that the grammar of the courtroom is not limited to explicitly stated rules, but also includes implicit rules. Official proceedings, standards of authentication and professional responsibilities are certainly part of the language-game and provide meaning to this particular deepfake. But so are implicit rules that determine what sort of things are appropriate and meaningful, and that guide the various actors' expectations of how matters are handled. Besides this, the Wittgensteinian perspective highlights that the context provided by the courtroom is dynamic in nature, as are all language-games. The courtroom-game is not an unchanging static background, but rather an active network of meanings, norms and practices that ultimately consists of individual instances of use of language and technology (including, but not limited to, deepfakes).

Understanding how the deepfake is shaped by its context also allows us to explore how this context may be affected by the deepfake. It becomes apparent that the deepfake challenges both explicit and implicit aspects of the language-game. Standards of authentication of digital materials are challenged as they do not seem up to par with the improving quality of deepfakes (see Pfefferkorn, 2019). At the same time, the possibility of any given piece of digital material being deepfaked challenges implicit rules. One intriguing example of this is captured by the claim that deepfakes give way to the so-called 'liar's dividend,' where a defendant claims that evidence is fake or constructed. Such claims have already become more prominent in recent years with cries of 'fake news' becoming more common (Schiff, Schiff & Bueno, 2021). However, deepfakes might provide the liar's dividend with more legitimacy in courtroom contexts, as the authenticity of digital material will become harder to verify.

---

[10] See, e.g., (*Batiashvili v Georgia*).

Finally, as the STOA report presents, deepfakes will have a wider impact that goes beyond individual court cases. As the impact on individuals and explicit and implicit courtroom proceedings cascades, courts and jurors may start "questioning the authenticity of every piece of evidence" (Huijstee et al., 2021, p. 55). As such, the culture or form of life that endows existing language-games in the courtroom with meaning and appropriateness may be challenged by the existence of deepfakes. This culture, grounded in (certainties like) a trust in information and a presumption of innocence, is importantly manifested through technology-games involving digital material. When such technology-games are altered by the introduction of deepfakes, the forms of life manifested through them transform also.

Deepfake pornography

Another scenario considers the use of deepfake technology for creating false pornographic material. A female investigative journalist finds herself the subject of a fake pornographic video, using her face and the body of someone else, intended to harm and discredit her and her work. The video is convincing to the public and the journalist has to deal with the consequences. Like the previous scenario, this scenario is not fictional. Indian journalist Rana Ayyub and Brazilian journalist Patricia Campos Mello were both victims of deepfake pornography intended to silence them (Ayyub, 2018, Posetti et al., 2021). Again, the immediate impact on the individual is relatively clear. The deepfake was created with the intention to harm the reputation of the journalist, and that is most likely how it will impact the journalist in question. Beyond this, being scripted into fake pornography without their consent undermines the journalists' integrity and human agency, and can cause feelings of embarrassment and shame, possibly resulting in serious mental harm.

To understand the broader impact, the deepfake figuring in this scenario may also be understood as existing within a language-game. Mark Coeckelbergh introduces the notion of gender-games to address how the use of technologies is embedded in gender norms and cultures (Coeckelbergh, 2019, p. 31). A gender-game, then, is a language-game involving the meanings of sex and gender that shapes how words and technologies are used and developed. It becomes clear that the deepfake in the scenario in question is shaped by a gender-game when one considers how the scenario would have played out if the victim were male. It is entirely unlikely that the deepfake would have the same consequences, showing that social and cultural hierarchies involving gender shape the use of deepfakes.

While gender-games give meaning to how words and technologies are used, they are at the same time constituted thereby. As such, deepfakes are not only gendered, but also gendering, meaning that they impact the meanings of sex and gender over time. By allowing users to harm women through deepfake technology, deepfakes are likely to strengthen the grammar of sex and gender as a weapon against women.

Rae Langton and Caroline West argue that pornography, by introducing certain presuppositions about women (e.g., that 'no' actually means 'yes'), reinforces sexual language-games that silence and subordinate women (Langton & West, 1999, p. 11). Such games are themselves, then, embedded in an existing culture of sex and gender, or, a particular form of life.

While deepfakes appear to strengthen them, it is possible that technologies which provide novel uses may challenge existing language-games and forms of life involving the meanings of sex and gender. The oral contraceptive pill, for example, has been argued to promote women's sexual autonomy by separating sex from reproduction (Ketting, 2000). Similarly, one may look at the recent advances made in the development of artificial wombs, which could potentially impact gender roles in potentially positive ways (Hermann, 2019). What remains crucial is that the development and use of technological artifacts cannot be understood as separate from its social and cultural contexts, wherein the meanings of sex and gender plays an important role. Understanding the Infopocalypse, then, must include an appreciation of the genderedness and gendering qualities of deepfakes.

False political statement

Finally, the STOA report describes a scenario wherein deepfake technology is used to undermine trust in European politics. A deepfake is created that depicts several EU health ministers confidentially deliberating to withhold vaccine supplies. The deepfake is spread on social media platforms and quickly goes viral, as many social media users believe the video to be authentic. What makes this scenario unique from the previous two is that in it the deepfake was not created and used to harm any particular actor. Whereas the professional and private lives of the individual politicians depicted in the deepfake are certainly affected, the intent extends beyond harming the reputation of these individuals to harming the reputation of the political organization they are a part of. As such, understanding this deepfake necessarily involves the broader context of this organization, and how trust in such an organization is established.

Again, this broader context may be understood in terms of language-games and forms of life. As such, actions that express trust in the content one comes across on social media may be regarded as particular moves in a language-game, itself given meaning by what it means to trust something in a form of life. What becomes clear from this analysis is that trusting a piece of digital media content is not an isolated act based solely on features of that content, but rather is shaped by its social and cultural context. This is in line with Anton Vedder's work on digital spread of misinformation, wherein he states that assigning trust is based on "an intricate complex of backgrounds of all kinds of manifest or latent recognition procedures for persons and organizations, traditions of reputations, and usage (Vedder, 2001, p. 4). Moreover, empirical research has indicated that the likelihood of a person trusting internet information

is in part dependent on their socio-cultural background, further justifying a Wittgensteinian perspective (Strang, 2018).

Besides being embedded in language-games and forms of life involving trust, deepfakes will likely impact those language-games and forms of life. Robert Chesney and Danielle Keats Citron argue that deepfakes will lead to "eroding trust in institutions" precisely because of the kind of scenario described here (Chesney & Citron, 2019, p. 1779). This erosion of trust can be accounted for by a Wittgensteinian perspective. When trust is no longer the default attitude towards particular institutions, the grammar that governs whether acts exhibiting trust are appropriate transforms. David Lewis aptly suggests that what is appropriate in a given language-game "[tends] to evolve in such a way as is required in order to make whatever occurs count as correct play" (Lewis, 1979, p. 347). This is to say, when people are given reasons not to trust, as deepfakes appear to do, the act of trusting itself acquires a new meaning. Crucial for understanding the infopocalypse, then, is understanding the social and cultural contexts that embed trust in deepfakes.

Conclusion

The Wittgensteinian perspective on technologies in use, introduced in this chapter, provides and understanding of the use of technological artifacts as embedded in pre-existing social and cultural context. I have applied this perspective to three scenarios involving deepfakes: manipulated court evidence; deepfake pornography; and a false political statement. By doing so, I have attempted to overcome Postphenomenology's apparent shortcomings regarding addressing the importance of social and cultural contexts. The Wittgensteinian perspective can address social and cultural contexts by understanding them as language-games and forms of life, which stand in a holistic relationship with technology use. The three scenarios moreover lead to the important insights that the contexts that embed technology use are typically implicit and dynamic, and that understanding them is essential for understanding the Infopocalypse.

CHAPTER 4   DEEPFAKES IN USE

This chapter will draw upon insights from chapters two and three to reflect on the main project of this thesis. As such, it will provide a conclusive analysis of the 'infopocalypse' as discussed in chapter one, drawing from both Postphenomenology and Wittgenstein to assess in which ways the notion is misleading. I conclude this section by stating that the notion of the infopocalypse is misleading in that it overemphasizes the role of information, wrongly supposes an apocalypse-like scenario, and generalizes the various kinds of impact that deepfake technologies have. Hereafter, I turn towards the philosophical ideas and concepts introduced throughout the thesis, and reflect on the questions as to whether they are compatible and how they might complement each other. I argue Wittgenstein is compatible with Postphenomenology with regard to both its pragmatic and phenomenological commitments, and that the Wittgensteinian understanding of technology use as embedded in language games and forms of life can complement the Postphenomenological framework by emphasizing the social and cultural contexts of human-technology-world relations, and may provide nuance to notions such as macroperception, multistability and background.

Part 1 - Rethinking the Infopocalypse

In this section, I revisit the popular understanding of the societal impact of deepfake technology as presented in chapter one under the label *Infopocalypse*. I suggested that this notion, although fueled by legitimate concerns, is lacking with regard to its philosophical underpinnings. Having provided concepts and tools in chapters two and three that, optimistically, provide the basis for a proper philosophical understanding, I now revisit the Infopocalypse. Using insights from Postphenomenology and Wittgenstein, I illuminate three ways in which the notion is misleading. Firstly, the infopocalypse overemphasizes the importance of information, in turn neglecting other aspects of the experience of deepfaked media. Secondly, the infopocalypse implies that the impact of deepfake technology can be understood as singular, as such crudely generalizing the many different types of impact it has at different levels. And finally, the infopocalypse suggests an 'apocalyptic' scenario, whereby in a society post deepfake-saturation concepts like trust or truth have no meaning. This neglects the dynamic character of society.

<u>Information</u>

The term 'infopocalypse' is a portmanteau of the words 'information' and 'apocalypse.' This leads to the assumption that the societal impact of deepfakes is mostly, if not entirely, related to information, and thus proper analysis should be information-centric. This is in line with some of the academic literature (see Rini, 2019, Fallis, 2020). A closer look at concerns in popular media, however, shows that deepfakes may affect human beings in ways more complex. Through introducing Postphenomenology, I suggest that the relationship between human beings and deepfakes is more complex and nuanced than that implied by the infopocalypse. The human subject does not simply encounter a deepfake as containing isolable (mis)information, but rather sustains complex human-deepfake-world relationships.

Solely considering information leads to a neglect of how deepfakes may affect people that are not actively engaging with them, which is implied by much of the popular media surrounding the infopocalypse. The very possibility of deepfakes being present in one's surroundings, even if not directly, constitutes a background relationship, whereby human experience and interaction occurs in the context of deepfake technology. Understanding this aspect of human-deepfake relations is crucial for a proper appreciation of the long-term societal impact of deepfakes, and requires a scope that goes beyond human-media interaction as ascertaining information.

Generalization

If it is at all possible to legitimately speak of an Infopocalypse, it must be clear that there is not 'one' Infopocalypse. The impact of deepfake technology is noticed in different contexts and at different levels. As demonstrated in chapther three, deepfakes take on distinct meanings in the contexts of courtrooms, pornography and politics, and have disparaging impacts at different levels. All of the ways in which deepfakes impact human beings ultimately cascade into something that could be called the infopocalypse, but merely considering this larger level is misleading to the nuances of human-technology relations. Through introducing a Wittgensteinian perspective on technology use in chapter three, I argue that the individual, social and cultural impact of deepfakes should be understood as holistically interrelated, whereby individual use is both embedded in and constitutive of social and cultural contexts. Adopting this perspective renders the popular understanding of the infopocalypse as simple and deceptive. Much like it is impossible to isolate the impact of a deepfake from its particular context, so too is it impossible to describe that impact in a way that is applicable to all possible contexts.

Apocalypse

Finally, the notion of 'infopocalypse' is misleading in suggesting that deepfake technology will result in an apocalyptic scenario, as such announcing the *death of truth* according to some (Schwartz, 2018). This conception of deepfake technology as signifying an apocalypse, whether with regard to truth, trust, democracy or something else, grossly misunderstands the dynamic nature of cultures and social practices. By understanding social practices and cultures in terms of language-games and forms of life, they appear as ever changing as sustained by shared practices. Whereas it is possible, even likely, that deepfake technologies will facilitate practices that challenge existing norms surrounding reliability and trust, this does not mean that those aspects of contemporary culture will lose their meaning entirely. In fact, those norms, as well as the deepfakes themselves, are already part of a dynamic web of practices and cultural beliefs that change continuously. The impact of deepfakes, then, is by no means inconsequential or uninteresting, but at the same time not apocalyptic. Deepfakes do not eliminate norms surrounding reliability and trust, but are given meaning by them and may, in some cases, challenge and disrupt them from within.

The new picture

The insights provided by Postphenomenology and Wittgenstein lead to a rethinking of the Infopocalypse. The impact of deepfake technology is not to be understood as relating solely to information, cannot be described in abstract terms, and should not be regarded as an apocalyptic scenario. Rather, deepfakes should be understood as mediating all aspects of the ways human beings experience and interact with digital media in complex ways, and as necessarily embedded in dynamic social and cultural contexts. The new picture of the Infopocalypse, then, refers to a complex, multifarious and ongoing process of social transformation.

Part 2 - Wittgenstein in Postphenomenology

In this section I reflect on the philosophical view presented throughout this thesis project. Specifically, I consider the decision to adopt a Wittgensteinian understanding of technology use from a postphenomenological perspective.

2.1 - Wittgenstein's views are compatible with Postphenomenology

In the second chapter of this thesis, I described Postphenomenology as defined by two core commitmentments. In the first place, Postphenomenology builds on traditional Phenomenology by

considering the human being and their experienced world as co-constitutive. However, where traditional Phenomenology of technology has a tendency to discuss technology in abstract and monolithic terms, Postphenomenology draws upon pragmatic ontology to shift the discussion towards individual artifacts and uses thereof. As such, determining whether the views of the later Wittgenstein are compatible with postphenomenology requires consideration of Wittgenstein's philosophy in relation to these two core commitments. I will do so by first considering the pragmatic commitment, and hereafter discussing the phenomenological commitment.

The pragmatic commitment

The relationship between the later Wittgenstein and Pragmatism has been subject of ongoing debate among Wittgenstein scholars for decades, spearheaded by Hilary Putnam's 1995 lecture 'Was Wittgenstein a Pragmatist?' (Putnam, 1995). In this lecture it is argued that, although Wittgenstein cannot be said to have been a pragmatist in the strict sense, he shares with Pragmatism a central emphasis on what Putnam calls "the primacy of practice" (Putnam, 1995, p. 52).[11] What this means is that Wittgenstein resists the idea that language-games and forms of life can be described 'from-the-outside-in,' and instead always require an appreciation of the use of words and practices within said language-games and forms of life. As such, Wittgenstein's views, like those of pragmatists such as John Dewey and William James, prioritize the importance of practice over theory. It is precisely this characteristic of pragmatic philosophy that Ihde claims inspired him to move past traditional Phenomenology in conceiving Postphenomenology (Ihde, 2009).

Putnam's discussion does not by any means settle the question as to what exactly is the relationship between Wittgenstein's later philosophy and Pragmatism, and many have challenged it for various reasons (cf. Hensley, 2012, Moyal-Sharrock, 2017).[12] What it does demonstrate, however, is that there is a clear way to interpret and employ the Wittgensteinian notions of language-games and forms of life that is compatible with pragmatic ontology. The analysis offered in the previous chapter does this by looking at 'deepfake-games' as consisting of individual scenarios involving actual use of deepfakes, rather than being defined by some external conditions. In general, the Wittgensteinian understanding of technology use appears to be pragmatic in the same way that Postphenomenology is, by electing to ground analysis of

---

[11] The notion of 'the primacy of practice' has also been used by Anthony Rudd to discuss the comparisons between the philosophy of the later Wittgenstein and Heidegger (Rudd, 2005).
[12] Judy Hensley's article 'Who's Calling Wittgenstein a Pragmatist' was part of a 2012 symposium on Wittgenstein and Pragmatism organized by the European Journal of Pragmatism and American Philosophy.

technologies primarily in empirical input provided by actual cases rather than the theoretical application of abstract philosophical notions.

This pragmatic outlook moreover sets the analysis provided in this thesis apart from the existing literature discussed at the start of chapter two. Don Fallis' appreciation of the Infopocalypse, for example, bases its argument on a general theory of information carrying that was developed entirely separate from consideration of actual interactions with deepfakes (Fallis, 2020). This approach appears untenable when considering the technological shaping of experiences and actions, as per Postphenomenology, as well as language-games and forms of life, as per Wittgenstein. In discussing the methodology of Postphenomenology, Rosenberger and Verbeek make this point by presenting the following example: "Rather than analyzing the imaging devices on the Mars orbiter in terms of existing theories of perception or knowledge development, [Postphenomenology] investigates how this device establishes a new kind of sensorial relationship between scientists and the planet Mars, which is a new basis for knowledge development" (Rosenberger & Verbeek, 2015, pp. 30-31). In the same way, this thesis does not study deepfakes in terms of existing theories of knowledge and information, but instead investigates how deepfakes establish new kinds of human-world relations and language-games as a basis for new kinds of knowledge and information practices.

The phenomenological commitment

While the connection between the later Wittgenstein and Pragmatism is somewhat obvious and was even recognized by Wittgenstein himself (see Wittgenstein, 1969, § 422), this is not the case for Wittgenstein's relationship with Phenomenology. Throughout his career, Wittgenstein seldom interacted with the works of prominent figures in Phenomenology like Edmund Husserl, Maurice Merleau-Ponty and Martin Heidegger, despite them finding popularity around the same time as Wittgenstein himself in the first half of the 20th century. Even after the publications of Wittgenstein's final works impacted numerous aisles of philosophical study in the years after his death, phenomenologists took little notice. I believe that this lack of interaction can be largely attributed to the analytic/continental split that places Wittgenstein strictly in the camp of analytic philosophy and Phenomenology on the side of continental philosophy, with little to no communication between them.

As the analytic/continental split has steadily lost its influence over past decades, an increasing number of scholars have started connecting the views of Wittgenstein to those of the phenomenologists mentioned, perhaps most thoroughly in Nicholas F. Gier's book *Wittgenstein and Phenomenology: A Comparative Study of the Later Wittgenstein, Husserl, Heidegger, and Merleau-Ponty* (Gier, 1981). Although it is far beyond the scope of the present thesis to discuss this discourse in detail, it is important to

point out that there do exist significant agreements between the later Wittgenstein and the phenomenological tradition that warrants considering them as participating in a joint program. Wittgenstein's philosophy mirrors Phenomenology not only in rejecting both Cartesian Dualism and Logical Positivism, but it has moreover been argued that Wittgenstein's positive account offered in response is similar also.

Søren Overgaard, for example, points out that Wittgenstein's discussions of pain and subjectivity in the *Philosophical Investigations* (in particular § 281 until § 289) lend itself to an account of human beings as necessarily *living human beings* (Overgaard, 2006, p. 63). This is to say that, according to Wittgenstein, human beings are not merely 'minds' that can be isolated from body and world, but are rather living human beings that are "already bodily and worldly" (Overgaard, 2006, p. 67). The resulting picture is strikingly similar to that of Phenomenology, which likewise understands the human being and their environment as inseparable from each other, or, to put it in phenomenological terms, *co-constitutive*. It is thus not only due to his pragmatic stance that Wittgenstein opts to consider language as it occurs in everyday life, but moreover due to an understanding of human life as necessarily situated in their everyday lives (Overgaard, 2006, pp. 64-65).

Overgaard's conclusion that Wittgenstein's understanding of human beings is "remarkably similar to the one we find in phenomenologists such as Husserl, Merleau-Ponty, and Heidegger" does not mean that Wittgenstein should be considered as a phenomenologist (Overgaard, 2006, p. 67). Nonetheless, I do believe that this similarity between Wittgenstein and Phenomenology justifies considering Wittgenstein as compatible with the phenomenological commitment of Postphenomenology, where a human being and their environment are understood as co-constitutive. It moreover shows that notions such as language-games and forms of life can be understood in phenomenological terms, insofar as they are taken to refer to the embeddedness and situatedness of human life, human language, and human actions. Therefore, I believe that it is worthwhile to consider whether and how these notions may be complementary to the postphenomenological program, which I do in the following section.

2.2 - Wittgenstein's views are complementary to Postphenomenology

In the final section of chapter two, I suggested that Postphenomenology is perceived to be lacking with regard to its capacity for addressing and understanding the social and cultural embeddedness of human-technology relations. That thought was continued in chapter three, where I suggested that a Wittgensteinian understanding of technology use is exceptionally capable of doing this, and may therefore complement the postphenomenological perspective. This suggestion raises two important questions. First, it should be considered to what extent Postphenomenology already addresses social and cultural embeddedness, as it is

unfair to claim that such themes have never been discussed. I suggest that one can appeal to the notions of *multistability* and *macroperception* to demonstrate Postphenomenology's acknowledgement of social and cultural embeddedness. Second, it should be considered how the notions of language-games and forms of life can be incorporated into the Postphenomenological framework.

Multistability and macroperception

The simple observation that the context within which technological artifacts reside matters for how they appear and are used is by no means novel for Postphenomenology. Ihde himself appreciates this already in 1999 when he writes: "No technology is 'one thing,' nor is it incapable of belonging to multiple contexts" (Ihde, 1999, p. 47). What this means is that an artifact cannot be identified with a single human-world relationship, but that it may facilitate various relationships that come about in different circumstances. Even something that appears to be fairly straightforward like a hammer might be used to hammer in nails in one context, and figure as a murderweapon in the next. The notion of multistability is intended to capture this aspect of technology, signifying that human beings can always relate to artifacts in different ways. At the same time, it also emphasizes that a single artifact facilitates only a limited set of human-technology relations due to its materiality. A hammer might be used for various purposes, but can never be used as a water bottle, a wallet, or a pair of headphones; its materiality limits its use. As such, any artifact is multistable in the sense that there exist multiple stable relations that human beings can have towards it, limited by its materiality (Rosenberger, 2014, p. 377).

What is not captured by the notion of *multistability*, however, is why certain stable relations are more or less dominant than others in particular contexts. While the materiality of artifacts might determine which stabilities can exist, it does not determine which stabilities will exist in a given context. To account for this, postphenomenologists refer to Ihde's distinction between *microperception* and *macroperception*, where the former is to be understood as "sensory-bodily" and the latter as "cultural-hermeneutic" (Ihde, 1995, p. 76). Importantly, Ihde holds that the two cannot be separated from one another, suggesting that microperception only ever exists in the context of macroperception. This cultural macroperception, then, is what postphenomenologists understand as structuring multistabilities, ultimately determining which human-technology relation occurs in different cultural contexts (cf. Hasse, 2008).

The role of social and cultural embeddedness manifests itself in Postphenomenology through the notions of multistability and macroperception. As such, Postphenomenology aligns itself with the Wittgensteinian understanding by acknowledging the importance of the social and cultural context of technology use. What remains lacking, however, is a clear indication as to *how* macroperception structures multistabilities. It is unclear how social and cultural beliefs and practices stand in relation to individual

human-technology relations. In the following section I suggest that Wittgensteinian notions can be adopted to illuminate this relation.

Language-games and forms of life as macroperceptual structuring

Insofar as the Wittgensteinian view provides an understanding of the social and cultural embeddedness of human-technology relations it might be considered as offering an account of the macroperceptual structuring of multistabilities. This to say that the use of an artifact is not merely determined by its materiality, but moreover by the grammar of a particular language-game. To go back to the hammer, it is not just that a hammer is predominantly used to hammer in nails because of its material features, it is moreover due to a grammar of socially and culturally embedded practices of hammer-use. The notions of language-games and forms of life, then, become answers to the question as to why certain stable human-technology relations become dominant over others.

The Wittgensteinian view moreover allows for an understanding of the inverse relation between individual human-technology relations and their social and cultural contexts. Much like Ihde states that there does not exist macroperception "without its microperceptual foci," so too does a Wittgensteinian perspective suggest that there are no language-games or forms of life without individual uses of language, and no technology-games without individual uses of technology (Ihde, 1990, 29, Coeckelbergh, 2018b, p. 1506). As a result, the way that individual human-technology relations affect macroperceptions is *from within*. The rules of the grammar that structure multistabilities are manifested in how technologies are used. Thus, when human beings start using technologies differently, those rules will change in effect. New patterns will arise, and those will guide technology-use going forward, but it is ultimately use that enacts it. As Wittgenstein writes: "it is our acting, which lies at the bottom of the language-game" (Wittgenstein, 1969, § 204).

Language-games and forms of life as backgrounds

Another way in which the notions of language-games and forms of life can be included in the postphenomenological framework is as elaborations of the use of the word *background*. As discussed in chapter two, the 'background relation' is part of the basic framework vocabulary of Postphenomenology. At the same time, however, it is perhaps Postphenomenology's least explored and employed notion, at least in my personal experience of reading postphenomenological works. While the embodiment, hermeneutic, and alterity relations have all enjoyed thorough scrutiny in past decades, the background relation seems to have been left by the wayside (see, e.g., Verbeek, 2008, Ihde, 2012, Wellner, 2014). Where

Postphenomenology as a field is introduced, background relations are often an afterthought, or not discussed at all. I believe that the field of Postphenomenology would benefit greatly from a closer examination of background relations. The Wittgensteinian perspective offers a great avenue for carrying out this project.

The idea of a background is critical for Wittgenstein. He writes in On Certainty: "But I did not get my picture of the world by satisfying myself of its correctness: nor do I have it because I am satisfied of its correctness. No: it is the inherited background against which I distinguish between true and false" (Wittgenstein, 1969, § 94). This background can be interpreted as the grammar of language-games and forms of life. It is, as Wittgenstein says, *against* this background of grammar that human beings are able to make sense of their environments, and may be understood as the foundation of meaning (cf. Moyal-Sharrock, 2015, p. 25). As much as grammar is foundational, it is also dynamic. As I have discussed in this thesis, language-games, forms of life, and their grammars are affected by technologies. Even some certainties, which form the "unmoving foundation of [our] language-games" (Wittgenstein, 1969, § 403) are arguably subject to technological mediation (Hermann, unpublished).

If appreciating the background of everyday language and action is crucial for understanding its meaning, then so too should be understanding how technological artifacts become part of this grammar. The background relations of Postphenomenology, then, may be understood to refer to cases where technological artifacts are not only embedded in the grammar of its social and cultural contexts, but come to shape this very grammar. As illustrated in part two of chapter three, how deepfakes are used and interacted with is not only guided by pre-existing norms and practices, but moreover challenges those norms and practices to eventually shape new ones. In the case of deception, deepfakes are subject to norms of trust in particular institutions. But as they challenge these norms, deepfakes will eventually lead to new norms and practices.

Conclusion

The philosophical notions introduced in this thesis, provided by the postphenomenological and Wittgensteinian perspectives on technologies in use, lead to a rethinking of the Infopocalypse. Rather than information-centric, described in abstract terms, and regarded as an apocalyptic scenario, the resultant view of deepfake technology's societal impact is complex, multifarious, and ongoing. Furthermore, I argued that the Wittgenstienian perspective is not only compatible with Postphenomenology, but can moreover provide elaboration of macroperceptual structuring of multistabilities and background relations, as such complementing the postphenomenological approach.

CONCLUSION

In this thesis, I have presented an understanding of the societal impact of deepfake technology as based on the philosophical perspectives of Postphenomenology and the later Wittgenstein. The postphenomenological perspective allows for an understanding of human-deepfake relations that takes into account the nuances of human experiences. As a result, it demonstrates that the impact of deepfake technology is not merely a question of acquiring information from digital media, but rather a question of the way that deepfakes shape human experiences. The Wittgensteinian perspective supplements Postphenomenology by addressing the social and cultural embeddedness of technology use. As such, it allows for an understanding of the use of deepfakes as shaped by social and cultural contexts, and provides insights as to how deepfakes may in turn challenge the grammar of those contexts.

The analysis offered in this thesis leads to two main insights. First, it shows that the notion of the Infopocalypse is misguided with respect to its focus on information, its generalization of technological impact, and its understanding of social practices and cultures as static. The societal impact of deepfake technology must rather be understood as an ongoing dynamic between individual human-technology relations, social practices and cultural meanings, that continuously reinforce and challenge each other. Second, this thesis suggests that the Wittgensteinian perspective on technologies in use is both compatible with and complementary to the Postphenomenological tradition, by offering new ways to understand the macroperceptual structuring multistabilities and background relations.

There exist various avenues for further research in regards to this thesis project. It might be worthwhile, for example, to consider how the insights produced in this thesis could be helpful in developing an ethics of deepfake technology. In particular, I suggest that an understanding of the dynamic relationship between individual uses of deepfakes and existing norms and practices may be helpful in understanding how deepfakes can affect such norms and practices in morally relevant ways. As such, the analysis provided in this thesis can be considered as adjacent to recent work on *techno-moral change*, which addresses exactly this question (see Swierstra et al., 2009, Nickel et al., 2021). It may also be constructive to consider how my attempt at incorporating Wittgensteinian notions into the postphenomenological perspective relates to other recent developments in Postphenomenology, such as Kudina's articulation of the *lemniscate principle* (Kudina, 2021).

REFERENCES

Achterhuis, H. J. (2001). *American philosophy of technology: The empirical turn*. Indiana University Press.

Ayyub, R. (2018). I Was The Victim Of A Deepfake Porn Plot Intended To Silence Me. Accessed via: https://www.huffingtonpost.co.uk/entry/deepfake-porn_uk_5bf2c126e4b0f32bd58ba316

BBC News. (2021, March 15). Mother "used deepfake to frame cheerleading rivals." Accessed via: https://www.bbc.com/news/technology-56404038

Biletzki, A., & Matar, A. (2002). Ludwig Wittgenstein. Accessed via: https://plato.stanford.edu/entries/wittgenstein/?fbclid=IwAR3Zuz7cDWwdoXnnlzrWESuRW8KnBK7RB7_AcKQatGaFDcZy5kVEg0MPMgs

Brey, P. (2010). Philosophy of technology after the empirical turn. *Techné: Research in Philosophy and Technology*, *14*(1), 36-48.

Chesney, B., & Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *Calif. L. Rev., 107*, 1753.

Coeckelbergh, M. (2018a). Technology and the good society: A polemical essay on social ontology, political principles, and responsibility for technology. Technology in Society, 52, 4-9.

Coeckelbergh, M. (2018b). Technology games: Using Wittgenstein for understanding and evaluating technology. Science and Engineering Ethics, 24(5), 1503-1519.

Coeckelbergh, M., & Funk, M. (2018). Wittgenstein as a philosopher of technology: tool use, forms of life, technique, and a transcendental argument. Human Studies, 41(2), 165-191.

Coeckelbergh, M. (2019). Technology Games/Gender Games. From Wittgenstein's Toolbox and Language Games to Gendered Robots and Biased Artificial Intelligence. In *Feminist philosophy of technology* (pp. 27-38). JB Metzler, Stuttgart.

De Vries, K. (2020). You never fake alone. Creative AI in action. Information, Communication & Society, 23(14), 2110-2127.

Facebook. (2018, April 17). You won't believe what Obama says in this video. Accessed via: https://www.facebook.com/watch/?v=10157675129905329

Fallis, D. (2020). The Epistemic Threat of Deepfakes. *Philosophy & Technology*, 1-21.

Floridi, L. (2018). Artificial intelligence, deepfakes and a future of ectypes. *Philosophy & Technology*, *31*(3), 317-321.

Gier, N. F. (1981). *Wittgenstein and phenomenology: A comparative study of the later Wittgenstein, Husserl, Heidegger, and Merleau-Ponty*. SUNY Press.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. Advances in neural information processing systems, 27.

Gui, J., Sun, Z., Wen, Y., Tao, D., & Ye, J. (2020). A review on generative adversarial networks: Algorithms, theory, and applications. arXiv preprint arXiv:2001.06937.

Hacker, P. (2015). Forms of life. *Nordic Wittgenstein Review*, 1-20

Hasse, C. (2008). Postphenomenology: Learning cultural perception in science. *Human Studies*, *31*(1), 43.

Heidegger, M. (1954). The question concerning technology. Technology and values: Essential readings, 99, 113.

Heidegger, M. (1962). *Being and Time*. Translated by John Macquarrie and Edward Robinson.

Hensley, J. M. (2012). Who's Calling Wittgenstein a Pragmatist?. *European Journal of Pragmatism and American Philosophy*, *4*(IV-2).

Henry, N., Powell, A., & Flynn, A. (2018). AI can now create fake porn, making revenge porn even more complicated. *The Conversation*, *28*.

Hermann, J. (2019). The Potential Mediating Role of the Artificial Womb. *Justice Everywhere*. Accessed via: http://justice-everywhere.org/gender/tba/

Hermann, J. (unpublished). When the ground starts to shake. *unpublished manuscript*, 1-16.

Hopster, J. (2021). What are socially disruptive technologies?. *Technology in Society*, *67*, 101750.

Husserl, E. (1970) *Logical Investigations*. Humanities Press International

Ihde, D. (1990). Technology and the lifeworld: From garden to earth. Bloomington: Indiana University Press.

Ihde, D. (1995). *Postphenomenology: Essays in the postmodern context*. Northwestern University Press.

Ihde, D. (1999). Technology and Prognostic Predicaments. *AI & Society, 13*, 44-51

Ihde, D. (2009). *Postphenomenology and technoscience: The Peking university lectures*. Suny Press.

Ihde, D. (2012). Postphenomenological re-embodiment. *Foundations of Science*, *17*(4), 373-377.

Introna, L. D., & Ilharco, F. M. (2004). The ontological screening of contemporary life: a phenomenological analysis of screens. *European Journal of Information Systems*, *13*(3), 221-234.

Irwin, S. O. N. (2016). *Digital media: Human–technology connection*. Rowman & Littlefield.

Kerner, C., & Risse, M. (2021). Beyond Porn and Discreditation: Epistemic Promises and Perils of Deepfake Technology in Digital Lifeworlds. *Moral Philosophy and Politics*, *8*(1), 81-108.

Ketting, E. (2000). De invloed van orale anticonceptie op de maatschappij. Netherlands Tijdschrift Geneeskunde 144, 238-286.

Kienzler, W. (2017). Wittgenstein's philosophical development. *A Companion to Wittgenstein*, 23-40.

Kudina, O. (2021). "Alexa, who am I?": Voice Assistants and Hermeneutic Lemniscate as the Technologically Mediated Sense-Making. *Human Studies*, 1-21.

Langton, R., & West, C. (1999). Scorekeeping in a pornographic language game. Australasian Journal of Philosophy, 77(3), 303-319.

Lewis, D. (1979). Scorekeeping in a language game. In *Semantics from different points of view* (pp. 172-187). Springer, Berlin, Heidelberg.

Loiperdinger, M., & Elzer, B. (2004). Lumière's arrival of the train: Cinema's founding myth. *The Moving Image*, *4*(1), 89-118.

Merleau-Ponty, M. (1965). *Phenomenology of perception*. Translated by Colin Smith.

Moyal-Sharrock, D. (2005). Unravelling certainty. In *Readings of Wittgenstein's on certainty* (pp. 76-99). Palgrave Macmillan, London.

Moyal-Sharrock, D. (2015). Wittgenstein on Forms of Life, Patterns of life, and ways of living. *Nordic Wittgenstein Review*, 21-42.

Moyal-Sharrock, D. (2017). Can Wittgenstein be called a pragmatist?. *PARADIGMI*. 67-85. 10.3280/PARA2016-003005.

Moyal-Sharrock, D. (2020). Wittgenstein's grammar: Through thick and thin. In *WITTGENSTEINIAN (adj.)* (pp. 39-54). Springer, Cham.

Nickel, P. J., Kudina, O., & van de Poel, I. Moral Uncertainty in Technomoral Change: Bridging the Explanatory Gap.

Ovadya, V. (2018). What's Worse Than Fake News? The Distortion Of Reality Itself. *New Perspectives Quarterly* 35, no. 2, 43-45.

Overgaard, S. (2006). The problem of other minds: Wittgenstein's phenomenological perspective. *Phenomenology and the Cognitive Sciences*, *5*(1), 53-73.

Patrini, G. (2019). Mapping the deepfake landscape. *Deepfake Labs*, *10*.

Pfefferkorn, R. (2019). " Deepfakes" in the Courtroom. *BU Pub. Int. LJ, 29*, 245.

Posetti, J., Shabbir, N., Maynard, D., Bontcheva, K., & Aboulez, N. (2021). The Chilling: Global Trends in Online Violence Against Women Journalists.

Putnam, H. (1995). Pragmatism: An open question. *Wiley*.

Rini, R. (2020). Deepfakes and the Epistemic Backstop. *Philosopher's Imprint*, *20*(24).

Rosenberger, R. (2009). The sudden experience of the computer. *AI & society*, *24*(2), 173-180.

Rosenberger, R. (2014). Multistability and the agency of mundane artifacts: From speed bumps to subway benches. *Human Studies*, *37*(3), 369-392.

Rosenberger, R., & Verbeek, P. P. (2015). A field guide to postphenomenology. Postphenomenological investigations: Essays on human-technology relations, 9-41.

Schiff, K. J., Schiff, D., & Bueno, N. S. (2021) The Liar's Dividend: How Deepfakes and Fake News Affect Politician Support and Trust in Media. Accessed via: https://osf.io/f75v2/download

Schwartz, O. (2018). You thought fake news was bad? Deep fakes are where truth goes to die. *The Guardian*, *12*.

Strang, K. D. (2018). Socio-cultural distrust of Internet information. *International Journal of Trust Management in Computing and Communications, 4(2)*, 95-110.

Ströker E. (1997) Intentionality and Constitution: Changes in the Concept of Intentionality in Edmund Husserl's Philosophy. In: The Husserlian Foundations of Science. Contributions to Phenomenology, vol 30. Springer, Dordrecht. https://doi.org/10.1007/978-94-015-8824-9_4

Stupp, C. (2019). Fraudsters used AI to mimic CEO's voice in unusual cybercrime case. *The Wall Street Journal*, *30*.

Toews, R. (2020). Deepfakes are going to wreak havoc on society. We are not prepared. Forbes. Accessed via: https://www.forbes.com/sites/robtoews/2020/05/25/deepfakes-are-going-to-wreak-havoc-on-society-we-are-not-prepared/?sh=1ca7d4574940

Twitter. (2018, April 17). You won't believe what Obama says in this video! Accessed via: https://twitter.com/BuzzFeed/status/986257991799222272

Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media+ Society*, *6*(1), 2056305120903408.

Van Huijstee, M., Van Boheemen, P., Das, D., Nierling, L., Jahnel, J., Karaboga, M., & Fatun, M (2021). Tackling deepfakes in European policy. Accessed via: https://www.europarl.europa.eu/RegData/etudes/STUD/2021/690039/EPRS_STU(2021)690039_EN.pdf

Vedder, A. H. (2001). Misinformation through the internet: epistemology and ethics. Accessed via: https://pure.uvt.nl/ws/portalfiles/portal/419310/wrdreli.pdf

Verbeek, P. P. (2001). Don Ihde: The Technological Lifeworld. In *American Philosophy of Technology: The Empirical Turn.* (pp. 119-146). Indiana University Press.

Verbeek, P. P. (2008). Cyborg intentionality: Rethinking the phenomenology of human–technology relations. *Phenomenology and the Cognitive Sciences*, *7*(3), 387-395.

Verbeek, P. P. (2015). COVER STORY beyond interaction: a short introduction to mediation theory. *interactions*, *22*(3), 26-31.

Victor, D. (2021, March 10). Your Loved Ones, and Eerie Tom Cruise Videos, Reanimate Unease With Unease With Deepfakes. *New York Times*.

Wellner, G. (2011). Wall-window-screen: How the cell phone mediates a worldview for us. *Humanities and technology Review*, *30*.

Wellner, G. (2014). The quasi-face of the cell phone: Rethinking alterity and screens. *Human Studies*, *37*(3), 299-316.

Whiting, D. (2012). Languages, language-games, and forms of life. Accessed via: https://eprints.soton.ac.uk/343022/1/Languages%252C_language_games_and_forms_of_life_-_Daniel_Whiting.pdf

Winner, L. (1983). Technologies as forms of life. In *Epistemology, methodology, and the social sciences* (pp. 249-263). Springer, Dordrecht.

Wittgenstein, L. (1953). *Philosophische untersuchungen/philosophical investigations*, Revised 4th ed. (G. E. M. Anscombe, P. M. S. Hacker & J. Schulte, Trans.). Malden, MA: Wiley 2009.

Wittgenstein, L. (1969). *On certainty* (D. Paul & G. E. M. Anscombe, Trans.). Oxford: Basil Blackwell.

YouTube. (2018, April 17). You won't believe what Obama says in this video! Accessed via: https://www.youtube.com/watch?v=cQ54GDm1eL0