Erasmus University Rotterdam

In deep(fake) water?
Online news media framing of artificial intelligence technology

Master thesis
MSc Sociology – Engaging Public Issues

Name: Femke van Bruggen

Student number: 469358

Thesis supervisor: Willem Schinkel

Erasmus University- Erasmus School of Social and Behavioural Sciences

Date of submission: 16-06-2022

Acknowledgements

This Master thesis is part of the final process of graduating from the MSc Sociology, with a specialization in Engaging Public Issues, at Erasmus University Rotterdam. This project focuses on deepfake technology and media framing. The finalization of this project would not have been possible without the help of several people that I would like to thank.

First, I would like to thank my thesis supervisor Willem Schinkel for his critical insights, extensive academic knowledge, attention to detail, constructive feedback and interest in the development of the topic. Thank you so much. Secondly, I would like to thank my fellow classmates, who provided me with much needed feedback, motivation, and inspiration. Lastly, I want to mention the support of my family and friends who fully supported me during the last phase of the academic year and final part of my 'university life'.

STATEMENT OF ORIGINALITY

This document is written by student Femke van Bruggen, who declares to take full responsibility for the contents of this document. I declare that the text and the work presented in this document is original and that no sources other than those mentioned in the text and its references have been used in creating it.

In deep(fake) water?
Online news media framing of artificial intelligence technology

ABSTRACT

Recent technological advances in the field of artificial intelligence have given rise to the phenomenon of deepfake technology, which involves the creation of fake, yet highly realistic digital content, mostly videos. Since this technology is still new, studies investigating deepfakes are still relatively rare. Therefore, this project provides an interesting contribution to the literature on framing and social imaginaries. The overall aim of this study is to investigate how online news media frame deepfake technology through a qualitative content analysis (CQA). The content analysis of several major online news platforms allowed for the emergence of four prominent frames across the data: the future frame, the threat frame, the responsibility frame, and the corporate frame. Firstly, the *future frame* reveals that journalists tend to frame deepfake technology in relation to the future through hypothetical scenarios and questions. Secondly, *the threat frame* involves the tendency of online news media to present deepfake technology as a significant threat or risk to society through associations of words and metaphors. Third, the *responsibility frame* presents deepfake technology as a 'wicked problem', as a result (implicit) questions are raised in online news media about which actor(s) should take responsibility for deepfakes. This study found that law- and policymakers and tech corporations are especially mentioned as important actors to solve the deepfake 'problem' across several domains, such as pornography. Fourth, the *corporate frame* highlights the tendency of online news media to highlight the impact that deepfake technology has/will have on economic processes. This impact is proposed to be optimistic and positive for the purpose of growth, optimization, and development of businesses. Lastly, the study revealed that journalists fail to include relevant deepfake news stories from non-Western countries, which indicates that knowledge production and framing in online news media suppresses certain perspectives and reinforces existing power structures.

Key words: Artificial intelligence, content analysis, deepfake technology, media framing theory, social imaginaries

TABLE OF CONTENTS

1.1. INTRODUCTION

Communication does not stand apart from reality. There is not, first, reality and then, second, communication. Communication participates in the formulization and change of reality. (Ericson, 1998)

In 2018, a video of Barack Obama calling Donald Trump a 'total and complete dipshit' went viral on the internet (BuzzFeed, 2018). A few minutes in, Obama makes a few out-of-character statements, after which the video screen is split into two; revealing that the claims are not actually made by the former US president, but by actor and comedian Jordan Peele. It becomes clear that while Peele is talking, the fake "Obama" lip-syncs. As the video goes on, Peele warns viewers to remain careful when it comes to trusting content and media on the internet.

This particular type of artificial intelligence technology can be characterized as 'deepfake' and was coined in 2017 on Reddit, a news and discussion website for online communities (Somers, 2020). In its beginning phase, the software was mostly used to edit porn videos with the use of open-source face-swapping technology (Somers, 2020). However, as deepfake usage matured and was adopted by other users, its application and usage broadened. Importantly, deepfakes offer a significant number of opportunities, such as usage in the fashion industry, in which customers can try on certain clothes virtually or project them onto their bodies or for educational purposes (Godulla et al., 2021). In addition, deepfakes can be an important part of a healthy and democratic space of creative freedom and relatively harmless satire. However, it is becoming apparent that as deepfake technology develops further, it will become increasingly difficult to distinguish such manipulated media from real content (Vizoso et al., 2021).

As with any type of emerging information technology, moral and practical concerns have been raised about the potential negative, immediate consequences of deepfakes (De Ruiter, 2021; Fallis, 2020). Most importantly, research shows that about 96 percent of all deepfakes appear to be created for non-consensual pornographic purposes and revenge porn (Deeptrace, 2019). As for the remaining deepfakes, it has been suggested that the technology may be used by actors with malicious intentions for purposes ranging from identity fraud to financial market disruption, to political disruption, terrorism, and intimidation (Chesney & Citron, 2018). What sets deepfakes apart from other AI technologies, is the democratization and ease of the production and dissemination of deepfakes; this type of video content can go

'viral' in mere hours through algorithms on far reaching social networks, accessed by millions of viewers (De Ruiter, 2021; Vizoso et al., 2021). One could imagine how deepfakes, especially those conveying controversial issues or 'alternative facts', have the potential to go viral, since audiences might be less motivated to check the veracity depending on the subject sensitivity and the period in which the deepfake is released. For example, manipulated online political messages during times of conflict between nations or during elections may spread online and lead to increased agitation and potentially damaging consequences.

In order to investigate how deepfake technology can be understood as a public issue, it is important to consider the role of online news media framing. According to Entman (1993), news media have the potential to play an important role in framing issues and may influence citizens' perceptions with regards to emerging technologies such as deepfake. This 'framing' can be best understood as the existence of particular words or images that consistently come up in media in relation to a particular group, person, or larger issue. Online news media specifically, have enabled instant communication to global audiences. This is particularly relevant in the case of deepfakes, since these are spread on social media, through which they reach a wide audience. In this way, one could see how online news media can play a role in the construction and reproduction of certain narratives and stories on the internet. Based on these insights, the following research question is posed: *How do online news media frame stories about deepfake technology?*

1.2. Scientific relevance

Since deepfake technology is still a relatively new phenomenon, not much research has been published yet. Most of the existing research comes from the field of law and computer science, with a few exceptions of ethical philosophy and political and communication science (Godulla et al., 2021). According to the researcher's investigation, no studies on the framing of deepfakes in online news media have been conducted yet. There are studies about framing related to other technologies or artificial intelligence in general, however the only study that deals with this to an extent comes from Wahl-Jorgensen and Carlson (2021). The authors (2021) investigated journalistic discourses on deepfakes as the future of fake news. However, this project aims to go beyond the perspective of deepfakes in relation to fake news and consider a broader, critical approach, considering for example pornographic deepfakes and larger narratives and visions about the futures and actors surrounding the deepfake debate. Therefore, there is a substantial research gap, which this study aims to fill.

1.3. Societal relevance

Even though deepfake technology is still developing and there have been little case studies of deepfakes disrupting society to a significant extent, it seems likely that the technology will become more easily accessible and thus pervasive. As a result, emerging deepfake technology is likely to hold major implications for media and journalism, bodily autonomy, regulation, politics and economies worldwide. Since framing of emerging technologies and other public issues by online news media has the potential to shape the ways in which citizens make sense of the world around them (Entman, 1993) and potentially how the technology will be implemented and regulated, it is relevant to analyze and uncover potential storytelling narratives and power dynamics. It is important for example, to consider who or what is included, excluded, or obscured in stories about deepfake artificial intelligence and what stakeholders, and mechanisms are involved in the production of such stories. Taking this into account, we can see that politics and framing of AI may be characterized by patterns of a certain kind of domination or even a restriction of the imagination. As, Ouchchy et al. (2020) mention, there is a strong need for a more sophisticated and holistic discourse with regards to the topic of AI in media. So, through an analysis of deepfake framing, it could become more clear how knowledge and a public understanding and an imaginary of deepfake technology is created, which may inform the formation of public opinion and public acceptance and legitimization of policy regarding possible 'deepfake' futures.

2. THEORETICAL FRAMEWORK

2.1. Contextualizing deepfakes

2.1.1. Deepfake pornography

One of the most relevant concerns in the case of deepfake technology is non-consensual pornographic content in which (mostly) womens' faces are morphed onto another body in a video. In fact, 96 percent of deepfake content can be characterized as such (Deeptrace, 2019). From a historical perspective, it could be argued that deepfake pornography perpetuates a psychological and social construction of power in which a 'gazer' has power over the person captured in the gaze. In this way, an individual depicted in a deepfake video is reduced to a passive, defenseless state without any form of consent. According to Raffaghello et al. (2019) then, the overwhelming majority of deepfakes may thus be viewed as an expression of patriarchal power and a specific type of sexual violence.

2.1.2. Deepfake, fake news and politics

Besides pornography concerns, a particular warning has been issued that deepfakes may threaten politics and the democratic process in general. In fact, recent findings by Vaccari and Chadwick (2020) may provide some evidence for this claim: the authors (2020) found that the presence of deepfakes created a sense of uncertainty among citizens, which reduced trust in social media. From this, one could see how deepfakes may have the potential to undermine a shared sense of reality or 'truth' in media and politics among citizens about what is 'real' (Chesney & Citron, 2018). In line with this, Fallis (2020) describes deepfakes as a serious epistemic threat to our ability to acquire knowledge and trust 'the facts' presented to us. These concerns regarding deepfakes might be best understood in the aftermath of the 2016 US presidential election, which directed attention towards fake news and trust in the media. In fact, while contested, some researchers have claimed the emergence of a so-called post-factual or post-truth paradigm, in which perceived boundaries between real and fake are becoming increasingly blurred (Lewandowsky et al. 2017).

Interestingly, several U.S. states, such as California have now even attempted to implement policies to ban deepfakes (Paul, 2020). However, not much evidence has yet been found that deepfakes, as tools to spread disinformation have actually been used to influence elections. Overall, since this type of artificial intelligence technology is in its infancy, it is still hard to grasp how exactly deepfakes will be used in the future.

2.2. Framing theory and media

2.2.1. Framing theory

According to framing theory, frames are used in everyday life to organize experiences and make sense of these experiences (Goffman, 1974). To be more precise, framing theory emphasizes the process of selecting and making salient several aspects of a perceived reality in a communication context (Entman, 1993). In a practical sense, frames can be detected by looking for particular words or images that consistently come up in a narrative and which convey themes across news media. The repetition and reinforcement of certain ideas then creates one interpretation that is more easily comprehensible and memorable than others. Following this logic, one could see how online news media outlets have the ability to influence the ways in which audiences may respond to and construct opinions about various public issues (Entman, 1993). Moreover, frames can be understood as entities that are situated in different locations in various communication processes: in text, in a culture, in the receiver of the message, and in the sender.

To illustrate the process of framing, Entman (1993, p.52) gives the example of the cold war, which "highlighted certain foreign events-say, civil wars-as problems, identified their source (communist rebels), offered moral judgments (atheistic aggression), and commended particular solutions (U.S. support for the other side)". In essence, frames can provide a focus or boundary about what will be discussed, how it will be discussed, and most importantly, how it might not be discussed (Altheide, 1997). So, by directing attention to one dimension of a controversy over another, media framing shapes citizens' judgements of particular issues (Mueller et al., 2003).

As Borah (2011) mentions, the debate is ongoing about the conceptualization of frames; the field is described as a 'fractured paradigm' with roots in various disciplines. The author (2011) mentions that there is a desire for a clear outline of demarcated and specific frames to guide the research in this field. However, this study takes the approach that it might be more fruitful to instead consider frames based on the concept of *ideal type* by Weber (as mentioned in Swedberg, 2017) to understand social science research. In this way, frames can be used as methodological devices or tools to grasp social reality. It is impossible to establish a perfect concept that truly captures 'what is real'. So, it is important to realize that multiple frames can exist in one text and arguably, frames might even overlap. Nonetheless, framing theory is useful for finding general patterns within the data.

2.2.2. Imaginaries

Another way of looking at framing as a type of cultural conditioning and knowledge creation, is by considering the concept of *imaginary* (Mordini, 2007). According to this theory, stories and narratives become the lens or 'imaginary' through which citizens make sense of the world and the symbols around them, including emerging artificial intelligence technologies such as deepfake. A similar mechanism might be identified in the work of Jasanoff and Kim (2015) on *socio-technical imaginaries*. The authors (2015), argue that multinational corporations, such as international news media organizations, increasingly frame and act upon certain imagined understandings of what the world is like and what it ought to be. In this sense, news media may engage with fears and hopes of their audiences and in this process, reproduce certain narratives around technology (Jasanoff & Kim, 2015).

An explanation for this process may be distilled from Altheide (1997), who suggests that mass media tend to operate and are structured around an entertainment format, through which a certain media logic has emerged. In practice, this means that due to commercialization of news media, complex events are framed as problems to appeal to audiences and capture their attention. In this way, news stories tend to be characterized as entertainment commodities that need to be sold. Moreover, the author (1997) argues that due to this media logic, mass media now function as a 'fear machine' in news production to keep audiences engaged. It seems likely that, particularly in the early stages of the introduction of an unfamiliar, yet powerful technology such as deepfake, audiences may feel uncertain or anxious. Considering this and the fact that media logic works best with 'new' or unique issues (Altheide, 1997), one could see how news media organizations might capitalize on fears surrounding deepfakes in online news stories.

2.3. Framing of artificial intelligence

2.3.1. Artificial intelligence framing research

Since most people lack direct and explicit experience with or knowledge of AI, the public tends to rely on journalists reporting on AI. In fact, as Köstler and Ossewaarde (2022) note, governments, media, and tech companies have the power to create a certain linguistic hegemony, meaning that they perpetuate certain metaphors and myths when it comes to AI. Unfortunately, not much research has been carried out yet on those metaphors and myths in relation to AI. However, a study by Chuan et al. (2019) that focuses on the longitudinal framing of AI in American newspapers found that the topic of ethics has been increasingly

11

discussed in recent years. Moreover, while not specifically placed in the framing paradigm, Wahl-Jorgensen and Carlson (2021) found that journalistic coverage of deepfakes is highly speculative, and largely paints a dystopian picture of the consequences of deepfake technology. In addition, Sun et al. (2020) found that while some news articles focus on the melodramatic and sensational side of AI technologies, the data in their study points to the fact that benefits-oriented argumentation patterns (e.g. euphoric, pragmatic, and economically optimistic arguments) were more prevalent than patterns highlighting risks or limitations (e.g., economically pessimistic and relativizing arguments). In addition, Sun et al.'s analysis (2020) uncovered an international competition frame. The authors argue that due to the major impact and uncertainty associated with AI technologies, news articles from news outlets in their sample pointed to potential fundamental shifts in the future geopolitical landscape and potential power relationships.

2.3.2. Envisioning AI futures

Building on more general, relevant theories, historians of media and technology have shown that the introduction of a new technology, such as AI, tends to be accompanied by a projection of a wide range of hopes and fears such as speculations, fantasies, and hints at the future (Sturken et al., 2004). For example, it was found that mainstream media journalists had an influence in terms of shaping the public imagination during the early launch of computers through the use of misleading metaphors and technical exaggerations (Natale & Ballatore, 2017). It is particularly relevant here to note that stories or narratives dealing with artificial intelligence often rely on claims about the future development of the field and associated consequences (Natale & Ballatore, 2017). Through these predictions and perceptions of the future, ideas about a technology then form a larger narrative or story that tells us what is desirable, but also tells us about the risks or hazards that may be associated with these new developments.

When considering the role of risks or benefits, it is not only important to investigate to what extent these are emphasized in media, but it is also of key interest to investigate how they are presented. For example, research by Slovic (2000) points to the fact that most risks can be categorized into groups: the degree to which a risk is 'dreaded' and the degree to which a risk is unknown. As it turns out, people tend to overestimate unusual and spectacular risks such as floods or tornadoes and underestimate more well-known, familiar risks such as cancer or strokes. Moreover, spectacular hazards may be more easily accessed on a cognitive level by subjects and thus remembered, which explains why such hazards tend to be perceived

as riskier (Kahneman and Tversky, 1984). These insights are likely to be applicable to the framing of deepfake technology since the technology presents an unknown risk. As a result, journalists may overestimate the risk of type of AI and frame it accordingly.

Moreover, from a future oriented perspective, we may consider the work of Neiger (2007), who argues that much of the news discourse can be understood through the concept of *conjectured future*. To illustrate, this type of storytelling is aimed at speculation about future developments or events, focusing on questions such as: "What will happen" or "What is likely to happen"? In this sense, journalistic writing allows for an opportunity to discuss uncertain futures and potential worst-case scenarios. This might be especially prominent in the case of deepfake technology, since the technology is still new and not many extreme incidents of deepfake interventions have been identified yet, leaving ample room for interpretation and conjecture about the future. So, it seems likely that framing of artificial intelligence and thus deepfake technology is focused on the future, rather than on the present state and might include attempts to predict this through models and scenarios.

2.3.3. Performativity

Lastly, this attention to the potential of online news media to create visions of the future highlights how such narratives might be mobilized by certain stakeholders with interests and as a result, these stories directly shape and construct present reality (Mager & Katzenbach, 2021). Similarly, Oomen et al (2021) describe the concept of *techniques of futuring*. These techniques are defined by the authors as practices that bring together actors around one or more imagined futures and through which actors come to share orientations for action. From this we can see that frames do not only influence sense-making processes of artificial intelligence technologies; they also actively shape the world in which these technologies will play a role. In line with Merton (1948), it becomes clear that public prophecies or predictions become an important part of a situation or issue and as a result may affect subsequent developments. Overall, it becomes clear that imagined futures are socially performed, meaning that images and expectations for the future become real in that they influence decision or policy making and the ways in which social life is organized.

3. METHOD

In this section, the relevant methodological approaches will be introduced and discussed. In the context of this this study, a mostly inductive qualitative research approach is selected, more specifically a framing analysis. The collected data consists of online news media articles. In terms of data analysis, this process involves a qualitative content analysis (QCA). Because of this, the data can be analysed in a structured way whilst making use of existing literature to guide the process of interpretation. In the following sections, an elaboration on all parts will be discussed.

3.1. Data collection

For the data collection, the final selection included 93 randomly selected online news articles collected through the LexisNexis database. A selection of news, feature, and opinion or editorial articles from several major English language publications from the United States, Canada, the United Kingdom, Australia and India were chosen. In terms of selection, 'elite'/high quality newspapers were picked, since these are most likely to influence other media and lead the public debate (Entman, 2008). To identify relevant online news media articles, the keywords deepfake, deepfakes, deepfake technology were used. Importantly, articles that contain less than 300 words were excluded, since it seems unlikely that frames could be adequately identified from such short pieces. The wordcount in the selected articles ranged from 300 to 2500 words. Additionally, the timeframe from which articles are chosen ranged from June 2019 to April 2022 (end of data collection). This timeframe rationale was chosen, because the coverage of deepfakes strongly increased in mid-June 2019, after the Intelligence Committee of the US House Of Representatives organized a hearing focusing on deepfakes and artificial intelligence in a framework that was called a 'post truth future' (CBS News, 2019). Importantly, data was collected until no new themes or codes are found.

3.2. Analysis

In terms of the methodological approach, it is important to note that there is no unified paradigm or framework when it comes to framing research (Entman, 1993). However, for this project, a qualitative content analysis (QCA) was implemented. Content analysis is a method of analyzing written, verbal or visual communication messages (Cole, 1988). Overall, the aim of this approach is to build a model including several concepts and categories related to the phenomenon in question. In the context of this study, qualitative content analysis allowed for

identifying what is included in a text or image and consequently, what might also be underrepresented in naturalized news media narratives about deepfake technology. This is relevant, since news media offer the potential for some perspectives and voices to be emphasized, while others might be structurally left out.

In line with Altheide (1996), it is vital in qualitative research to understand various characteristics represented in documents and what these may represent in the broader social context. So, following Altheide (1996), counting or coding techniques, while helpful, are not the key drivers of qualitative analysis but are used for quantitative analysis. Similarly, Reese (2001, p.8) suggests that: "The collapsing of media texts and discourses into containers based on size or frequency might obscure embedded meanings, while a qualitative approach allows for analysis of ambiguity, historical contingency, and an emphasis on meaning making". Therefore, the qualitative approach will allow for a deeper understanding of underlying patterns in deepfake technology stories and allows for a deeper investigation into journalists' and editors' cultural beliefs, journalistic norms, and organizational constraints.

This project takes news articles as the unit of analysis and takes a mostly inductive approach in terms of identifying frames. While the study does build on prior theories, it is expected that new or unique categories and patterns arise due to the fact that deepfake technology is such a new phenomenon. Therefore, it is important to allow frames to emerge in an inductive manner while the research progresses (Elo & Kyngäs, 2008). QCA is known for flexibility and thus fits well with the objective of this study. So, in line with Schreier (2012), some categories of the coding frames emerge from relevant literature, while others emerge from the news articles themselves. Lastly, the tool that is used to implement the coding process and analysis progress is the computer-assisted qualitative data analysis software ATLAS.ti.

3.3. Operationalization

In terms of operationalization, a useful way to identify frames in the text and images is to utilize and distil so-called framing devices from the data, such as metaphors, historical examples from which lessons are drawn, catchphrases and visual images (Haller & Ralph, 2001). Moreover, it is also vital to consider the sources mentioned in the news story, direct and indirect quotes and particular language and terminology (Haller & Ralph, 2001).

3.4. Coding

After the collection of data, the relevant articles were coded in Atlas.Ti. For alternative deepfake frames to emerge from the data, an inductive coding process was employed (Hsieh & Shannon, 2005). First, initial coding was carried out, which involved writing notes and highlighting elements in the data while reading it. The researcher coded the indicators such as metaphors or catchphrases that serve to interpret the news article and help understand what kind of frame is being used. Next, data and notes were read through again, and as many keywords and key phrases as necessary were created. After generating the initial codes, similar open codes were grouped together and put into tentative sub-categories. The purpose of creating such tentative sub-categories with similar codes was to increase a better understanding of the different aspects of the phenomenon in question. The formulation and decision in the creation of which codes to put together in categories was done through the researcher's interpretation. During the final phase, the tentative sub-categories were cut down further into final categories. Each of these categories was named based on a short name that reflected the category or theme/pattern. More generally, a coding manual was kept, so that all category names, examples, and definitions were recorded and kept in a consistent manner.

3.5. Ethics

Most ethical concerns in qualitative research centre around having participants as a sample and thus involve not harming others and considering ethical concerns. Since this study focused on online news media articles, it is unobtrusive by nature and ethical concerns are limited. Online news articles are publicly accessible and thus do not require practices such as informed consent or confidentiality (Shaw, 2008). However, it is important that during the analysis, explanation and accuracy of the data that is being used is important: no interpretations and steps should be left out of the process. Thus, transparency should be a main concern, to demonstrate how conclusions are found (Flick, 2011).

4. RESULTS

This chapter reports the findings of the qualitative content analysis of the online news frames found in the news articles on deepfake technology by several major online news media platforms. Several frames emerged from the data and were named by the researcher as following: future frame, threat frame, responsibility frame and corporate frame. All four frames highlight different patterns and reflect the dominant ways of online framing of deepfake technology. Arguably, these frames can be considered as being part of the deepfake technology *imaginary* (Jasanoff & Kim, 2015) and may shape the way the technology is perceived and implemented. While these frames are not *ideal types* (Swedberg, 2017), or perfect examples, they do set the stage for the identification of important patterns in the data that can be used to better understand the topic at hand.

Before explaining and illustrating the four dominant frames, it is crucial to note an important overarching finding that emerged from the data as part of this larger imaginary involving deepfake technology. Most of the discussions about existing or hypothetical deepfakes involved actors and events in northern U.S., Australia, or Western Europe. For example, most stories involved well-known actors such as Barack Obama, Tom Cruise, Nancy Pelosi or Mark Zuckerberg. Meanwhile, there were also major deepfake incidents involving political actors in countries such as Brazil, Gabon and Myanmar. However, these were mentioned only sparingly. The implications of this finding will be further analyzed in the discussion section.

Future frame

The first most pervasive and overarching frame that emerged from the data is the 'future frame'. This frame is characterized by a focus on the future of society and the changes that might occur due to the rise of deepfake technology. This can be identified in quotes such as: "We are entering a new world", "The future will be synthesized" or "The age of the deepfake". Here, we can see that journalists attempt to announce or indicate the start of a new era; there exists a clear distinction between a specific understanding of the world before and the world after the introduction of deepfake technology. Moreover, many of the writers for online news articles specifically pose large questions about the future and consequences for our understanding of truth and information. For example, one journalist from Australian news outlet Crikey asks: "We know from Brexit and Donald Trump the potential of fake news; we probably know from our own families the ability of people to take as read whatever their

Facebook feed tells them. What happens to democracy when anybody can make anybody say literally anything?" This asking of questions clearly aligns with Neiger (2007) and his conceptualization of much of the news as *conjectured future*. Storytelling of this kind involves speculation about future developments or events, focusing on questions involving potential what-if scenarios. This last point is especially relevant here, since throughout the sample, several hypothetical scenarios are presented to the audience. For example, one journalist for the Guardian writes: "Fast forward and your dad, who is 75 but looks like a 45-year-old third-degree-burns victim, will sign over your inheritance to his "girlfriend", who will turn out to be a deepfake Lady Gaga who's been scamming him on Facebook. The terrifying Russell T Davies dystopia will seem like a gentle parochial comedy." While this quote might be considered tongue-in-cheek, it still serves to illustrate how an imagined future is constructed in which everyday life rituals and practices are not the same as they were in the past.

Moreover, it is interesting to note that these hypothetical scenarios extend towards political figures or to geo-political relationships and patterns. For example, Australian news outlet The Mandarin sketches a hypothetical deepfake of the US president claiming that China can no longer be trusted as a nation-state: "The president says that after consulting its great allies, including Australia and the UK, the US has no choice but to make a powerful pre-emptive military strike on China. The president concludes: "God bless our troops". Now consider what happens next. In the uneasy minutes that follow, how would Chinese authorities react? Are we sure verification would precede retaliation?" This example clearly illustrates how the construction of a particular future aims to educate audiences about the potential impact deepfakes may have on a global level. At the same time, it also paints a picture of the nation state China as a potential enemy, re-producing and predicting a particular narrative about which actors may be involved in potential conflicts.

Importantly, it is not just potential conflict situations that are being discussed, the future of actual, real-life conflicts is discussed as well. For example, some journalists mention the deepfake of Ukrainian president Zelensky that was created and spread in March 2022 during the current Ukraine war. This deepfake was of low quality and therefore did not make much of an impact, but as one journalist concluded on the Zelensky situation: "But what happens in coming years as the technology improves, as it will, and videos are created for an audience who are not only more credulous but actively seeking to have their prejudices confirmed?" Here, we can see that in this case, the future of online content creation and political attitudes is questioned, as well as the ability of audiences to think critically. In

addition, the journalist alludes to the fact that in the future, cyberwars may emerge in addition to traditional wars. Overall, it thus becomes clear that journalists do not only frame current events, but explicitly frame future events and developments as well, speculating about the trajectory and implications relating to deepfake technology.

Threat frame

The second frame that was extracted from the data is the 'threat frame'; this frame generated the most codes. To illustrate, the threat frame can be defined by a strong emphasis in online news media articles on the threat or risk that deepfake technology may pose to society, which oftentimes involves the future of society. It seems to be the case that there is a perception of threat and a looming sense of crisis among journalists reporting on deepfakes, combined with a sense of uncertainty. In some cases, the threat is even presented as existing beyond our own imagination. We can see this reflected in a statement such as: "It's something that is likely more important and more ominous than we can even imagine". These findings are in line with prior research from Wahl-Jorgensen and Carlson (2021, who note that journalists construe fearful futures in the context of deepfake technology. To illustrate the threat frame further, several key words and terms were identified multiple times across the sample such as: worry, ominous, disrupt, scary, terror, concern, danger, unnerving, terrifying and pandora's box has been opened. These key words indicate that online news media can have an influence; using these words and associations, deepfake technology is framed as a dark, dangerous force.

Interestingly, the 'deepfake threat' is presented in different forms. In some cases, the threat is framed as something that affects the individual. In a sense, one could argue that this frame bears some resemblance to what Mills (1959) might call a private trouble; a problem that pertains to the individual. To illustrate, a piece from the Canadian National Post mentions how deepfake technology will: "Turn you into a digital puppet". In this sense, we can see how this journalist described the process of agency being taken away from the individual and essentially leaves them as helpless. Similarly, another article describes the personal impact as following: "Deepfake skims the very magic of you: the way your eyes narrow as you reach for a word; the angle at which your tongue skims your teeth as you inhale; the precise way your nostrils flare as you reach a punchline. Your unique one-in-7.53 billion human face. It extracts these things, plasters them on to someone else's body and causes havoc". Here, we can see that the loss or risk of losing one's identity is emphasized. However, in other cases the deepfake threat is presented on a more collective level that affects society at large, and even government and resembles what Mills (1959) would define as a public issue. For instance: "In

national security and intelligence circles, alarm bells were ringing". In this case, you could argue that the threat is presented as a public issue that is in the process of becoming institutionalized.

Another interesting subcategory that emerged in this frame is the *war metaphor*. The language used in some news articles reflects words and expressions that indicate that there is a conflict or a war to be fought against deepfake technology. Terms such as: "frontier", "arms race" and "under attack" can be found in a significant number of articles. According to these articles, there is an urgent battle against this threat, and we should prepare, as a society, to fight back against it. It might also be viewed in the context of potentially preparing the public for difficult times ahead or by emphasizing the seriousness of the situation. Importantly, the use of war metaphors is a common strategy for representing significant challenges. (Flusberg et al., 2018) One could think of the war against poverty, crime, drugs, cancer, or inflation. Moreover, according to Flusberg et al. (2018), this metaphor is even used for non-obvious enemies such as traffic jams, plastic bags, or videogames. The authors (2018) suggest that war metaphors are so common since they are based on fairly simple and widely shared schematic knowledge that structures our ability to make sense of a situation and because the metaphor expresses an urgent, negative emotional tone that captures attention.

It is important to note that while almost all news outlets use the threat frame, there are some news outlets that challenge this frame and suggest that deepfakes may not be as dangerous as we think. For example: "It should also be pointed out that, although the technology is very impressive, the idea that a deep fake could literally be indistinguishable from real footage is still only theoretical. Examples such as the one above are clearly fake and incongruous, while an attempt to actually deceive using deep fakes would require incredibly advanced software, meticulous data and a way to imitate the subject's voice as well as their face". Here, one can see that sophisticated deepfake production is framed as practically unobtainable as of now.

Responsibility frame

The third frame that emerged from the data is what can be characterized as the 'responsibility frame'. It seems to be the case that deepfake technology is framed by online news media journalists as a so-called *wicked problem* (Rittel & Webber, 1973), meaning that it is complicated and difficult or even impossible to solve. As a result of this perceived complexity or 'wickedness', (implicit) questions are raised in online news media about which actor(s) should take responsibility or be held accountable for deepfake as a wicked problem and its

potential solutions. To further support this perspective, news media tend to frame deepfake technology as operating in an 'ecosystem', which indicates that a complex system with various actors, nodes and links is apparently in place.

Across the sample, it becomes clear that the question of responsibility is traced back to a variety of different actors and stakeholders. First, law- and policymakers are brought into the larger picture a significant number of times. In fact, many of the articles emphasize that there is a strong need for actors in this field to address the grey area that currently exists with regards to legal regulation of deepfake technology. This is especially clear in the case of deepfake crime such as scamming and identity theft. One journalist asks: "And what of the legality of this process? It seems to me that active and aware lawmakers would take immediate steps to make the unauthorized production of AI deepfakes a felony offence, at least in the case where the fake is being used to defame, damage, or deceive. And it seems to be that we should perhaps throw caution to the wind and make this an exceptionally wide-ranging law". Here, it becomes clear that a call is being made to point to the fact that lawmakers should take initiative and expand on existing regulatory infrastructures. Interestingly, one could see here how an overall critique of the larger regulatory system is missing and blame is specifically placed on law and policymakers. So, interestingly, responsibility is not placed on the 'system' as a whole.

Similarly, in the case of deepfake pornography, responsibility also tends to be placed on actors in the legal system. However, some online news media present a narrative that places responsibility upon actors within the tech industry. For example, one journalist from The Telegraph mentions: "What I am calling for is something that will stop the perpetual catch-up that we have to play with the tech industry where the law is lagging behind the imagination of IT entrepreneurs who seem to want to make profit out of the humiliation, particularly, of women". This framing of IT entrepreneurs as taking advantage of women is thus in line with Raffaghello et al. (2019), who suggest that deepfakes may be viewed as an expression of patriarchal power and a specific type of sexual violence. Moreover, actual narratives or interviews with female victims of deepfake porn are not found in any of the articles in this sample; illustrating that these individuals are not necessarily given a voice.

Building on this role that tech companies play in terms of dealing with deepfake technology, these organizations are mentioned a significant number of times across the data. Oftentimes, the narrative takes the form of highlighting strategies that these tech corporations undertake and integrate into their business models to detect deepfakes. For example: "In August, TikTok announced a ban on misleading "synthetic or manipulated content. Facebook

announced a specific ban on deepfakes in January with exceptions for parody and satire, while Twitter's policy is to label tweets with manipulated or synthetic media, but only to remove them if they are likely to cause harm". Moreover, one journalist mentions: "Internet companies prepare to fight the deepfake future". So, using these examples, one could argue that major tech organisations tend to be framed as heroes that are battling a metaphorical deepfake 'enemy' and are considered part of the solution to the problem.

Moreover, tech companies are also presented as honourable actors taking on the responsibility to educate the public about how to identify deepfakes. For example: "Microsoft has also worked with the University of Washington to improve media literacy to help people sort disinformation from genuine facts. 'Practical media knowledge can enable us all to think critically about the context of media and become more engaged citizens while still appreciating satire and parody,' the firm wrote in a blog post. 'Though not all synthetic media is bad, even a short intervention with media literacy resources has been shown to help people identify it and treat it more cautiously.' Microsoft has launched an interactive quiz for voters in the upcoming US election to help them learn about synthetic media, develop critical media literacy skills and get a deeper understanding of the impact deepfakes can have on democracy". In this way, Microsoft is presented as an organization that provides citizens with the right tools and information for deepfake identification. Additionally, the responsibility of recognizing a deepfake is placed on the individual.

However, despite efforts to educate the public, one could question whether these tech organizations might actually also be a part of the problem and facilitate the spread of disinformation. In fact, a journalist from The Guardian critically analyzes Instagram's response to a deepfake of Mark Zuckerberg and notes that: "Unlike previous deepfake iterations, which have been glitchy or badly dubbed, the Zuckerberg one is sophisticated and smooth. And it presented the real Zuckerberg with a dilemma: would the platform remove it? It has been allowed to remain. Instagram's reasoning is that the deepfake did not break its content moderation policies. You might say this is a positive for free speech, given that the video portrays Zuckerberg in a negative light. But you could also say it goes against the tech titan's recent promises to tackle disinformation and fake news online. Which it absolutely does." From this, it becomes clear that some news media outlets acknowledge that Instagram is not consistent in the application of its policies, operating in a grey area, thus avoiding responsibility to a certain extent. However, there is also the acknowledgement that there should also be space for free speech and potentially satire, which adds an ethical dimension to the frame. So, a contradiction and perhaps a sense of hypocrisy can be extracted here: big tech

platforms allow some deepfakes to remain, while at the same time they try to educate audiences about how to identify deepfakes. Most journalists do not seem to acknowledge the fact that tech companies could be a part of the problem by allowing certain deepfakes to remain and driving up engagement on their platforms. Interestingly, this illustrates a neo-liberal framework in which responsibility without condemnation is placed on these organizations to create their own regulation, which gives them a lot of power. In this way, the infrastructures that govern the power of tech companies is reinforced.

Corporate frame

The fourth and last frame that emerged from the data is the corporate frame. This frame encompasses the tendency of online news media to focus on the impact that deepfake technology has on economic processes and how it may change the landscape of certain industries. While most articles are categorized in technology or news sections, many journalists place business-related elements at the core of the article. For example, a significant number of articles discuss the benefits of deepfake technology for the advertising and music/film industry. This is illustrated well in the following statement: "But Xia predicts that it could become a publicity tool putting consumers into their favourite movies. 'If rights owners like Disney or Netflix wanted to use this as a marketing exercise, I could totally see that happening. It would be very smart.' Indeed, film-makers have been dreaming about this interface since the earliest days of cinema." So, deepfake technology is presented here as a useful future tool to create more customer engagement and drive creativity in filmmaking. Similarly, another interesting, perhaps less obvious point to make is the observation that certain news outlets describe how deepfake technology can promote a sense of efficiency and optimize business processes. This can be observed in a statement that was included in an article involving the CEO of a deepfake start-up: "Let's say you have 3,000 warehouse workers in North America," he says. "Some speak English, but some may be more familiar with Spanish." "If you need to communicate complex information to them, a four-page PDF is not a good way. It would be much better to make a two- or three-minute video, in English and Spanish. "If you had to record every one of these videos, it would be a huge job. Now we can do that with a [small] production cost and the time it takes someone to write the script. That exemplifies very well how technology is used today." From this statement, one could see how deepfake is presented as a positive development and as a solution to further lower costs and promote economic growth. However, this narrative leaves out the fact that these industries standardize and de-personalize interactions with workers, which may not be beneficial from a

long-term perspective. In this way, it becomes clear that most journalists operating in this frame do not tend to provide a critical account on how deepfake technology may exacerbate the neo-liberal capitalism ideology fuelling various industries.

However, some journalists do point to potential ethical issues that may arise with such optimization and automation processes. This becomes clear from a statement from a news article about automated deepfake recruitment processes on LinkedIn. To illustrate: "DiResta said that some companies want to 'automate away the manual labour of making all those phone calls or sending all those messages". The practice is not illegal, although it does breach LinkedIn's policies. But is it immoral?" Here, automation is linked to a matter of morality and is not left unquestioned. Another ethical question is raised in the context of deepfake use in the entertainment industry: "It's not hard to foresee, though, how such deepfakes could lead to ethical and intellectual property issues. If you didn't want to pay the market rate for using an established artist's music in a film, TV show or commercial, you could create your own imitation. Streaming services could, meanwhile, pad out genre playlists with similar sounding AI artists who don't earn royalties, thereby increasing profits. Ultimately, will streaming services, radio stations and others increasingly avoid paying humans for music?" Once again, the emphasis is placed on morality and the extra power that streaming services and other actors in the entertainment industry may gain and what consequences this may have for individual artist.

Along those lines, another journalist addressed a case involving the power of credit card companies facilitating payment for deepfake pornography and manipulated nude images. "Verotel, the Netherlands-based payment processor being used to facilitate the card payments, says on its website that it has enabled more than 50,000 "high-risk merchants", such as porn sites, to build their online businesses. When asked to clarify why it processed payments from the deepfake site, when the arrangement began, and whether it had informed Mastercard or Visa about the relationship, it did not respond." Here, it again becomes clear that journalists do sometimes shed light on how big corporations may perpetuate harmful practices. So, this corporate frame highlights that some of the business-related news is focused on ethics.

5. CONCLUSION AND DISCUSSION

5.1. Conclusion

The overall aim of this project was to uncover patterns in the ways in which online news media frame deepfake technology. Through a qualitative content analysis (QCA), four main frames emerged from the data: the future frame, the threat frame, the responsibility frame, and the corporate frame. First, the *future frame* was the most dominant and revealed that journalists tend to communicate about the ways in which deepfake technology will impact the future. More specifically, journalists mention the arrival of 'a new era' and questions are raised about how society should understand the construction of truth and reality. In addition, an important feature of the future frame is the presence of hypothetical scenarios and questions involving speculation about events and actors. Particularly prevalent are hypothetical scenarios involving political figures, politics in general, geo-political relationships and conflict and war. So, deepfake framing and reporting is not merely focused on the present but is also heavily skewed towards the future.

Secondly, the *threat frame* involves the tendency of online news media to present deepfake technology as a significant threat or risk to society, on both an individual and collective level. There seems to be a perception of a looming sense of crisis, which is combined with an emphasis on the uncertainty of the technology's implications. Words and metaphors such as ominous, disrupt, scary, terror, concern and Pandora's box are used often and support the construction of this frame. In addition, journalists tend to use specific words and expressions that indicate an understanding that a conflict or a war that is going to be waged against deepfake technology. So, this finding highlights the presence of a so-called war metaphor in this frame. Overall, we can see that the threat frame emphasizes the dangers of an emerging artificial intelligence technology, in this case deepfake.

Third, the *responsibility frame* presents deepfake technology as a 'wicked problem' that is almost impossible to solve, as a result (implicit) questions are raised in online news media about which actor(s) should take responsibility or be held accountable for deepfake and potential solutions. This study found that, likely due to the many crime incidents involving deepfakes, law- and policymakers are especially mentioned as important actors to solve the deepfake 'problem' in several domains, including pornography. Besides actors in the legal system, responsibility is also placed on actors within the tech industry. Oftentimes, journalists describe deepfake detection strategies that these tech corporations undertake and integrate into their business models. As a result, tech corporations are presented as fighting the deepfake

'enemy', educating the public with knowledge about deepfakes and are thus framed as the solution to the problem. However, most journalists fail to include the fact that these organizations also facilitate disinformation to continue through unclear policies. Therefore, it seems to be the case that many online news media fail to acknowledge the ethical side of tech companies' responsibility: while tech corporations strive to educate audiences about how to detect deepfakes, at the same time, these organization allow certain deepfakes to remain on their platforms. Overall, this frame does not include one actor that should take responsibility for the deepfake problem, rather multiple actors and their roles are described. According to the analysis, in this study, there is no clear solution or suggestion as to who should tackle the issue.

Lastly, the *corporate frame* involves the tendency of online news media to focus on the role that deepfake technology may play in terms of economic processes and changing the landscape of several industries such as the music industry, the film industry or even in production processes. This study found that most articles that feature the corporate frame are written from an optimistic and positive perspective, emphasizing the benefits and potential of deepfake technology, such as optimization and increased effectiveness of operations. Some journalists, however, do note potential ethical issues that may arise in the context of businesses using or indirectly supporting deepfake technology in their daily operations, however this perspective is rare. Thus, we can see that the corporate frame generally presents deepfake as a promising technology that will help businesses move forward in the future.

5.2. Discussion

Theoretical implications

This study was able to contribute to the field of online news media framing of artificial intelligence in various ways. First, the presence of the future frame and threat frame confirm and build on prior studies and findings in the literature. In line with Sturken et al. (2004), this study was able to show that the rise and introduction of a new technology, in this case deepfake is indeed framed through the projection of a wide range of hopes and fears such as speculations, fantasies and hints that involve the future (Sturken et al., 2004; Natale & Ballatore, 2017). The findings do also differ from prior research, for example, the framing in this study contrast with the work of Sun et al. (2020) on the framing of artificial intelligence technology. Interestingly, Sun et al.'s (2020) data pointed to the fact that benefits-oriented argumentation patterns were more prevalent than patterns highlighting risks or limitations.

However, while deepfake technology is also mentioned in tandem with some benefits, especially in the corporate frame, it is more so perceived as an unsolvable, wicked, dangerous threat to society. Most importantly, these articles highlight a considerable degree of uncertainty about the precise nature and impact of deepfake technology and there is a tendency to frame deepfake technology as a phenomenon that needs to be fought against in a metaphorical 'war'. In this way, it seems to be the case that mass media tend to operate according to a certain media logic in which complex events are framed as problems to appeal to audiences and capture their attention. Certainly, Altheide's (1997) proposition that mass media function as a 'fear machine' to keep audiences seems quite applicable in the case of the introduction of an unfamiliar, yet powerful technology such as deepfake. It seems to be the case that indeed, online news media organizations might capitalize on fears surrounding deepfakes in their articles.

In addition, this study was able to tentatively confirm findings from a study by Chuan et al. (2019) who found that the topic of ethics in relation to AI has been increasingly discussed in recent years. Across the sample, various ethical concerns were mentioned a significant number of times. While certainly not all articles feature such ethical concerns, several journalists do hint at potential moral or ethical issues that may arise with the further introduction of deepfake technology in society across various domains.

Moreover, as becomes clear from the analysis, the four identified frames provide their own distinct meaning and key description and traits. However, since these are *ideal types* (Swedberg, 2017) this study suggests that these frames may also overlap and can be viewed as an integrated 'wholeness' which might best be understood as being part of the larger imaginary (Jasanoff & Kim, 2015). Since media are part of the social imaginary, stories and frames about deepfake technology further shape the imaginary. It seems to be the case that framing research might benefit from taking this approach of viewing the frames as tools to find patterns, but not to take these frames as the 'full truth' so to speak.

Furthermore, this study was able to reveal certain critical perspectives and uncover power relationships, which seems to be missing in most of the existing framing literature. According to Entman (2007), media might play a pivotal role in the distribution of political or intellectual power to groups, causes, or individuals. In this way, there exists some form of content bias. Entman (2007), suggests that to reveal media content biases, patterns of slant should be identified that have the potential to prime audiences, consciously or unconsciously to support the interests of particular holders or seekers of political power. In the case of deepfake technology, the data revealed that actors and events involving deepfake technology

happening outside of the U.S. or Europe are barely discussed in the online news media sample in this study. Incidents with deepfakes involving politicians from Myanmar, India and Brazil are discussed in only nine articles across the sample. One could argue that this clearly reflects a distortion of knowledge, a colonial knowledge mindset and emphasis on Western politics end events, thus shaping society's understanding of what is considered 'relevant' deepfake news. In this sense, it turns out to be almost impossible to even understand framing of deepfake technology from different places and global perspectives, when these perspectives and events are not included in the frame in the first place. Moreover, this study was able to show that online news media report a positive bias towards big tech corporations such as Facebook and Instagram. Interestingly, journalists tend to not take a critical perspective towards those tech organizations, but rather frame them as the solution in helping society battle the deepfake technology issue. Lastly, the analysis revealed that while there is attention for the ways in which female identifying people fall victim to pornographic deepfakes, their personal perspectives are not actually included in those narratives, thereby potentially leaving out important perspectives. Future research would benefit from the integration of such critical perspectives to allow underlying power structures to emerge.

Societal implications

As for societal implications, various suggestions can be made. First, since this research indicates that deepfake technology is framed as a threat, it seems likely that such online news media framing may negatively influence the general public's perception of this phenomenon and create a culture of AI fear. It is important to note however, that the framing of artificial intelligence through future oriented, fear inducing storytelling techniques might not accurately reflect the relevant aspects of the technology. In fact, as Brooks (2021) argues, we should be wary of such doom scenarios. He suggests that we are not going to be overpowered or surprised by super intelligence. In fact, such intelligence will likely evolve over time, other intelligences will appear, and the world will have more experience with such technologies. Building on this, we can see that this is not the first time in history in which people have been worried about the veracity of videos or images. For example, many people thought that the videos of the Apollo moon landings were faked (Villard, 2004). So, in this sense, deepfakes do not necessarily pose a brand-new threat.

Secondly, it is important to consider that the framing of deepfake technology can be performative and may lead to the enactment and reinforcement of such imagined futures and power structures in social life and policy decisions. For example, by focusing on deepfake

regulation and detection in online news media, attention may get directed away from other relevant courses of action such as education about deepfake awareness. Another example could be that through the lack of attention for deepfake stories outside of the 'Global North', these regions may receive less support or funding for the process of regulating this technology.

5.2.3. Limitations

Here, several limitations will be briefly presented and discussed. First of all, the framing literature emphasizes the importance of studying a consistent set of frames. However, there is a tendency for researchers to generate a unique set of frames for every study. This 'lack of disciplined approach' may lead to a situation in which researchers can be subjective and may find the evidence they are looking for (Borah, 2011). It seems to be the case that the examination or emergence of specific frames has value in understanding a particular issue; however, these unique frames should be connected to the larger implications for framing theory (Borah, 2011). Similarly, in this study, unique frames were developed to make sense of deepfake technology. Thus, these unique frames may be considered limited in empirical relation to the connection of prior findings within the literature. However, since deepfake technology is such a new phenomenon, this approach helps to build a new framework for understanding,

Secondly, it was unfortunately beyond the scope of this project to use inter-coder reliability. Intercoder-reliability involves a more than one-person perspective on how to code the data; this allows for more transparency in terms of data collection, dialogue among researchers and improved validity and credibility (O'Connor & Joffe, 2020). Therefore, this project would potentially have benefitted from multiple perspective on the data, so that perhaps other patterns could have been identified as well.

As a final point, it is relevant to be aware of the term 'deepfake' used in this thesis project and in a more general sense and how this term may influence the way we think about this technology. By its nature, deepfake technology resonates mostly with associations of 'fake' news and (mis)information, especially in the context of the 2016 presidential election campaign (Kerner & Risse, 2020). However, as discussed earlier, deepfake technology is not only relevant in this context, but rather deals with a much larger set of contexts and issues, such as non-consensual pornography. Therefore, it might be useful to attempt to shift from the term deepfake technology to the term *synthetic media*, as suggested by Kerner and Risse (2020). In this way, this umbrella term can serve to create new narratives and imaginaries and

avoid a reductionist view of the implications of this artificial intelligence technology. For the purpose of this study however, deepfake was used since this is the most common terminology that is being used when speaking about this phenomenon. Future research might benefit from expanding on this concept further to broaden the larger discourse and create new associations.

References

Altheide, D. L. (1997). "The News Media, the Problem Frame, and the Production of Fear."
     The Sociological Quarterly, 38(4), 647–668. https://doi.org/10.1111/j.1533-
     8525.1997.tb00758.x

Borah, P. (2011). Conceptual Issues in Framing Theory: A Systematic Examination of a
     Decade's Literature. *Journal of Communication, 61*(2), 246–263.
     https://doi.org/10.1111/j.1460-2466.2011.01539.x

Buzzfeed. (2018, April 17). *You Won't Believe What Obama Says In This Video!* [Video].
     YouTube. https://www.youtube.com/watch?v=cQ54GDm1eL0

Brooks, R. (2021, October 20). *The Seven Deadly Sins of AI Predictions*. MIT Technology
     Review. https://www.technologyreview.com/2017/10/06/241837/the-seven-deadly-
     sins-of-ai-predictions/

CBS News. (2019, June 13). *House holds hearing on "deepfakes" and artificial intelligence
     amid national security concerns.* https://www.cbsnews.com/news/house-holds-
     hearing-on-deepfakes-and-artificial-intelligence-amid-national-security-concerns-
     live-stream/

Chesney, R., & Citron, D. K. (2018). Deep Fakes: A Looming Challenge for Privacy,
     Democracy, and National Security. *SSRN Electronic Journal.*
     https://doi.org/10.2139/ssrn.3213954

Chuan, C. H., Tsai, W. H. S., & Cho, S. Y. (2019). Framing Artificial Intelligence in
     American Newspapers. *Proceedings of the 2019 AAAI/ACM Conference on AI,
     Ethics, and Society.* https://doi.org/10.1145/3306618.3314285

Cole, F. L. (1988). Content Analysis. *Clinical Nurse Specialist, 2*(1), 53–57.
     https://doi.org/10.1097/00002800-198800210-00025

Deeptrace. (2019, September). *The State of Deepfakes: Landscape, Threats and Impact.*
     https://regmedia.co.uk/2019/10/08/deepfake_report.pdf

De Ruiter, A. (2021). The Distinct Wrong of Deepfakes. *Philosophy & Technology, 34*(4),
     1311–1332. https://doi.org/10.1007/s13347-021-00459-2

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced
     Nursing, 62*(1), 107–115. https://doi.org/10.1111/j.1365-2648.2007.04569.x

Entman, R. M. (1993). Framing: Toward Clarification of a Fractured Paradigm. *Journal of
     Communication, 43*(4), 51–58. https://doi.org/10.1111/j.1460-2466.1993.tb01304.x

Entman, R. M. (2007). Framing Bias: Media in the Distribution of Power. *Journal of Communication, 57*(1), 163–173. https://doi.org/10.1111/j.1460-2466.2006.00336.x

Ericson, R. V. (1998). How Journalists Visualize Fact. *The ANNALS of the American Academy of Political and Social Science, 560*(1), 83–95. https://doi.org/10.1177/0002716298560001007

Fallis, D. (2020). The Epistemic Threat of Deepfakes. *Philosophy & Technology, 34*(4), 623–643. https://doi.org/10.1007/s13347-020-00419-2

Flick, U. (2011). *Introducing Research Methodology: A Beginner's Guide to Doing a Research Project* (First ed.). SAGE Publications Ltd.

Flusberg, S. J., Matlock, T., & Thibodeau, P. H. (2018). War metaphors in public discourse. *Metaphor and Symbol, 33*(1), 1–18. https://doi.org/10.1080/10926488.2018.1407992

Godulla, A., Hoffmann, C. P., & Seibert, D. (2021). Dealing with deepfakes – an interdisciplinary examination of the state of research and implications for communication studies. *Studies in Communication and Media, 10*(1), 72–96. https://doi.org/10.5771/2192-4007-2021-1-72

Goffman, E. (1974). *Frame analysis: An essay on the organization of experience*. Harvard University Press.

Haller, B., & Ralph, S. (2001). Not Worth Keeping Alive? News framing of physician-assisted suicide in the United States and Great Britain. *Journalism Studies, 2*(3), 407–421. https://doi.org/10.1080/14616700119330

Hsieh, H. F., & Shannon, S. E. (2005). Three Approaches to Qualitative Content Analysis. *Qualitative Health Research, 15*(9), 1277–1288. https://doi.org/10.1177/1049732305276687

Jasanoff, S., & Kim, S. (2015). *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power* (1st ed.). University of Chicago Press.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist, 39(4)*, 341–350.

Kerner, C., & Risse, M. (2020). Beyond Porn and Discreditation: Epistemic Promises and Perils of Deepfake Technology in Digital Lifeworlds. *Moral Philosophy and Politics, 8*(1), 81–108. https://doi.org/10.1515/mopp-2020-0024

Köstler, L., & Ossewaarde, R. (2021). The making of AI society: AI futures frames in German political and media discourses. *AI & SOCIETY, 37*(1), 249–263. https://doi.org/10.1007/s00146-021-01161-9

Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the "post-truth" era. *Journal of Applied Research in Memory and Cognition, 6*(4), 353–369. https://doi.org/10.1016/j.jarmac.2017.07.008

Mager, A., & Katzenbach, C. (2021). Future imaginaries in the making and governing of digital technology: Multiple, contested, commodified. *New Media & Society, 23*(2), 223–236. https://doi.org/10.1177/1461444820929321

Merton, R. K. (1948). The Self-Fulfilling Prophecy. *The Antioch Review, 8*(2), 193. https://doi.org/10.2307/4609267

Mills, C. W. (1959). *The sociological imagination*

Mordini, E. (2007). Technology and fear: is wonder the key? *Trends in Biotechnology, 25*(12), 544–546. https://doi.org/10.1016/j.tibtech.2007.08.012

Mueller, C., Ferree, M. M., Gamson, W. A., Gerhards, J., Rucht, D., & Gerhards, J. (2003). Shaping Abortion Discourse: Democracy and the Public Sphere in Germany and the United States. *Contemporary Sociology, 32*(4), 461. https://doi.org/10.2307/1556562

Natale, S., & Ballatore, A. (2017). Imagining the thinking machine: Technological myths and the rise of artificial intelligence. *Convergence: The International Journal of Research into New Media Technologies, 26*(1), 3–18. https://doi.org/10.1177/1354856517715164

Neiger, M. (2007). Media oracles. *Journalism, 8*(3), 309–321. https://doi.org/10.1177/1464884907076464

O'Connor, C., & Joffe, H. (2020). Intercoder Reliability in Qualitative Research: Debates and Practical Guidelines. *International Journal of Qualitative Methods*, *19*, 160940691989922. https://doi.org/10.1177/1609406919899220

Oomen, J., Hoffman, J., & Hajer, M. A. (2021). Techniques of futuring: On how imagined futures become socially performative. *European Journal of Social Theory, 25*(2), 252–270. https://doi.org/10.1177/1368431020988826

Ouchchy, L., Coin, A., & Dubljević, V. (2020). AI in the headlines: the portrayal of the ethical issues of artificial intelligence in the media. *AI & SOCIETY, 35*(4), 927–936. https://doi.org/10.1007/s00146-020-00965-5

Paul, K. (2020, December 15). *California makes 'deepfake' videos illegal, but law may be hard to enforce*. The Guardian. https://www.theguardian.com/us-news/2019/oct/07/california-makes-deepfake-videos-illegal-but-law-may-be-hard-to-enforce

Raffaghello, I., Kastalio, L., Kalf, S., & Paisley, E. (2019). What Does a Feminist Approach to Deepfake Pornography look like? Retrieved from:https://mastersofmedia.hum.uva.nl/blog/2019/10/24/what-does-a-feminist-approach-to-deepfake-pornography-look-like/

Reese, S. (2001). *Framing Public Life: Perspectives on Media and Our Understanding of the Social World*. Routledge.

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences, 4*(2), 155–169. https://doi.org/10.1007/bf01405730

Schreier, M. (2012). *Qualitative Content Analysis in Practice* (1st ed.). SAGE Publications Ltd.

Shaw, I. (2008). Ethics and the Practice of Qualitative Research. *Qualitative Social Work, 7*(4), 400–414. https://doi.org/10.1177/1473325008097137

Slovic, P. (Ed.). (2000). *The perception of risk*. Earthscan Publications.

Somers, M. (2020, July 21). *Deepfakes, explained*. MIT Sloan. https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained

Sun, S., Zhai, Y., Shen, B., & Chen, Y. (2020). Newspaper coverage of artificial intelligence: A perspective of emerging technologies. *Telematics and Informatics, 53*, 101433. https://doi.org/10.1016/j.tele.2020.101433

Sturken, M, Thomas, D, Ball-Rokeach, S (2004) *Technological Visions: The Hopes and Fears That Shape New Technologies*. Philadelphia, PA: Temple University Press.

Swedberg, R. (2017). How to use Max Weber's ideal type in sociological analysis. Journal of *Classical Sociology, 18*(3), 181–196. https://doi.org/10.1177/1468795x17743643

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), 205630512090340. https://doi.org/10.1177/2056305120903408

Villard, R. (2004). Did NASA fake the moon landing? *Astronomy, 32*(7), 48–53.

Vizoso, N., Vaz-Álvarez, M., & López-García, X. (2021). Fighting Deepfakes: Media and Internet Giants' Converging and Diverging Strategies Against Hi-Tech Misinformation. *Media and Communication, 9*(1), 291–300. https://doi.org/10.17645/mac.v9i1.3494

Wahl-Jorgensen, K., & Carlson, M. (2021). Conjecturing Fearful Futures: Journalistic Discourses on Deepfakes. *Journalism Practice, 15*(6), 803–820. https://doi.org/10.1080/17512786.2021.1908838

CHECKLIST ETHICAL AND PRIVACY ASPECTS OF RESEARCH

INSTRUCTION

This checklist should be completed for every research study that is conducted at the Department of Public Administration and Sociology (DPAS). This checklist should be completed *before* commencing with data collection or approaching participants. Students can complete this checklist with help of their supervisor.

This checklist is a mandatory part of the empirical master's thesis and has to be uploaded along with the research proposal.

The guideline for ethical aspects of research of the Dutch Sociological Association (NSV) can be found on their website (http://www.nsv-sociologie.nl/?page_id=17). If you have doubts about ethical or privacy aspects of your research study, discuss and resolve the matter with your EUR supervisor. If needed and if advised to do so by your supervisor, you can also consult Dr. Jennifer A. Holland, coordinator of the Sociology Master's Thesis program.

PART I: GENERAL INFORMATION

Project title: Framing of artificial intelligence technologies: deepfake disruption?

Name, email of student: Femke van Bruggen, 469358fb@student.eur.nl

Name, email of supervisor: Willem Schinkel, Schinkel@essb.eur.nl

Start date and duration: March 20th – June 19th

Is the research study conducted within DPAS   **YES** -NO

If 'NO': at or for what institute or organization will the study be conducted?
(e.g. internship organization)

PART II: HUMAN SUBJECTS

1.   Does your research involve human participants.   YES -**NO**

   If 'NO': skip to part V.

   If 'YES': does the study involve medical or physical research?   YES-NO
   Research that falls under the Medical Research Involving Human Subjects Act (WMO) must first be submitted to an accredited medical research ethics committee or the Central Committee on Research Involving Human Subjects (CCMO).

2.   Does your research involve field observations without manipulations
   that will not involve identification of participants.   YES-NO

   If 'YES': skip to part IV.

3.   Research involving completely anonymous data files (secondary
   data that has been anonymized by someone else).   YES - NO

   If 'YES': skip to part IV.

PART III: PARTICIPANTS

1.   Will information about the nature of the study and about what
   participants can expect during the study be withheld from them?   YES - NO

2.   Will any of the participants not be asked for verbal or written
   'informed consent,' whereby they agree to participate in the study?   YES - NO

3.   Will information about the possibility to discontinue the participation
   at any time be withheld from participants?   YES - NO

4.   Will the study involve actively deceiving the participants?   YES - NO
   Note: almost all research studies involve some kind of deception of participants. Try to think about what types of deception are ethical or non-ethical (e.g. purpose of the study is not told, coercion is exerted on participants, giving participants the feeling that they harm other people by making certain decisions, etc.).

5.   Does the study involve the risk of causing psychological stress or
   negative emotions beyond those normally encountered by
   participants?   `   YES - NO

6.   Will information be collected about special categories of data, as
   defined by the GDPR (e.g. racial or ethnic origin, political opinions,
   religious or philosophical beliefs, trade union membership, genetic
   data, biometric data for the purpose of uniquely identifying a person,
   data concerning mental or physical health, data concerning a person's
   sex life or sexual orientation)?   YES - NO

7. Will the study involve the participation of minors (<18 years old) or
    other groups that cannot give consent?   YES - NO

8. Is the health and/or safety of participants at risk during the study?   YES - NO

9. Can participants be identified by the study results or can the
    confidentiality of the participants' identity not be ensured?   YES - NO

10. Are there any other possible ethical issues with regard to this study?   YES - NO

If you have answered 'YES' to any of the previous questions, please indicate below why
this issue is unavoidable in this study.

_____
_____
_____
_____

What safeguards are taken to relieve possible adverse consequences of these issues
(e.g., informing participants about the study afterwards, extra safety regulations, etc.).

_____
_____
_____
_____

Are there any unintended circumstances in the study that can cause harm or have
negative (emotional) consequences to the participants? Indicate what possible
circumstances this could be.

_____
_____
_____
_____

Please attach your informed consent form in Appendix I, if applicable.

Continue to part IV.

PART IV: SAMPLE

Where will you collect or obtain your data?

_____

_____

Note: indicate for separate data sources.

What is the (anticipated) size of your sample?

_____

_____

Note: indicate for separate data sources.

What is the size of the population from which you will sample?

_____

_____

Note: indicate for separate data sources.

Continue to part V.

Part V: Data storage and backup

 Where and when will you store your data in the short term, after acquisition?

Data (news articles) collected from LexisNexis will be stored on my personal computer in Atlas.Ti and potentially on an extra server online

_____

_____

Note: indicate for separate data sources, for instance for paper-and pencil test data, and for digital data files.

Who is responsible for the immediate day-to-day management, storage and backup of the data arising from your research?

I am

_____

_____

How (frequently) will you back-up your research data for short-term data security?

I aim to automatically back-up my data every day

_____

_____

In case of collecting personal data how will you anonymize the data?

-

_____

_____

_____

_____

Note: It is advisable to keep directly identifying personal details separated from the rest of the data. Personal details are then replaced by a key/ code. Only the code is part of the database with data and the list of respondents/research subjects is kept separate.

PART VI: SIGNATURE

Please note that it is your responsibility to follow the ethical guidelines in the conduct of your study. This includes providing information to participants about the study and ensuring confidentiality in storage and use of personal data. Treat participants respectfully, be on time at appointments, call participants when they have signed up for your study and fulfil promises made to participants.

Furthermore, it is your responsibility that data are authentic, of high quality and properly stored. The principle is always that the supervisor (or strictly speaking the Erasmus University Rotterdam) remains owner of the data, and that the student should therefore hand over all data to the supervisor.

Hereby I declare that the study will be conducted in accordance with the ethical guidelines of the Department of Public Administration and Sociology at Erasmus University Rotterdam. I have answered the questions truthfully.

Name student: Femke van Bruggen   Name (EUR) supervisor: Willem Schinkel

Date: March 19th   Date: March 19th

APPENDIX I: Informed Consent Form (if applicable)

-