Bachelor Thesis

University of Twente, Enschede & Universität Münster

THE DUAL NATURE OF DEEPFAKES

An Analysis of the Deepfake Discourse in the USA: Benefits, Threats, Challenges, and Solutions

Johanna Patz

Programme: Public Governance across Borders (MST)

Supervisors: Dr. Ringo Ossewaarde, Dr. Veronica Junjan

Date of Presentation: 3rd July 2024

Word Count: 11.932

ABSTRACT

This bachelor thesis aims to investigate how political actors in the USA perceive deepfakes in the period from 2020 to 2024 with a view to benefits, threats, and challenges, as well as potential solution strategies to combat the threats and challenges posed by deepfakes. To achieve results and identify direct meanings as well as uncover underlying assumptions, qualitative content analysis will be used on different types of documents, such as policy documents, legal texts, transcripts, and media articles. A coding scheme will be provided that operationalizes the core concepts, and the software ATLAS.ti will be used to analyze the collected data. The findings of the analysis uncover a dual nature of deepfakes: while deepfakes may have beneficial uses in the entertainment and healthcare industries, political actors predominantly emphasize the threats posed by deepfakes to democracy, society, and individuals, for example, threats to national security, manipulation, disinformation, psychological damage to individuals, or the erosion of trust from society in democracy. The threats and challenges arising from deepfakes are to be combated by a multi-faceted solution strategy that includes more than a single solution. Political actors highlight collaboration, education campaigns, regulatory measures, etc. as solutions.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1 BACKGROUND AND STATE OF THE ART

Today, we live in a so-called post-truth era, marked by spreading misinformation to manipulate public opinion poses a significant threat (Masood et al., 2023; Westerlund, 2019). A dangerous manifestation of this is manipulated images and videos of people increasingly found online, resulting in a new phenomenon called ‘deepfakes‘ in which individuals say or do something they never did (Vasist & Krishnan, 2022; Mustak et al., 2023). Originally, deepfakes were used to insert public figures into pornographic content. However, with advances in machine learning and artificial intelligence, they are increasingly used for political sabotage, terrorist propaganda, and the spread of fake news (Westerlund, 2019; Kshetri, 2023; Mustak et al., 2023). These highly realistic and credible videos make it increasingly difficult for people to distinguish them from authentic media, posing serious consequences and threats to societal trust and democratic processes (Mustak et al., 2023).

Despite the critical implications and growing prominence of deepfakes, research on this topic is still scarce (Vasist & Krishnan, 2022; Westerlund, 2019). Nevertheless, social scientists are slowly beginning to explore this area through research and published literature on the three main topics: 1. Deepfake Detection and Creation, 2. Risks and Threats to Society and Democracy, 3. Ethical Aspects of Deepfake Technologies (Vasist & Krishnan, 2022, Misirlis & Munawar, 2023). While both scholars identify these focus points, Misirlis & Munawar (2023) assert that regarding the USA, much literature concentrates on risks and threats to democracy and national security.

However, empirical research on this phenomenon has significant gaps. Apart from all the research and literature on the risks and threats that deepfakes can lead to, there is an evident lack of research on the beneficial applications of deepfakes, the challenges arising from the dissemination of these videos, and potential solutions and strategies to combat their misuse. Filling these gaps is urgent to entirely recognize the impact of deepfakes, in both a positive and negative sense. By identifying these gaps, there will be a substantial possibility to develop strategies to fully promote their positive potential and at the same time to mitigate their dangerous risks for democracy, society, and individuals. This research aims to fill these critical gaps by examining how political actors in the USA perceive, envision, and engage with deepfakes. On the one hand, this research is novel and highly important for the practical implementation of policies, as the results of the analysis can inform policymakers for making the right policy decisions and implementing strategies that address the urgency to combat the

dangers of deepfakes and tackle the challenges arising from deepfakes. On the other hand, the research has a social relevance by recognizing dangers and identifying challenges and solution strategies to transform these threats into an opportunity to make society and democracy more resilient. Therefore, this research makes a worthwhile contribution to this under-researched field by identifying the perspectives of political actors regarding deepfakes and based on these making recommendations to address one of the biggest challenge and danger of our time.

1.2 RESEARCH QUESTIONS

To fill this knowledge gap, this descriptive research aims to explore the perspectives of political actors in the USA regarding deepfakes. Therefore, this thesis's research question is: **how do political actors in the USA envision deepfakes in the deepfake discourse from 2020 to 2024?** Political actors refer to decision-makers such as politicians, which includes the presidents, their vice presidents, and the cabinet, as well as members of the US Congress, the members of the House of Representatives and the Senate, who have given a speech or have already made indirect initiatives such as passing laws related to deepfakes. The aim is to find out whether they tend to focus rather on advantages, threats, challenges, and/or possible solution strategies.

To better illustrate the research objective and make the main research question more concrete, particular attention will be paid to the following sub-question: **in what ways do political actors in the USA envision the benefits, threats, and challenges of deepfakes in the deepfake discourse?** This question aims to identify what sorts of aspects the political actors mention, particularly which advantages, threats, and challenges they address. The expected advantages could include the use of deepfakes in film production, education, or advertising for large businesses. Deepfakes are also expected to be seen as threats, such as the spread of misinformation, negative impact on citizen trust, or undermining democratic elections. Expected challenges could include the expansion of detection methods. Given the disadvantages and challenges that arise from the use of deepfakes, it is interesting to recognize which solution strategies political actors emphasize in the deepfake discourse. This leads to another sub-question: **in what ways do political actors in the USA envision potential solutions to combat the threats and challenges arising from deepfakes in deepfake discourse?** Potential solution strategies may include regulation through laws, public education campaigns to create awareness or the development of anti- deepfake technology.

These questions are particularly urgent and important because they address the critical gap in research and shed light on how political actors in the US perceive and engage with deepfakes

in the deepfake discourse in terms of benefits, threats, challenges, and potential solutions. By examining and uncovering the viewpoints, perceptions, understandings, concerns, and strategies of political actors, this research provides valuable insight into their underlying assumptions, meanings, and assessments. This thesis will reveal underlying policy priorities regarding deepfakes and the readiness of political actors to address this phenomenon. Understanding the perspectives of political actors is crucial for grasping the nuances of this issue and informing policymakers to develop suitable policies. These might include regulations, strategies to minimize negative effects- such as preserving democratic processes and public trust- and supporting the beneficial uses of deepfakes. Therefore, the results of this research will contribute to a deeper understanding of political actors' positioning regarding deepfakes and support the development of effective policymaking.

1.3 RESEARCH APPROACH

The approach used in this bachelor's thesis is an interpretative approach to explore and interpret how political actors in the USA perceive and envision deepfakes. Interpretive social sciences "investigate meaning behind the understanding of human behavior, interactions, and society" (Pulla & Carter, 2018, p. 10). As this bachelor's thesis aims to develop an interpretation of what political actors in the USA explicitly or implicitly say about deepfakes in the deepfake discourse in terms of benefits, threats, challenges, and solution strategies, the interpretative approach is the most appropriate one. To develop an interpretation, content analysis will be used in this thesis as an interpretive method that corresponds to the interpretive approach. It fits with the interpretive approach because it allows for the uncovering and interpretation of the meanings and underlying assumptions of statements made by political actors in the USA in the period from 2020 to 2024 in different data, that are part of the deepfake discourse, like policy documents, political speeches, legal texts or transcripts of hearings and interviews. To identify recurring themes and patterns in the deepfake discourse, the data will be examined using a coding scheme and the coding software ATLAS.ti. This bachelor thesis is structured as follows: first, the research question is theorized, followed by a description of the methodological procedure of this research. After the analysis, the results are explained and discussed, ending with a conclusion.

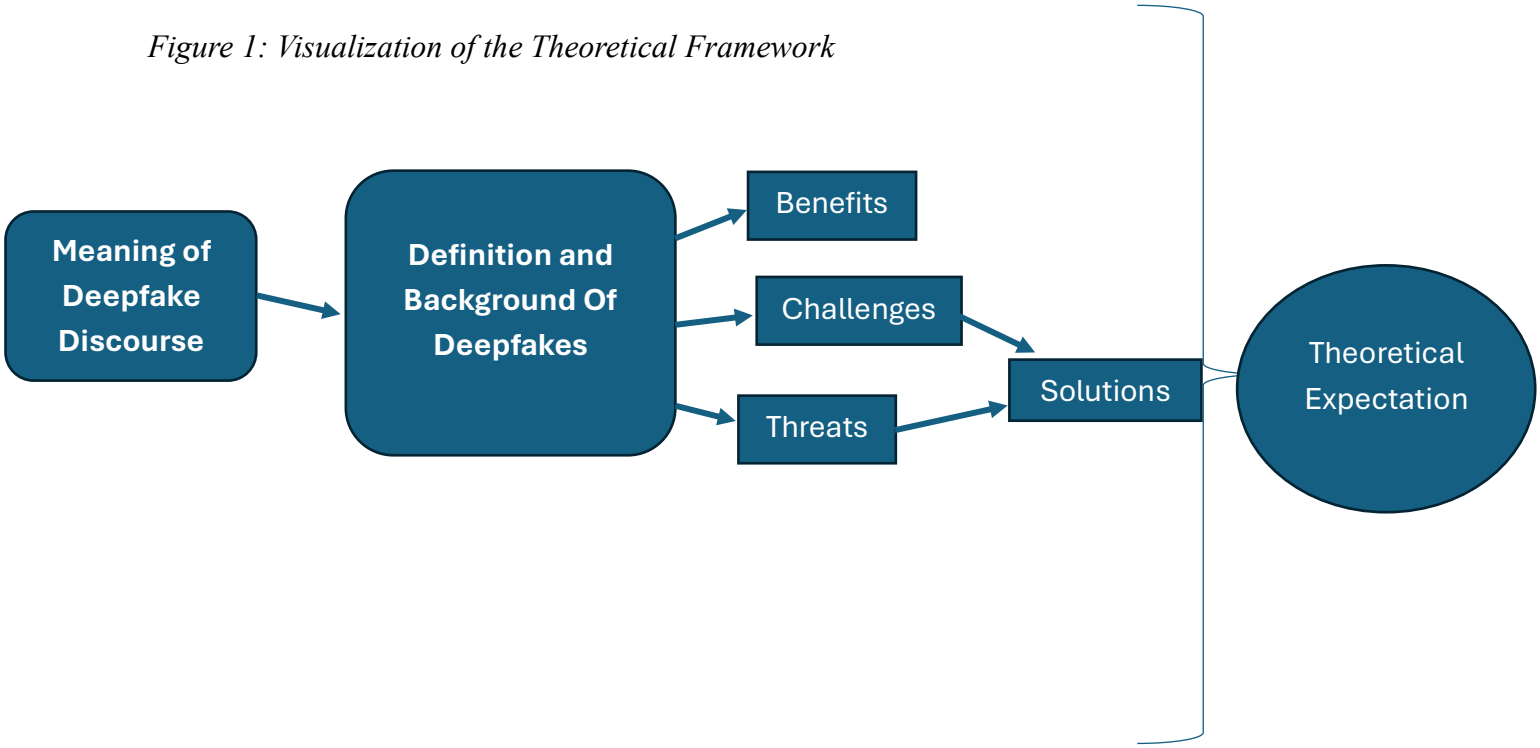
2. THEORY

2.1 INTRODUCTION

This chapter establishes a theoretical framework to support the research. It provides the reader with the context of the research and background information to make it easier to understand the

topic. It also intends to support the research question by narrowing it down through clearly defined concepts. Further, the theoretical framework is helpful for the following analysis of the deepfake discourse among political actors in the US by providing a clear understanding of the phenomenon of deepfakes, focusing on the discourse around it and the technology itself, acting as a lens through which to interpret data and results. In the following, the core concepts of the research question, as well as others that are important to explain in the context of this thesis, will be explained and discussed. First, based on existing academic literature, the meaning of 'deepfake discourse' will be defined. Next, the technology itself and further insights will be explained and explored. After identifying general advantages, but also threats and challenges that this technology brings with it and the corresponding solutions, theoretical expectations will be finally pre-formulated based on the findings of this chapter. Visualized this chapter is organized as follows:

Figure 1: Visualization of the Theoretical Framework



2.2 MEANING OF DEEPFAKE DISCOURSE

Different scholars have different understandings of the term *discourse* (Guardado, 2019). While in linguistics, "discourse refers to a unit of language longer than a single sentence (Nordquist, 2009), in this thesis it "means anything from a historical monument, a lieu de mémoire, a policy, a political strategy, narratives in a restricted or a broad sense of the term, text, talk, a speech, topic-related conversations, to language per se" (Wodak & Meyer, 2009, p.2f.). According to Cook (1989), it refers to the totality of all these elements interacting. The term *discourse* comes

from the Latin prefix *dis,* which means ‘away‘. The root word comes from *currere*, which means ‘to run‘. Literally translated, discourse means ‘run-away‘. In the context of conversations, it can be seen as the way how conversations flow (Nordquist, 2009). Teun van Dijk, a Dutch linguist stresses that a discourse “is a form of language use” (van Dijk, 1997, p.2). He claims that “people use language in order to communicate ideas or beliefs (to express emotion), and they do so as part of more complex social events” (van Dijk, 1997, p.2). Van Dijk (1997) also states that language use does not only consist of spoken language but also of written language as in documents etc.

When analyzing a discourse, one usually examines the use of spoken or written language, within a context such as that of deepfake (Henry & Tator, 2002; Nordquist, 2009). The debate or documents can deal with various aspects of deepfakes: firstly, with the ethical implications by focusing on manipulation or deception. Secondly, with the social influence deepfakes may have like shaping public opinion, the spread of misinformation and the erosion of trust. Thirdly, it might examine legal issues like privacy laws, property rights and the update of AI-related legal frameworks. Lastly, it might be about the advantages, disadvantages, solutions, and challenges that this technology brings with it, such as benefits in various fields like entertainment, medical research, and business or the misuse for the spread of false information and the erosion of trust, and as solution strategies certain detection and verification tools or raising public awareness campaigns.

2.3 DEFINITION AND BACKGROUND OF DEEPFAKES

But what exactly are deepfakes? Manipulated images and videos of people are increasingly common on the internet due to the latest technological developments in artificial intelligence and machine learning (Mustak et al., 2023). This has led to the rise of a new phenomenon called ‘deepfakes‘ (Vasist & Krishnan, 2022). The word is made up of ‘deep learning‘, which stands for AI deep learning technology, an approach to machine learning in which multiple layers are processed (Johnson & Diakopoulos, 2021; Payne, 2024). ‘Fake‘ stands for the fact that the content of the result is not real (Payne, 2024). Initially, they were created to defame and embarrass public figures such as politicians or celebrities whose faces and bodies were depicted in porn (Westerlund, 2019; Masood et al., 2023). In 2018, with the release of deepfake tools such as DeepFaceLab, the number of deepfakes posted online was in the millions (Kshetri, 2023). Created by various actors such as governments, political activists, criminals, and foreign governments but also by legitimate actors like television companies, today they are often used to create humorous or political content, whereas malicious uses increase (Zannettou et al., 2019; Westerlund, 2019).

In deepfake discourses, there are many definitions of deepfakes from different scholars, but they only differ in small details. Misirlis & Munawar (2023), for example, define deepfakes as "resulting media from the synthesis of different persons' images and videos- mostly faces-, replacing a real one" (p.26). Vasist & Krishnan (2022) consider deepfakes to be "hyper-realistic synthetic media where an individual's face in a photo or video is swapped with that of another person" (p.591). With the help of artificial intelligence and machine learning algorithms, images or videos are created that look so real that they can hardly be distinguished from reality by the public, but also by detection tools (Mustak et al., 2023; Misirlis & Munawar, 2023; Chesney & Citron, 2018). These so-called hyper-realistic videos often depict a person saying or doing something that never happened without giving their consent (Johnson & Diakopoulos, 2021; Westerlund, 2019). While in deepfake discourses many scholars emphasize the indistinguishability between real and deepfakes material, as well as the creation of videos in which people are depicted without their consent, doing something they have never said or done, Vasist & Krishnan (2022) stress three other characteristics of deepfakes: first, that deepfakes are particularly easy to create because they require few skills and resources. Secondly, the simplified transmission of the content via social media. Third, the increasing number of digital materials that can be used as data to create these deepfakes (Vasist & Krishnan, 2022).

2.4 BENEFITS, THREATS, CHALLENGES AND SOLUTIONS OF DEEPFAKES

As mentioned before, deepfakes come with various benefits, threats, and challenges and there exist many solution strategies to combat them. The following table provides an overview, which will be explained in more detail during this chapter.

Table 1: Overview of Benefits, Threats, Challenges and Solutions

Benefits	**Threats**	**Challenges**	**Solutions**
Entertainment/ film industry	Erosion of Citizens' trust	Social Media Dependence	Legislation and Regulation
Education	Spread of Disinformation	Detection Difficulties	Corporate Policies
Medicine & Healthcare	Journalistic integrity		Education and Media Literacy

Business Sector	Manipulation of democratic elections		Anti- Deepfake Technology
	Political Sabotage		
	Societal Polarization		
	National and Global Security		
	Personal Harm through blackmail & bullying		

Benefits

In deepfake discourses, scholars typically cite the following advantages of deepfakes. Firstly, in the entertainment industry, where Verdoliva (2020) mentions some beneficial applications, such as the use of deepfakes in creative arts, for advertising, movie production and video games (Misirlis & Munawar, 2023; Westerlund, 2019). They are also used for visual effects, for digital avatar creation in, for example, video games, Snapchat filters, etc. (Nguyen et al., 2022). In the film industry, deepfakes might update series without having to reshoot anything, but also bring dead artists back to life (Diakopoulos & Johnson, 2021; Nguyen et al., 2022). This is also beneficial in education: by 'bringing back to life' historical figures, lessons can be made more interesting and livelier, for example when it comes to historical speeches (Payne, 2024; Chesney & Citron, 2018; Misirlis & Munawar, 2023). A further advantage is the use of deepfakes in medicine and healthcare (Misirlis & Munawar, 2023; Westerlund, 2019). Deepfakes can act as an ‘assistive technology’ for people who have lost their voice (Nguyen et al., 2022). Payne (2024) emphasizes the research on tumors which can be improved and therefore, making treatment easier. Lastly, deepfakes may be used for advertising purposes and for content creation to help launch more lively, multilingual advertising campaigns for businesses (Mustak et al., 2023; Diakopoulos & Johnson, 2021). Nevertheless, the malicious use of deepfakes largely outweighs the positive ones (Nguyen et al., 2022).

Threats

The negative consequences of deepfakes occur on micro, meso and macro levels (Vasist & Krishnan, 2022). They threaten democracy, society and the individual who becomes victim

(Johnson & Diakopoulos, 2021). In deepfake discourses, typically the same disadvantages are mentioned. One of them is the spread of false information, particularly political disinformation, (Misirlis & Munawar, 2023; Pawelec, 2022). While traditional fake news only focuses on the content, deepfakes convey a simulation of a speaker delivering the news, which makes it seem even more real (Vasist & Krishnan, 2022). According to Masood et al. (2023) the use of deepfakes for disinformation will lead to the next threat; the erosion of trust. Citizens' trust in information, technology, journalism and democracy is being lost (Misirlis & Munawar, 2023; Westerlund, 2019; Vasist & Krishnan, 2022; Verdoliva, 2020). Deepfakes pose a further threat, especially for democracies regarding the manipulation of political elections (Chesney & Citron, 2018). As politicians are also targets for deepfakes, this can have immense effects on the electoral process (Vasist & Krishnan, 2022). The misinformation that is spread via deepfakes, for example about a political actor or his opponent, could trigger a political scandal and in turn give another party advantages by changing voter preferences (Dobber et al., 2021; Diakopoulos & Johnson, 2021). Public opinion can easily be influenced and manipulated (Verdoliva, 2020; Diakopoulos & Johnson, 2021). Another threat to democracy poses political sabotage. By creating deepfakes with completely new content and by including controversial or hateful statements, political or division in society or even societal polarization can occur (Vasist & Krishnan, 2022; Chesney & Citron, 2018). This is particularly beneficial for terrorist propaganda, which threatens national security, especially in democracies (Westerlund, 2019). Deepfakes can be used there to announce and justify violence, crime or terrorism (Mustak et al., 2023; Masood et al., 2023). National but also world security is at risk when deepfakes are used on political world leaders with fake speeches in which violence or war is announced and the public is incited (Nguyen et al., 2022). Ultimately, without reference to threats to democracy, deepfakes represent a social threat to the individual. Women and girls, who are often the target group in pornographic deepfakes, have little chance of defending themselves against it (Kshetri, 2023). In addition, these videos can also be used for bullying or blackmail, which can lead to long-term psychological and social damage to the victims (Westerlund, 2019; Verdoliva, 2020).

Challenges

In deepfake discourses, deepfakes also pose challenges, for example on social media platforms on which citizens increasingly rely for information gathering and communication (Fletcher, 2018). In heated political debates, false information can easily be spread online and simply be believed. On the one hand, due to the increasing occurrence of other algorithms that are used to produce deepfakes, videos that look increasingly real are being created and therefore uneasy to identify as fake (Fletcher, 2018). On the other hand, solutions to combat deepfakes are still in

their early stages: Nguyen et al. (2022) claim that many methods have been proposed and tested, but none are efficient. Chesney & Citron (2018) emphasize that deepfakes are becoming increasingly resistant to detection due to their high quality.

Solutions

Solutions to combat the negative effects of deepfakes often require the efforts of multiple stakeholders, such as social media platforms, policymakers, lawmakers, etc. (Johnson & Diakopoulos, 2021). In deepfake discourses, typically four possible solutions can be found: firstly, legislation and regulation (Westerlund, 2019). Although many new laws have already been passed (Meneses, 2021), according to Ray (2021), more laws are still needed to combat political deepfakes. Secondly, there are corporate policies for developing solution strategies to combat deepfakes (Westerlund, 2019). Johnson & Diakopoulos (2021) particularly emphasize the role of technical experts. Their understanding and knowledge are needed to develop solutions and social responses in collaboration with other stakeholders. Thirdly, education and media literacy are a frequently mentioned solution approach to create awareness among the population, uncover and recognize deepfakes and know how to deal with them (Westerlund, 2019; Diakopoulos & Johnson, 2021). Ultimately, anti-deepfake technology is one of the most developed strategies to detect deepfakes (Westerlund, 2019). These include deepfake detection and content authentication (Chesney & Citron, 2018). While Verdoliva (2020) calls on the scientific community to develop reliable tools that automatically detect deepfakes, Johnson & Diakopoulos (2021) again emphasize the role of the technical expert who plays a role in designing, developing and testing these tools.

2.5 CONCLUSION

This section has presented a theoretical framework for defining and explaining the core concepts of the research question. The concept of ‘deepfake discourse‘ refers to the use of language in social interactions with which beliefs, values, etc., whether written or spoken, are conveyed in a deepfake context. Therefore, the discourse can deal with various aspects such as ethical implications, social influence and legal issues. Deepfakes are videos that use artificial intelligence and machine learning algorithms to accuse a person of saying or doing something that never happened, without getting their consent. Due to developments in machine learning and artificial intelligence, it is easier to manipulate media that looks real. While deepfakes have advantages in the film industry, education, as well as in medical and business sectors, threats can arise, particularly regarding democracy, through the spread of fake news, thereby reducing citizens' trust in democratic institutions and the media. In addition, deepfakes can manipulate

voter preferences and endanger national security, for example through terrorist propaganda. This results in challenges such as the rapid distribution of these videos, especially if the content consists of false information, or solutions to these threats that are still very early in their development. Approaches to solutions include laws, corporate policies, education and antideepfake technology such as detection methods.

Concerning the research question, a theoretical answer proposes that the deepfake discourse among political actors in the USA tends to dominantly focus on the threats and negative consequences that this technology poses. While also recognizing the beneficial use of deepfakes, political actors are expected to stress the dangers and threats that come with deepfakes for the democratic process, such as the spread of misinformation, erosion of citizens' trust in institutions and the manipulation of voter preferences, which are current due to the 2024 U.S. election year. They may also emphasize the urgency of combatting the threats through solutions. Possible mentioned solution strategies will be on the one hand, the development and implementation of anti- deepfake technologies to detect deepfakes, but on the other hand, there will be a strong call for laws and legislative measures to regulate the creation and spread of deepfakes, that they ensure the use of deepfakes rather in beneficial ways than the malicious use. To conclude, the theoretical expectation based on the concepts in this chapter is that negative consequences and threats that come with deepfakes will dominate the political deepfake discourse. At the same time, there will also be a call for technological solutions, as well as legislative measures to safeguard democratic principles and processes.

3. METHODS

3.1 INTRODUCTION

The purpose of the following chapter is to outline the methodological foundation of this research and therefore to explain how the research questions of this thesis will be answered. Furthermore, it serves as an instruction for my research to find out how political actors in the USA envision deepfakes in the deepfake discourse. While illustrating the methods that are applied in this thesis, it also allows someone else to replicate this study to ensure reproducibility. The method chapter is structured as follows: first, the selected cases of the deepfake discourse in the USA will be described and the reasons for choosing this case will be given. Subsequently, the data and their sources will be outlined, detailing the collection process and justifying the chosen method for this thesis. Additionally, the content analysis method will be explained, including how the data will be examined using this approach. While further explaining the process of analysis, there will also be a justification for why this is the appropriate method to answer the

research question. A coding scheme in which theoretical concepts are operationalized to conduct the analysis will be provided. The chapter ends with a conclusion.

3.2 DESCRIPTION OF THE CASE

This thesis examines how political actors in the USA envision deepfakes in the deepfake discourse in the USA in the period from 2020 to 2024. This is an illustrative and descriptive case study because this research aims to describe a phenomenon that occurs in a real-life context. The concrete object of research is the deepfake discourse constructed by political actors in the USA in the period from 2020 to 2024, which is analyzed to understand how political actors give the phenomenon of deepfakes meaning. The deepfake discourse with its geographical and temporal scope is the case studied. The deepfake discourse is about examining spoken and written language that can be found in documents and other data that are created by political actors between 2020 to 2024. The data is analyzed for what is explicitly or implicitly mentioned in terms of benefits, threats, challenges or potential solution strategies. The geographical scope here refers to the USA, as it is one of the pioneers in global politics, but also in terms of technological innovation. The period from 2020 to 2024 serves as the temporal scope here, as several important events took place during this period in the USA. On the one hand, there are the presidential elections in 2020 and 2024, where the use or misuse of deepfakes can have either a positive or negative impact on voter preferences. In addition, Donald Trump was still president until 2021, while Joe Biden has held the office of president since 2021. This also makes the period interesting to analyze, in a comparison of two presidents, how they envision deepfakes in the deepfake discourse, i.e. where there are similarities and where there are differences.

To justify the selection of this case, the USA is an interesting case due to its developments and the increased use of artificial intelligence, as well as the development regarding laws and regulations, but also deepfakes as a prominent topic of the political debate. Deepfakes are currently a critical topic in the USA, especially concerning the presidential elections, in which the use of deepfakes can be dangerous. Furthermore, the USA is considered a leading nation in the field of technology and has global political influence. Finding out how political actors in the US envision deepfakes in the deepfake discourse, and the threats, challenges, advantages or possible solution strategies that come with it, would provide valuable insights and could serve as an example to give other nations an idea of how they can develop strategies to deal with the phenomenon of deepfakes.

3.3 METHOD OF DATA COLLECTION

In this research, only secondary data will be used. The data used for the analysis include documents of various types: first, political speeches that were given by political actors and which offer immediate insight into the positions of political actors. Second, policy documents, such as official documents in which policies, suggestions and strategies are written down. Third, transcripts from interviews and hearings will be examined which reflect the different ways of thinking of various participants in the deepfake discourse. Fourth, media documents, such as (online) news articles will be used that report on deepfakes to recognize how it is written about the deepfake discourse from a neutral, external perspective. Ultimately, the dataset includes legal texts such as bills and laws that were proposed and passed to regulate the use of deepfakes showing what measures are taken to address the challenges posed by deepfakes. In this analysis, around 60 documents with a total of approx. 5000 pages will be analyzed, some documents with more, some with fewer pages. Authors of these documents include the legislature, the Senate and Congress, journalists and political actors themselves, which makes the data particularly relevant and credible. These types of data are particularly suitable for analyzing and answering the research question because the different types of text contain a lot of content and thus provide insight into different perspectives, attitudes and perceptions regarding deepfakes. The diversity of sources reflects many views from various perspectives to provide a good representation of the deepfake discourse while also making results more diverse, valid and representative. A solution can therefore be found indirectly in legal texts, while direct statements in speeches, for example, provide information about how political actors perceive deepfakes. Furthermore, the secondary data, especially policy documents and transcripts represent a credible and reliable source, particularly when they are collected and accessed from official platforms and archives. To collect the data, various sources are considered, such as government websites and archives of the White House to search for legal texts, policy documents, transcripts or political speeches that contain deepfakes. News portals and their databases are searched for news articles and headlines that revolve around the topic of deepfakes. Data collection is carried out using a systematic search. The search goes from large to small, i.e. from a Google search using keywords such as ‘Deepfake USA’ or ‘policy archive USA’ to websites like official government websites, databases, media articles and public archives, where documents are then filtered in the period from 2020 to 2024. This type of data collection may take some time, as the internet and the websites described all must be searched, but it ensures a diverse and rich result of coverage of the deepfake discourse among political actors in the USA, thus ensuring a reliable, valid and authentic result. This method accumulates a rich data set in which different

perspectives and viewpoints of political actors are presented, which is important for the interpretative analysis of this research to develop a nuanced understanding of how deepfakes are perceived and envisioned.

3.4 METHOD OF DATA ANALYSIS

Content analysis is the selected method for analyzing the data. It is defined as "intellectual process of categorizing qualitative textual data into clusters of similar entities, or conceptual categories, to identify consistent patterns and relationships between variables or themes" (Given, 2008, p.121). Content analysis can be used to identify key content, main topics, messages, statements, visions, ideologies or perspectives by recognizing direct meanings of expressions, but also by identifying and uncovering underlying meanings, attitudes and ideologies in statements. Therefore, content analysis is interpretative by nature as researchers interpret the meaning of the content.

This understanding also fits with my research goal, as this bachelor thesis aims to examine the perspectives of political actors on deepfakes in the deepfake discourse particularly their perceptions, key themes, messages, visions and implicit perspectives of deepfakes in terms of benefits, threats, challenges and solution strategies. Content analysis is a suitable research method for this as it is used to discover (hidden) meaning(s) of expressions by reading a text closely and recognizing the subjective interpretation and contextual dependency (Given, 2008). Content analysis can be done manually but is often conducted by using a research tool, particularly software. In this thesis, ATLAS.ti will be used, a research tool and qualitative data analysis software, that replaces the traditional way of using pen and paper to facilitate the analysis. It serves as a tool to organize, manage and code a large amount of qualitative data, whether textual, graphic, audio or visual material (Given, 2008). The software includes features that make it easier for researchers to annotate and explore their data to make important insights and interpretations from their data (Given, 2008). It offers advantages to the researcher such as speed, consistency, rigor and access to various analytical methods that are not possible by hand. By being able to efficiently upload, organize and analyze the data, and then efficiently code the data, the deepfake discourse can be explored in the best possible way.

With content analysis and ATLAS.ti, the data will be analyzed step by step using certain rules and a coding process. Coding is understood as “representing the operations by which data are broken down, conceptualized and put back together in new ways” (Flick, 2009, p.319). This procedure is a mixture of inductive and deductive coding, which means, on the one hand, using predetermined codes that emerge from the concepts to be found in the theory chapter of this thesis and then looking for excerpts that match these codes and, on the other hand, inductive coding where codes are developed from the data during analysis (Saldana, 2015). The data is

then divided into small analytical units based on the content and categorized into the concepts (Mayring, 2000). This approach offers flexibility and adaptability during the analysis when new results emerge from the data. Nevertheless, not all codes below are used in the analysis, as they did not have any suitable quotes when coding and are therefore not part of the result. After coding the data, all quotations will be collected, looking for patterns in certain codes and interpreting the results. Table 2 provides an overview of the original codes emerging from the concepts of the theory chapter, while Table 3 shows the final coding scheme with additional codes that emerged from the data. *Table 2: Original Coding Scheme*

Concept	**Codes**
Benefits	• Media and Entertainment • Education • Medicine and Healthcare
Threats	• Disinformation and Manipulation • Harm for Democracy and Individual
Challenges	• Social Media Dependence
	• Detection Difficulties
Solution Strategies	• Regulatory Measures • Education and Media Literacy • Anti- Deepfake Technology

Table 3: Final Coding Scheme

Concept	**Codes**
Benefits	• Media and Entertainment • Education • Medicine and Healthcare
Threats	• Disinformation and Manipulation • Harm for Democracy and Individual • Ease of Use • Harm for National Security • Harm for Vulnerable Groups • Pornographic Use • Trust and Integrity • Threat from Other Countries

Challenges	• Social Media Dependence • Detection Difficulties • Regulatory Challenges • Ease of Use
Solution Strategies	• Regulatory Measures • Education and Media Literacy • Anti- Deepfake Technology • Collaboration • Funding Research • Independent Oversight Entity • Reports and Recommendations • Watermarking

3.5 CONCLUSION

This research aims to examine the perceptions of political actors in the USA on deepfakes in terms of benefits, threats, challenges and solution strategies. In summary, the research activity consists of the following steps: first, secondary data is collected for the basis of the analysis, using the procedure described above. Various sources of written documents such as policy documents, news articles, legal texts and transcripts of hearings and speeches related to the deepfake discourse from 2020 to 2024 are used. Second, the coding scheme, which is created using concepts from the theory chapter, is developed to categorize data. Once the coding scheme has been established, the data will be uploaded to ATLAS.ti, a qualitative data analysis software and analyzed using content analysis and the coding scheme to find out what recurring perceptions, statements and (implicit) perspectives political actors have on deepfakes. Content analysis involves coding the data by categorizing important paragraphs based on the previously defined coding scheme, but also developing new codes that emerge from the data. This procedure will provide meaningful insights to answer the research questions. Once analyzed, the results will be collected, interpreted and described in the next chapter.

4. ANALYSIS

4.1 INTRODUCTION

In the following chapter, the results of the content analysis are presented and analyzed using key quotations, structured according to the sub-questions. As expected in the theory chapter, it is confirmed that political actors see deepfakes as having a dual nature, recognizing their beneficial use, as well as their negative connotations and threats on different levels. They also propose many solution strategies to mitigate the negative effects without impairing the positive ones, which shows their awareness of the threats and the benefits. The perception and response of political actors do not differ significantly, and awareness of this threat continues to grow, showing the value of democracy and preventing it from harm. Using the coding scheme developed earlier, various documents that are part of the deepfake discourse in the US between 2020 and 2024 were analyzed according to the benefits, threats, challenges and solution strategies of deepfakes mentioned by political actors. Following this examination of this deepfake discourse, an insight will now be given into the key results of this analysis, interpreting patterns as well as direct and indirect meanings of quotations. The chapter ends by answering the sub-questions.

4.2 DEEPFAKES AS BENEFITS FOR INDUSTRIES

This section aims to show how deepfakes are mentioned as beneficial for certain industries like entertainment and healthcare while political actors already recognize the threats they pose. This presents the dual nature of deepfakes. Based on the results of the analysis, it can be said that the favorable use of deepfakes has developed in two categories: media and entertainment, and medicine and healthcare. Like Westerlund (2019), political actors recognize the value of deepfakes for the media and entertainment industry, especially for humor and parody. According to David Scanlan (as cited in Iyer, 2024): "*Now, I know that there are instances*

where there's a parody and there's humor, and I've seen AI with prominent politicians doing funny things, and it is funny, but it's also quite obvious". This citation illustrates the absurdity of using deepfakes for parody when there is the paradox of using them also to deceive. What is shocking here is the paradox of the use of deepfakes to create a parody of political actors, while simultaneously creating the risk of deception and therefore dangers. Further, Secretary David Scanlan recognizes the humor and obviousness, that deepfakes can bring when it comes to parodying politicians, but his statement provides an important insight: that this obviousness does not always apply, especially with fraudulent deepfakes that are difficult to distinguish from the real.

Rick Allen, (as cited in *Advanced Technology: Examining Threats to National Security: Hearing before the Subcommittee on Emerging Threats and Spending Oversight, 2023)*, a US representative, emphasizes the importance of balancing the benefits of using deepfakes and the malicious uses of deepfakes: *"As I said before, our goal is not to make everything impossible, but our goal is to make malicious activity more difficult and more complicated, while allowing deepfakes for Hollywood movies or other types of entertainment applications to proceed"* (Allen, 2023, p. 23). Funnily enough, Allen sees AI and deepfakes as a normal thing for cultural industry and believes that deepfakes should continue to retain their positive contributions in these areas while aiming to find a balance between beneficial and malicious use. The key here is that this is a nuanced policy proposal to mitigate the risks that deepfakes pose, but on the other hand, maintain their legitimate use in the entertainment industry to encourage innovation, profit, and entertainment, while not considering protecting security. This reveals another key insight that the aim is not to ban deepfakes in principle but to suppress dangerous activities while recognizing the positive impacts, as well as deepfakes' beneficial uses are considered important to maintain.

"They are used to enhance video games and other forms of entertainment, and they are being used to advance medical research as well, but deepfake technology can also be weaponized and cause great harm" (as cited in Miller, 2023, p. 1). Just like Verdoliva (2020), Nancy Mace emphasizes deepfakes as a positive contribution to the entertainment industry and a clear advantage in medical research. What is unexpected here, however, like Rick Allen and David Scanlan, she also underlines the dual nature of deepfakes for beneficial and malicious uses. Above all, the versatility and complexity of deepfakes can be recognized here, as well as the relevance of distinguishing between advantages and disadvantages in the policy discourse. By discussing potential dangers and the metaphor of 'weaponization', Mace aims to set up

regulatory measures to curb misuse and promote beneficial innovations, as deepfakes are seen as a kind of weapon to attack democracy with.

Cathy McMorris Rodgers, a US representative, speaks in favor of deepfakes and concentrates on the benefits in the field of medical research: *"For example, AI technology and deep learning algorithms can help us detect cancers earlier and more quickly. Clinical trials are already underway and making major breakthroughs to diagnose cancers"* (as cited in *Americans At Risk: Manipulation And Deception In The Digital Age: Hearing Before The Subcommittee On Consumer Protection And Commerce,* 2020, p.4). As Westerlund (2019) shows and what is at the heart of the matter is that deepfakes are also helpful for a revolution in the medical field by advancing diseases and making diagnoses earlier - deepfakes can be a life-saving aid in this area, but also help to make a profit.

In sum, deepfakes in the deepfake discourse are particularly valued in the media and entertainment industry and are seen as significant contributions to medical research and practical application in this field. Nevertheless, despite their positive application, concerns in the deepfake discourse are already being expressed here by political actors, who are calling for consideration of the duality of deepfakes when developing solution strategies and setting up appropriate measures.

4.3 DEEPFAKES AS THREAT TO DEMOCRACY, SOCIETY AND INDIVIDUAL

The analysis already revealed some benefits of using deepfakes, but there are also many significant threats arising from their creation and use. The most important threats are analyzed and elaborated below, including threats to democracy, society and individuals. These include the erosion of trust, threats to national security, manipulation, the spread of false information and personal threats, showing that the threats outweigh the positive uses of deepfakes.

As Vasist & Krishnan (2022) already highlighted, the analysis found that the threats posed by deepfakes have developed at macro, meso and micro levels: threats to democracy, society and individuals. A key observation made during the analysis was that deepfakes can manipulate in various ways, for example through misinformation by containing convincing content that seems so real that it is hard to distinguish it from the truth. *"The threat that deepfakes could pose if used in misinformation campaigns is well known and well-documented. The Congressional Research Service (CRS) has noted that "state adversaries or politically motivated individuals could release falsified videos of elected officials or other public figures making incendiary comments or behaving inappropriately"* (as cited in Committee on Homeland Security and

Governmental Affairs, 2022, p2). Spreading misinformation can therefore have serious consequences for democracy, society, and individuals, leading to other threats.

Given the research of many scholars like Verdoliva (2020) or Diakopoulos and Johnson (2020), it was expected that political actors would see deepfakes as a threat, especially in terms of manipulating public opinion during the election process. For example, Donald Trump sees this above all as a threat to the democratic electoral process and warns against it, as it can influence voter behavior, public opinion and electoral integrity: *"As the world approaches another election cycle, Trump raises a pertinent concern about the malicious use of AI to influence voters. He warns that in an election year, AI could be used to maliciously generate images, voices and other forms of content to manipulate public opinion"* (Gil, 2024). This illustrates another key insight that deepfakes can spread misinformation, especially during election campaigns, which in turn can harm the electoral process, and confuse and mislead voters. Strangely enough, misinformation also leads to a loss of trust in democratic processes, political actors, etc. On the one hand, Donald Trump emphasizes the relevance of regulatory measures to precisely prevent this misuse of AI. On the other hand, it also shows how vulnerable democratic processes can be when it comes to technological manipulation.

As already surmised in the theory section, deepfakes pose a further threat, particularly to national security, as they have the potential to trigger physical conflict and violence: *"In the interview, Trump called for action regarding AI and AI-generated deepfakes, raising the concern that the technology could be used to start wars"* (as cited in Nelson, 2024a). Here too, Donald Trump deliberately emphasizes the relevance and urgency of acting against AIgenerated content. Vehemently, he also makes his concerns clear here that deepfakes represent much more than just influencing voter behavior but pose a real threat to a country and its security. As Mustak et al. (2023) argue, a key observation here is that deepfakes threaten national or global security by provoking and causing wars, physical violence, or diplomatic conflict. Donald Trump expects that if measures are not taken immediately against the creation of deepfakes, they could have the worst consequences. What is probably unexpected here is that deepfakes have more implications than one would probably think. On the one hand, one might think that deepfakes to start wars are a logical consequence of the spread of misinformation, but on the other hand, it is even more frightening how little has been done so far against deepfakes to prevent the greatest risks. The unexpected absurdity here is that deepfakes were initially "only" used for pornographic purposes, and now they pose such a great threat to national security that even wars could be started with them.

It was also expected that foreign actors and governments would use deepfakes to cause unrest in other countries. This was also confirmed by the analysis: "*Foreign governments and fraudsters have pursued political disinformation campaigns using new technology like deepfakes designed to sow civil unrest and disrupt democratic elections"* (as cited in *Fostering A Healthier Internet To Protect Consumers: Joint Hearing Before The Subcommittee On Communications And Technology And The Subcommittee On Consumer Protection And Commerce*, 2020, p.12). This citation by a US representative shows that deepfakes are strategically used by foreign governments to spread political disinformation campaigns, causing unrest among the population, but also to disrupt democratic elections, i.e. to manipulate and destabilize them. All of this in turn leads to a loss of trust. This fact, too, calls for a stronger response to protect democracy and to prevent and punish actors who create deepfakes for malicious acts. Nonetheless, ambiguity seems to prevail here, with no clear consensus yet on how to effectively combat deepfakes of this kind without restricting other rights such as free speech, etc.

Ultimately, a key insight is that deepfakes also pose a personal threat when individuals, especially vulnerable groups, are attacked, harassed and threatened by this AI-generated technology. As elaborated by Westerlund (2019) and Verdoliva (2020), deepfakes can lead to significant psychological and social damage, especially in pornographic use. *"The nonconsensual sharing of synthetic intimate images, often called deepfake pornography, is an exponentially growing form of abuse that humiliates, degrades and threatens women and girls, causing real and lasting damage to their mental health and wellbeing"* (as cited in *Congressman Joe Morelle Announces New Support For His Legislation To Stop Deepfake Pornography*, 2024). This quote emphasizes the use of deepfakes by publicly shaming vulnerable groups such as women, young girls, children, or other (religious) minorities without their consent. It also emphasizes the burden that this type of abuse brings with it, which can lead to long-term mental health consequences and threaten one's social life. The dissemination of non-consensual synthetic intimate images is therefore on the rise and so are the dangerous effects on their victims. Here too, a call is made for more regulatory or technological measures to be introduced to protect victims from this abuse and to support them in asserting their rights. Since this new form of online shaming also poses a great danger, especially for children and young girls who are not aware of it, it is even more frightening that there are few regulatory responses to this problem. Although awareness of the threats posed by deepfakes of all kinds is evident, too little has been done to curb the negative effects.

In sum, these are the threats posed by deepfakes that political actors recognize in the deepfake discourse in the USA. These have developed in two directions, namely threats to democracy, society and on the individual level, whereby misinformation is the biggest threat that leads to all other threats, such as loss of trust, psychological damage to individuals, manipulation or conflict. The important call for strategies and solutions to preserve democracy, ensure the safety of individuals and guarantee international stability and peace against the malicious use of deepfakes, is becoming ever clearer.

4.4 DEEPFAKE CHALLENGES: LEGAL, DETECTION AND ACCESSIBILITY ISSUES

During the analysis, three primary challenges emerged regarding deepfakes. regulatory challenges, easy access to create deepfakes, and the detection of deepfakes, whether by humans or by existing tools. Political actors see, as expected, the first key insight that existing laws and regulations are rather inappropriate and therefore new ones should be established: "*The interviewees, including Trump and Gensler, emphasize the inadequacy of existing legal frameworks in addressing the risks posed by deepfakes, signaling the need for innovative and adaptive regulatory approaches"* (Adebayo, 2024). Like Ray (2021), both Donald Trump and Cary Gensler recognize that legal frameworks are not sufficient to combat the threats posed by deepfakes. The challenge here for political actors and the legislative is therefore to draw up new and more effective laws and regulations. The absurdity here is that while they recognize the danger of deepfakes, regulatory measures to combat these threats are missing. Furthermore, the dual complexity of regulatory measures becomes visible: *"The hearing concluded with a call for action to address the challenges posed by deepfake technology, while also being mindful of protecting civil liberties and privacy"* (Miller, 2023). This citation illustrates the paradox that there is a consensus among political actors on the need for stricter regulations, but on the other hand, political actors, surprisingly, emphasize the relevance and specific challenge of balancing these regulators, as much emphasis is placed on the special consideration and protection of civil liberties and rights, as well as privacy. This also means that while deepfakes can be a major threat that can cause harm, political actors do not want to restrict civil rights and freedoms. We can therefore speak here of a dual nature of regulatory measures against deepfakes, which are almost paradoxical, as a call for strict regulations is usually always in conflict with the preservation of civil rights and freedoms. This also reflects the complexity of regulations etc., as it is necessary to balance different interests despite this danger. This can also lead to a real

challenge at this point, that an effective legal framework is set up to combat deepfakes but is also robust enough to preserve civil rights.

Further, the analysis of the deepfake discourse revealed that the detection challenge can move in two directions: on the one hand, the difficult for the human eye to detect deepfakes, and on the other hand, the inefficient and underdeveloped deepfake detection tools. This is accompanied by the observation of Nguyen et al. (2022). Although there is a lot of research in this area, they are becoming increasingly difficult to recognize due to their high quality: *"What you say at an interview with you almost doesn't matter anymore," Trump said. "They can change things around, and nobody can tell the difference; even experts can't tell the difference. This is a tremendous problem in terms of security"* (as cited in Nelson, 2024a). The shocking statement here is that his concern here is that in the end, no one can distinguish the real from the fake because their quality is so high - not even experts, which presumably include developers of deepfake detection tools or people with a high awareness of deepfakes. He particularly emphasizes the security challenges that arise from this. On the other hand, this quote from Trump also shows that if nobody can tell the difference, there is probably a lack of reliable deepfake detection algorithms that nobody seems to have access to or that there are no methods. What comes unexpectedly is the paradox that this also reveals no matter how educated the population is about deepfakes and especially their fraudulent dangers, it is of little use, as the deepfakes look deceptively real these days.

Tim Harper, a senior policy analyst told the Agency France Press: *"Of all the surfaces—video, image, audio—that AI can be used for voter suppression, audio is the biggest vulnerability. It is easy to clone a voice using AI, and it is difficult to identify"* (as cited in Chopra, 2024). This reveals the specificity of audio in the context of AI-generated fake material, but also how easily they can be created, making it impossible to distinguish between true and false. Although it was expected that there would be detection difficulties, it was rather unexpected that the focus was especially on audio. Therefore, unexpectedly Tim Harper draws attention to the dangers of audio deepfakes. Unexpectedly, this also presents a new challenge, namely the audio recognition of false material, which is often overseen. The paradox here is that while the technology continues to develop and improve in terms of quality and detection, as well as becoming more accessible to the general public, reliable tools for detecting them are still lagging far behind, as Chesney & Citron (2018) also state in their research.

The increasing ease of creating deepfakes also represents a major challenge for political actors. "*Recent technological advancements in artificial intelligence have opened the door for bad actors with very little technical knowledge to create deepfakes cheaply and easily"* (as cited in Miller, 2024). This reveals that, above all, access to powerful deepfake creation tools is becoming ever more widespread and access is no longer limited to experts. Paradoxically, while the accessibility of deepfake creation tools can be used for good, there is a potential for misuse by malicious actors. This challenge was initially unexpected in terms of theory, but again shows the paradox and dual nature of deepfakes, in that greater accessibility to the creation of deepfakes can be used for positive uses, for example for the media and entertainment industry, or to express one's creativity, but can also pose a great danger when dealing with maliciously created deepfakes. As a side point, unlike Fletcher (2018), political actors do not see the social media dependence for communication and information gathering as a challenge to them, that contributes to the fast spread of disinformation.

In sum, the content analysis uncovers three primary challenges in the deepfake discourse; political actors in the USA see regulatory challenges, which are mainly caused by insufficient existing legal frameworks, but also the call to protect civil rights and freedoms by the legislative and political actors. Secondly, they see detection difficulties due to the high quality, whether for the naked human eye or detection tools, and finally the increasingly easy access to platforms with which deepfakes can be created cheaply without any experience, which can lead to positive use, but also to a danger if deepfakes want to deliberately attack something.

4.5 SOLUTIONS: A MULTI-FACETED APPROACH TO COMBAT DEEPFAKES

This section shows that a multi-faceted approach is necessary, including collaboration, technological inventions, education, regulatory measures, and funded research to combat the negative effects of deepfakes. Several solutions have been identified and implemented in the deepfake discourse, while there are increasing calls for further action against deepfakes.

The first quote reveals some possible solutions to combat the malicious use of deepfakes: *"Collaboration between Congress and technology companies is essential, and I am glad to see that that is happening to address the challenges posed by deepfakes. Tech companies should be responsible for developing and implementing the policies to detect and mitigate this type of content, including what we were hearing today, on their platforms and sharing, most importantly, what they learn with others. […]. More robust privacy and consent laws are needed to protect individuals from using their likeness and voice in deepfake content without their permission, and continued research and development in AI deepfake technology are necessary,*

as is funding to counter deepfake misuse" (as cited in *Advances In Deepfake Technology: Hearing Before The Subcommittee On Cybersecurity, Information Technology, And Government Innovation*, 2023, p.8). As expected in the theory section, political actors place particular emphasis on regulatory measures and anti-deepfake technology when discussing solution strategies. Johnson & Diakopoulos (2021) highlight the efforts of multiple stakeholders, such as social media platforms, policy and lawmakers, while political actors consider collaboration with technical experts as crucial to mitigate the negative effects of deepfakes. The emphasis is particularly on cooperation, especially between the legislative branch and technology companies, and its relevance in combating deepfakes. While Westerlund (2019) marks that technology companies are needed to develop anti-deepfake technologies, such as deepfake detection tools, Johnson & Diakopoulos (2021) assert that they are necessary for developing appropriate policies. Strangely enough and paradoxically is that the responsibility in this citation to develop effective solutions and anti-deepfake technology is taken away from the government and shared with tech companies, which are responsible for a small part that deepfakes exist, can be created and spread. The aim is to work together to share knowledge and develop solutions, as is already the case in Congress through various hearings.

The second quote reflects the previously mentioned solution approach by Westerlund (2019) of anti-deepfake technology and deepfake detection tools and a call for legislative action. It also represents a new approach, which has also been recognized by scholars like Diakopoulos and Johnson (2020), namely the relevance of public awareness, education and digital literacy. "*There was a consensus on the urgent need for both public awareness and legislative action to combat the misuse of deepfake technology, with suggestions for incorporating digital literacy into education and developing technologies to detect and mark AI-generated content"* (Miller, 2023). This consensus encompasses the importance of digital literacy in educational measures to recognize deepfakes, become aware of them and thus prevent worse effects. The proposal to develop technologies that detect deepfakes, but also to label AI-generated content illustrates the efforts to ensure transparency and authenticity so that the population can take a critical stance towards this content. Still, the urgent call for legislative action and evolving technologies shows that these solutions are at an underdeveloped stage. Paradoxically, at least ten US states have already introduced several deepfake-related laws, such as the National Defense Authorization Act, the Deepfake Report Act, or the Identifying Outputs of Generative Adversarial Networks Act. This shows that political actors support education campaigns, legislative action and technological solutions.

In the marking of AI-generated content, political actors in the current government have a precise idea: "*In February, the Biden Administration said it would use watermarking and cryptography to fight political disinformation"* (Nelson, 2024b). The administration's use of watermarking and cryptography proves an active approach to maintaining authenticity and transparency in AI-generated content found online, reducing the potential for deepfakes to manipulate and influence the population. Surprisingly, watermarking is already one of the more advanced technologies with deepfake detection tools to reduce the far-reaching dangerous consequences of deepfakes. What is surprising here is that the administration itself is responsible for developing and specifying these methods. Paradoxically, however, the effectiveness and handling of these methods remain unclear, especially regarding their reliability and scalability.

A further quote illustrates other possible solutions: *"Specifically, the National Institute of Standards and Technology (NIST) U.S. AI Safety Institute will create guidelines, tools, benchmarks, and best practices to evaluate and mitigate risks using red-teaming and other methods. The Institute will develop technical guidance on issues such as authenticating content created by humans, watermarking AI-generated content, identifying and mitigating against harmful algorithmic discrimination and the creation of harmful synthetic content (such as deepfake image-based abuse), ensuring transparency, and enabling adoption of privacypreserving AI, and would support the development of safe and trusted AI and support enforcement agencies' efforts to protect the American public' rights and safety"* (The White House, 2024c). The National Institute of Standards and Technology (NIST) U.S. AI Safety Institute are considered an oversight entity that researches deepfakes and AI-generated content on behalf of the government, providing reports to inform the government and makes appropriate recommendations for action and best practices based on its observations, as well as evaluating these. NIST is also responsible for authenticating content, for example through watermarking, and thus developing technical assistance regarding progress in anti-deepfake technology. With the introduction of such an institution, it is recognized that something is already being done in this regard by implementing several solution strategies through one institution to combat the malicious use of deepfakes. This multi-faceted approach shows that multiple solutions are needed to curb these AI-generated media by establishing this institution. Political actors aim to develop AI that is privacy-preserving, secure and trustworthy. This means that political actors are also interested not in making AI disappear completely, but in exploiting its advantages and removing any negative consequences from the picture.

In sum, all quotes show that political actors are of the opinion that several solutions are needed to finally curb the malicious use of deepfakes or to minimize or even eliminate its negative effects, such as through cooperation between government and technical experts, regulatory measures such as laws, education and digital literacy for the population, technology that detects or flags deepfakes, or institutions that make recommendations and take action through funded research. These solution strategies emphasize once again the complexity of deepfakes and their threats and the effort required from various stakeholders to combat them.

4.6 WHAT ARE THE ENVISIONED BENEFITS, THREATS, CHALLENGES AND SOLUTIONS

To answer sub-question one regarding *benefits, threats and challenges that political actors envision about deepfakes*, the analysis reveals that political actors particularly recognize the dual nature of deepfakes. On the one hand, the positive impacts are in the areas of media and entertainment, as well as medicine and healthcare. In media and entertainment, they recognize the use of creating humorous and parodic content and advances in medical research as well as early cancer detection. Still, the malicious use of deepfakes and their threats are also repeatedly highlighted. On the one hand, there can be a threat to democracy and society, for example through misinformation that can lead to the manipulation of public opinion and thus the manipulation of the electoral process. Misinformation can also lead to a loss of public trust in democracy, political actors, and the media. Furthermore, deepfakes pose a threat if they are used by foreign governments and fraudsters to cause unrest and destabilization, or even to endanger national security by provoking conflicts and violence, including wars. On the other hand, political actors also recognize a threat to individuals in the discourse. For example, when a vulnerable group such as women, children, or other minorities fall victim to non-consensual pornographic content and thus suffer long-term psychological and social consequences.

Deepfakes come with some challenges when it comes to mitigating their threats through solution strategies. On the one hand, political actors still see a regulatory challenge, as they lament the inadequacy of the current legal framework and call for more efficient and adaptable legislative measures. However, they face the challenge of reconciling the complexity of enacting effective laws against malicious deepfakes with the protection of civil rights and freedoms. According to political actors, another difficulty is the detection of deepfakes. On the one hand, it is difficult for the human eye, even experts, to recognize deepfakes with everimproving quality, and on the other hand, there are underdeveloped and inefficient detection tools. This is especially true for audio deepfakes, which are particularly difficult to recognize.

Another challenge presents the ease of access to deepfake creation tools. It requires little technical knowledge; it is cheap and easy to use. This favors the creation of further deepfakes for malicious purposes.

Regarding sub-question two on *possible solutions to combat the threats and challenges posed by deepfakes*, which political actors envision in the deepfake discourse, it can be said that political actors identify a multifaceted approach that includes many possible solutions to combat deepfakes. There is no single solution, some of which have already been implemented more, others less. On the one hand, political actors emphasize cooperation in the implementation of solution strategies, especially between the legislative branch and technology companies or technical experts. Technology companies help to share their knowledge and thus develop best practices. Furthermore, technological solutions such as deepfake detection tools as well as the watermarking of AI-generated content are considered important. Watermarking is the keyword here to enable transparency and authenticity in digital media. Political actors repeatedly emphasize the relevance of more regulatory measures to protect individuals and prohibit deepfakes for malicious purposes. Nevertheless, some laws already exist at the government level to curb deepfakes and negative consequences, while at least ten US states are already enacting laws themselves. A less developed solution, but one recognized by many political actors, is to incorporate deepfakes into education and digital literacy to create public awareness, and educate the public about deepfakes, their potential negative effects, how to identify them, and how to deal with them. This can help individuals to think critically about digital content and minimize potential consequences as much as possible.

Another approach is to work with oversight entities, i.e. institutions that are mandated by law to carry out continuous research and development in the field of deepfake technology. The government also places great importance on promoting and investing in research. This helps to better understand deepfake technology and its impact, and based on these results, government institutions can provide reports with recommendations for action and general information on the status quo.

5. CONCLUSION

5.1 THE ENVISIONING OF DEEPFAKES BY POLITICAL ACTORS

So, what is the answer to the main research question of *how political actors in the USA envision deepfakes in the deepfake discourse in the period from 2020 to 2024?* In general, the results of the analysis show that political actors in the USA see deepfakes as dual in nature: on the one hand, they recognize their potential benefits, such as for media and entertainment, but also in medical research and healthcare, like Nguyen et al. (2022). On the other hand, the discourse predominantly focused on the threats posed by deepfakes that have many faces. This includes the threat to democracy, which can manipulate public opinion by spreading false information, disrupt democratic processes, threaten national security and cause a loss of public trust in democracy, describing the same what Misirlis & Munawar (2023) and Vasist & Krishnan (2022) describe. Furthermore, just like Kshetri (2023) underlines, personal threats are perceived in the discourse through the dissemination of non-consensual pornographic content that victimizes vulnerable groups and has long-term social and psychological consequences. As Diakopoulos & Johnson (2021) state, the discourse shows a dichotomous perspective, emphasizing the need to maintain the positive applications while urgently calling for action to combat the risks and challenges of deepfakes. They repeatedly emphasize the relevance of a robust strategy to prevent malicious use. This requires a multi-faceted approach with more than one solution. Their envisioned solution strategy includes collaboration between government and technology companies, technological innovation to develop deepfake detection tools and watermarking, regulatory measures to mitigate the negative consequences and protect victims, educational initiatives for society to raise awareness of deepfakes and research into the technology to identify potential risks, like Westerlund (2019) suggests.

Ultimately, however, the threats associated with deepfakes surpass their potential benefits, unlike the more balanced vision of Verdoliva (2020). The rapid development of this technology requires constant attention and quick action to prevent worst-case scenarios and to be able to react at an early stage. Although it was known, especially by experts and authors of scientific articles, that deepfakes have negative effects that have long since become reality, political actors

in the USA have now recognized the danger and are discussing possible solution strategies. The proactivity emanating from political actors also shows a willingness to contain this danger. Nevertheless, it remains important to keep an eye on this AI-generated technology to be able to constantly adapt.

5.2 SUGGESTIONS FOR FUTURE RESEARCH

How did this research fill the knowledge gap? This research conducted an analysis of how political actors in the USA envision deepfakes in the deepfake discourse. While previous research on deepfakes was rather limited, some social scientists slowly began to explore this topic. The three main areas of research and published literature were examined by Vasist & Krishnan (2022), as well as Misirlis & Munawar (2023): deepfake detection and creation, risks and threats to society and democracy, especially in the USA, and ethical aspects of deepfake technologies. This bachelor thesis delved deeper into perspectives of political actors in the US, including the benefits, threats and challenges they see in deepfakes, as well as solution strategies to combat negative effects. This research therefore focused more on positive impacts that deepfakes can have and solution strategies proposed by policy actors that are not only related to anti-deepfake technology, which has otherwise received less attention in research. This gap was filled by this research, which not only dealt with facts, but also with the perspectives of individuals, how they view deepfakes and not how deepfakes are empirically categorized as dangerous. The research conducted in this bachelor thesis has complemented the existing literature by providing a multifaceted view of deepfakes from political actors that deals with more than just threat and anti-deepfake technology. It showed the relevance that interdisciplinary approaches are important to understand the deepfake phenomenon in all its facets.

Like Diakopoulos & Johnson (2021), the result of the analysis illustrates the danger of deepfakes for democratic processes such as elections or national security, as well as the positive effects of deepfakes on medicine. Ray (2021) emphasizes in his research that deepfakes have already influenced elections. Although this fear applies to the US in the 2024 presidential election, it has not yet been proven to have influenced elections. Nevertheless, Chesney & Citron (2018) agree with the results of the analysis that deepfakes pose a threat to democracy through possible election manipulation and political sabotage. Furthermore, the analysis also agrees with Westerlund (2019) and Johnson & Diakopoulos (2021) as they support a multistakeholder approach, which the analysis also reveals. Unlike Fletcher (2018), the analysis

does not uncover the challenge of the reliance on social media platforms that contribute to the spread of misinformation.

For future research, it would be interesting to explore and evaluate the effectiveness and the success of solution strategies such as deepfake detection technologies, the use of educational campaigns or regulatory measures, as understanding the effectiveness of these strategies would provide valuable insights into how to combat the malicious uses of deepfakes in the best way. This can also be helpful for policymakers to provide guidance on refining and further improve their policies. Furthermore, there is also an interesting question for international co-operation. Whilst this bachelor thesis only focused on the USA, it would be interesting for the future to know how international cooperation is used to counter deepfake threats and which international policies are used for this purpose. This could show similarities and differences in governing and could pave the way for a common approach to manage the threat posed by deepfakes across borders.

5.3 PRACTICAL IMPLICATIONS FOR POLICY AND GOVERNANCE

So, what needs to be done by policymakers? Based on the insights developed, it is important that the government, including certain government agencies like ministries or departments, set up a deepfake policy. That includes promoting and subsidizing cooperation and involving technology companies such as Google, Meta or technical experts to develop strategies for mitigating deepfakes and develop detection tools. Furthermore, there should be support and subsidization for research done by universities and research centers on deepfake technology to understand its impact and develop effective countermeasures. Additionally, investments in education and digital literacy campaigns are essential, integrating digital literacy in schools and setting up public awareness campaigns. Therefore, education ministries are responsible for integrating media literacy in schools, while advertising agencies should set up public awareness campaigns to build awareness in public about deepfakes. The analysis has also shown that legislators, like the Senate should introduce and develop regulatory measures to address the threats posed by deepfakes while balancing civil rights. In addition, existing regulations should be revised and adapted by legislators. Regulatory bodies should act as agencies that enforce the regulations. Ultimately, the government should treat deepfakes seriously as a threat to democracy and individuals and act proactively to precisely address these threats to the democratic process. In general, policy development and evaluation should always be reflected upon, assessed for effectiveness, and updated or strategies should be sensibly adapted to keep

pace with the technological development of deepfakes and to be able to contain spreading threats in good time. In general, if the government takes appropriate measures in good time, the worst consequences such as the spread of misinformation and the resulting dangers for democratic processes and loss of trust should still be able to be stopped.

6. LIST OF REFERENCES

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 Cook, G. (1989). *Discourse*. Oxford University Press.

Diakopoulos, N., & Johnson, D. (2021). Anticipating and addressing the ethical implications of deepfakes in the context of elections. *New Media & Society*, *23*(7), 2072–2098. https://doi.org/10.1177/1461444820925811

Dobber, T., Metoui, N., Trilling, D., Helberger, N., & de Vreese, C. (2021). Do (microtargeted) deepfakes have real effects on political attitudes? *Politics [The International Journal of Press]*, *26*(1), 69–91. https://doi.org/10.1177/1940161220944364

Fletcher, J. (2018). Deepfakes, artificial intelligence, and some kind of dystopia: The new faces of online post-fact performance. *Theatre Journal*, *70*(4), 455–471. https://doi.org/10.1353/tj.2018.0097

Flick, U. (2009). *An Introduction to Qualitative Research* (4. Aufl.). SAGE Publications.

Given, L. (2008). *The SAGE encyclopedia of qualitative research methods*. SAGE Publications, Inc.

Guardado, M. (2019). *Discourse, ideology and heritage language socialization: Micro and macro perspectives*. De Gruyter Mouton.

Henry, F., & Tator, C. (2002). *Discourses of domination: Racial bias in the Canadian Englishlanguage press*. University of Toronto Press.

Johnson, D. G., & Diakopoulos, N. (2021). What to do about deepfakes. *Communications of the ACM*, *64*(3), 33–35. https://doi.org/10.1145/3447255

Kshetri, N. (2023). The economics of deepfakes. *Computer*, *56*(8), 89–94. https://doi.org/10.1109/mc.2023.3276068

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

Mayring, P. (2000). *Qualitative Content Analysis*. https://doi.org/10.17169/fqs-1.2.1089

Meneses, J. P. (2021). *Deepfakes and the 2020 US elections: what (did not) happen*. https://doi.org/10.48550/ARXIV.2101.09092

Misirlis, N., & Munawar, H. B. (2023). *From deepfake to deep useful: risks and opportunities through a systematic literature review*. https://doi.org/10.48550/ARXIV.2311.15809

Mustak, M., Salminen, J., Mäntymäki, M., Rahman, A., & Dwivedi, Y. K. (2023). Deepfakes: Deceptions, mitigations, and opportunities. *Journal of Business Research*, *154*(113368), 113368. https://doi.org/10.1016/j.jbusres.2022.113368

Vasist, P., Indian Institute of Management Kozhikode, Krishnan, S., & Indian Institute of Management Kozhikode. (2022). Deepfakes: An integrative review of the literature and an agenda for future research. *Communications of the Association for Information Systems*, *51*, 590–636. https://doi.org/10.17705/1cais.05126

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

Nordquist, R. (2009). *Definition and examples of discourse*. ThoughtCo. https://www.thoughtco.com/discourse-language-term-1690464 Last retrieved: 02 July 2024

Pawelec, M. (2022). Deepfakes and democracy (theory): How synthetic audio-visual media for disinformation and hate speech threaten core democratic functions. *Digital Society: Ethics, Socio-Legal and Governance of Digital Technology*, *1*(2). https://doi.org/10.1007/s44206-022-00010-6

Payne, L. (2024). Deepfake I History & Facts. In *Encyclopedia Britannica*. https://www.britannica.com/technology/deepfake Last retrieved: 02 July 2024

Pulla, V., & Carter, E. (2018). Employing interpretivism in social work research. *International Journal of Social Work and Human Services Practice*, *6*(1), 9–14. https://doi.org/10.13189/ijrh.2018.060102

Ray, A. (2021). Disinformation, deepfakes and democracies: The need for legislative reform. *The University of New South Wales law journal*, *44*(3). https://doi.org/10.53637/dels2700

Saldana, J. M. (2015). *The coding manual for qualitative researchers* (3. Aufl.). SAGE Publications.

van Dijk, T. A. (1997). *Discourse as structure and process*. http://ci.nii.ac.jp/ncid/BA30274676

Verdoliva, L. (2020). Media forensics and DeepFakes: An overview. *IEEE journal of selected topics in signal processing*, *14*(5), 910–932. https://doi.org/10.1109/jstsp.2020.3002101

Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology innovation management review*, *9*(11), 39–52. https://doi.org/10.22215/timreview/1282

Wodak, R., & Meyer, M. (2009). Critical discourse analysis: history, agenda, theory, and methodology. In R. Wodak & M. Meyer (Hrsg.), *Methods for Critical Discourse Analysis* (S. 33). Sage (2nd revised edition).

Zannettou, S., Sirivianos, M., Blackburn, J., & Kourtellis, N. (2019). The Web of false information: Rumors, fake news, hoaxes, clickbait, and various other shenanigans. *ACM Journal of Data and Information Quality*, *11*(3), 1–37. https://doi.org/10.1145/3309699

7. DATA APPENDIX

7.1 OVERVIEW OF DATA

POLICY DOCUMENTS

The White House. (2020). *Bill Announcement*. whitehouse.archives.gov. https://trumpwhitehouse.archives.gov/briefingsstatements/bill-announcement-122320/ Last retrieved: 02 July 2024

The White House. (2023). *Readout of White House State Legislative Convening on Non-Consensually Distributed Intimate Images*. whitehouse.gov. https://www.whitehouse.gov/briefing-room/statementsreleases/2023/04/26/readout-of-white-house-state-legislative-convening-on-nonconsensually-distributed-intimate-images/#:~:text=Today%2C%20Jennifer%20Klein%2C%20Assistant%20to,survivors%20of%20non%2Dconsensually%20distributed Last retrieved: 02 July 2024

The White House. (2024a). *A Call to Action to Combat Image-Based Sexual Abuse*. whitehouse.gov. https://www.whitehouse.gov/ostp/news-updates/2024/05/23/a-call-toaction-to-combat-image-based-sexual-abuse/ Last retrieved: 02 July 2024

The White House. (2024b). *A Call to Service for AI Talent in the Federal Government*. whitehouse.gov. https://www.whitehouse.gov/ostp/news-updates/2024/01/29/a-call-toservice-for-ai-talent-in-the-federal-government/ Last retrieved: 02 July 2024

The White House. (2024c). *FACT SHEET: The President's budget advances President Biden's unity agenda*. whitehouse.gov. https://www.whitehouse.gov/omb/briefingroom/2024/03/11/fact-sheet-the-presidents-budget-advances-president-bidens-unityagenda/ Last retrieved: 02 July 2024

The White House. (2024d). *U.S-EU Joint Statement of the Trade and Technology Council*. whitehouse.gov. https://www.whitehouse.gov/briefing-room/statementsreleases/2024/04/05/u-s-eu-joint-statement-of-the-trade-and-technology-council-3/ Last retrieved: 02 July 2024

MEDIA ARTICLES

Adebayo, O. (2024, February 3). Donald Trump brands AI as 'scary' amid rampant deepfakes. *UserInterface*. https://www.cryptopolitan.com/donald-trump-ai-scaryamid-rampant-deepfakes/ Last retrieved: 02 July 2024

Alexis, A. (2024, March 12). AI deepfakes targeted in Biden's State of the Union speech. *CFO Dive*. https://www.cfodive.com/news/ai-deepfakes-targeted-bidens-stateunion-speech-artificial-intelligence-technology/709952/#:~:text=%E2%80%9CBan%20AI%20voice%20impersonations%20and,wreak%20havoc%20in%20U.S.%20elections. Last retrieved: 02 July 2024

Briscoe, S. (2021, January 12). U.S. Laws Address Deepfakes. *ASIS*. https://www.asisonline.org/security-management-magazine/latestnews/today-in-security/2021/january/U-S-Laws-Address-Deepfakes/ Last retrieved 02 July 2024 Last retrieved: 02 July 2024

Burga, S. (2024, February 1). How a New Bill Could Protect Against Deepfakes. *Yahoo!Movies*. https://ca.movies.yahoo.com/news/bill-could-protectagainstdeepfakes%20214513999.html?guccounter=1&guce_referrer=aHR0cHM6Ly9 3d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAH5SoK49VR-YeUq5hNLwHVOILfaZhYpkS5sILmD8FwA1WnXZ0K5Q8798gqQLVRLzYvG5ANrsKGpooYpa-_mtm919Wf1NQ-XgPFsy3zxi76XmFCsRam7ZXEbRoDKpHrbKfJg4ubTJN__egPYi3nl17MVG4oSZNbhb92pwC0fv1Xeb%20Last%20retrieved%2002%20July%202024 Last retrieved: 02 July 2024

Chopra, A. (2024, February 5). Biden Robocall: Audio Deepfake Fuels Election Disinformation Fears. *Barron's*. https://www.barrons.com/news/biden-robocall-audio-deepfake-fuelselection-disinformation-fears-aa5baf03 Last retrieved: 02 July 2024

Durbin, R. J., Graham, L., & The bipartisan Disrupt Explicit Forged Images and Non-Consensual Edits Act of 2024. (2024). *The DEFIANCE Act of*

2024. https://www.durbin.senate.gov/imo/media/doc/defiance_act_of_2024.pdf Last retrieved: 02 July 2024

Edelman, A. (2024, January 25). *Growing concerns amid rise of deepfakes of 2024 presidential candidates* [Video]. NBC News. https://www.nbcnews.com/politics/states-turnattention-regulating-ai-deepfakes-2024-rcna135122 Last retrieved: 02 July 2024

Federal Trade Commission. (2024, February 15). *FTC proposes new protections to combat AI impersonation of individuals*. https://www.ftc.gov/news-events/news/pressreleases/2024/02/ftc-proposes-new-protections-combat-ai-impersonation-individuals Last retrieved: 02 July 2024

Ferraro, M. F. (2020, December 29). Congress's deepening interest in deepfakes. *The Hill*. https://thehill.com/opinion/cybersecurity/531911-congresss-deepening-interestin-deepfakes/ Last retrieved: 02 July 2024

Galston, W. A. (2020, January 8). Is seeing still believing? The deepfake challenge to truth in politics. *Brookings*. https://www.brookings.edu/articles/is-seeing-still-believing-thedeepfake-challenge-to-truth-in-politics/ Last retrieved: 02 July 2024

Gil, J. (2024, February 4). Donald Trump warns of the dangers of Artificial Intelligence and Deepfakes. *Geek Metaverse News*. https://www.geekmetaverse.com/donald-trumpdangers-artificial-intelligence-deepfakes/ Last retrieved: 02 July 2024

Harris, E. (2024, May 23). *Deepfake Laws: A Comprehensive Overview*. Plural Policy. https://pluralpolicy.com/blog/deepfake-laws/ Last retrieved: 02 July 2024

Higham, A. (2024, March 4). Donald Trump deepfakes with Black voters fuel outrage. *Newsweek*. https://www.newsweek.com/donald-trump-deepfakes-blackvoters-1875610 Last retrieved: 02 July 2024

Johnson, D. B. (2024, February 5). Deepfakes, dollars and 'deep state' fears: Inside the minds of election officials heading into 2024. *CyberScoop*. https://cyberscoop.com/deepfakesdollars-deep-state-fears-election-officials-concerns-2024/ Last retrieved: 02 July 2024

Johnson, T. (2023, October 30). Joe Biden talks about watching an AI generated deepfake of himself: "I said, 'When the hell did I say that?'" — update. *Deadline*. https://deadline.com/2023/10/ai-joe-biden-executive-order-

1235586979/ Last retrieved: 02 July 2024

Lynn, B. (2024, January 24). 'Deepfake' of Biden's voice called early example of US election disinformation. *Voice of America*. https://learningenglish.voanews.com/a/deepfake-ofbiden-s-voice-called-early-example-of-us-election-disinformation/7455392.html# Last retrieved: 02 July 2024

Mirza, R. (2024, February 17). *How AI-generated deepfakes threaten the 2024 election*. The Journalist's Resource. https://journalistsresource.org/home/how-ai-deepfakes-threatenthe-2024-elections/ Last retrieved: 02 July 2024

Mulvihill, G. (2024, January 31). *What to know about how lawmakers are addressing deepfakes like the ones that victimized Taylor Swift | AP News*. AP News. https://apnews.com/article/deepfake-images-taylor-swift-state-legislationbffbc274dd178ab054426ee7d691df7e Last retrieved: 02 July 2024

Nelson, J. (2024a, February 3). Deepfake news: Donald Trump says AI is 'So scary'. *Decrypt*. https://decrypt.co/215556/donald-trump-ai-scary-dangerous Last retrieved: 02 July 2024

Nelson, J. (2024b, March 8). Biden targets AI deepfakes in state of the Union, pushes for stronger privacy laws. *Decrypt*. https://decrypt.co/220907/biden-state-union-aideepfakes-privacy Last retrieved: 02 July 2024

Nt, S. (2024, February 20). Here's a Look at US Govt's New Legislations to Tackle Deepfakes, AI-generated explicit Content. *MediaNama*. https://beta.medianama.com/2024/02/223us-govt-deepfake-legislation/ Last retrieved: 02 July 2024

Overly, S. (2023, October 8). Washington grapples with AI deepfakes on the campaign trail. *Politico*. https://www.politico.com/news/2023/08/10/congress-ai-deepfakescampaign-trail-00110578 Last retrieved: 02 July 2024

Torkington, S. (2024, February 27). The US is drafting new laws to protect against AI-generated deepfakes- Google Search. *World Economic Forum*. https://www.google.com/search?client=safari&rls=en&q=The+US+is+drafting+new+laws+to+protect+against+AIgenerated+deepfakes+simon+torkington&ie=UTF-8&oe=UTF-8 Last retrieved: 02 July 2024

Congressman Joe Morelle Announces New Support For His Legislation To Stop Deepfake Pornography (2024, 24. April). Representative

Morelle. https://morelle.house.gov/media/press-releases/congressman-joe-morelleannounces-new-support-his-legislation-stop-deepfake Last retrieved: 02 July 2024

Weiner, D. I. (2023, December 5). Regulating AI Deepfakes and Synthetic Media in the Political Arena. *Brennan Center for Justice*. https://www.brennancenter.org/ourwork/research-reports/regulating-ai-deepfakes-and-synthetic-media-political-arena Last retrieved: 02 July 2024

TRANSCRIPTS

Advanced Technology: Examining Threats to National Security: Hearing before the Subcommittee on Emerging Threats and Spending Oversight, 118th Cong. [Hearing Transcript]. (2023). https://www.hsgac.senate.gov/wp-content/uploads/CHRG118shrg53708.pdf Last retrieved: 02 July 2024

Advances In Deepfake Technology: Hearing Before The Subcommittee On Cybersecurity, Information Technology, And Government Innovation. 118. Cong. [Hearing Transcript]. (2023). https://www.congress.gov/118/meeting/house/116559/documents/HHRG-118GO12-Transcript-20231108.pdf Last retrieved: 02 July 2024

AI And The Future Of Our Elections: Hearing Before The Committee On Rules And Administration United States Senate. 118 Cong. [Hearing Transcript]. (2023). https://www.rules.senate.gov/imo/media/doc/d041300e-a43b-9c54-b43cc86a983abc9a/AI%20and%20the%20Future%20of%20Our%20Elections.pdf Last retrieved: 02 July 2024

Americans At Risk: Manipulation And Deception In The Digital Age: Hearing Before The Subcommittee On Consumer Protection And Commerce. 116. Cong. [Hearing Transcript]. (2020). https://www.congress.gov/116/chrg/CHRG-116hhrg44716/CHRG-116hhrg44716.pdf Last retrieved: 02 July 2024

Artificial Intelligence and Human Rights: Hearing Before The Subcommittee On Human Rights And The Law. 118 Cong. [Hearing Transcript]. (2023). https://www.congress.gov/118/chrg/CHRG-118shrg52709/CHRG-

118shrg52709.pdf Last retrieved: 02 July 2024

Biden, J. (2024). Remarks of President Joe Biden- State of the Union Address As Prepared for Delivery: [Speech Transcript]. In *The White House*. https://www.whitehouse.gov/briefing-room/speeches-remarks/2024/03/07/remarks-of-president-joe-biden-state-of-the-union-address-asprepared-for-delivery-2/ Last retrieved: 02 July 2024

Biden, J. & Harris, K. (2023). Remarks by President Biden and Vice President Harris on the Administration's Commitment to Advancing the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence: [Speech Transcript]. In *The White House*. https://www.whitehouse.gov/briefing-room/speeches-remarks/2023/10/30/remarks-by-president-biden-and-vice-president-harris-on-theadministrations-commitment-to-advancing-the-safe-secure-and-trustworthydevelopment-and-use-of-artificial-intelligence/ Last retrieved: 02 July 2024

Connolly, G. E. (2024). *Ranking Member Connolly's Opening Statement at 2nd Subcommittee Hearing Examining Deepfakes | The Committee on Oversight and Accountability Democrats*. oversightdemocrats.house.gov. https://oversightdemocrats.house.gov/news/pressreleases/ranking-member-connollys-opening-statement-2nd-subcommittee-hearingexamining Last retrieved: 02 July 2024

Fostering A Healthier Internet To Protect Consumers: Joint Hearing Before The Subcommittee On Communications And Technology And The Subcommittee On Consumer Protection And Commerce. 116th Cong. [Hearing Transcript]. (2020). https://www.govinfo.gov/content/pkg/CHRG-116hhrg43533/pdf/CHRG-116hhrg43533.pdf Last retrieved: 02 July 2024

Governing AI Through Acquisition And Procurement: Hearing Before The Committee On Homeland Security And Governmental Affairs United States Senate, 118. Cong. [Hearing Transcript]. (2023). https://www.congress.gov/118/chrg/CHRG118shrg53707/CHRG-118shrg53707.pdf Last retrieved: 02 July 2024

Iyer, P. (2024, 29. April). *Transcript: US Senate Hearing on Oversight of AI & Election Deepfakes*. Tech Policy Press. https://www.techpolicy.press/transcript-us-senatehearing-on-oversight-of-ai-election-deepfakes/ Last retrieved: 02 July 2024

Miller, G. (2023). *Transcript: US House Hearing on "Advances in Deepfake Technology"* Tech Policy Press. https://www.techpolicy.press/transcript-us-house-hearing-on-advancesin-deepfake-technology/ Last retrieved: 02 July 2024

Miller, G. (2024). *Transcript: US House Subcommittee Hearing on the Harms of Deepfakes*. Tech Policy Press. https://www.techpolicy.press/house-hearing-addressing-deepfakeharms/ Last retrieved: 02 July 2024

Securing America's Elections: Hearing Before The Committee On The Judiciary House of Representatives. 116. Cong. [Hearing Transcript]. (2021). https://www.congress.gov/116/chrg/CHRG-116hhrg45285/CHRG-116hhrg45285.pdf Last retrieved: 02 July 2024

Securing The Future: Harnessing The Potential Of Emerging Technologies While Mitigating Security Risks: Hearing Before The Subcommittee On Cybersecurity, Infrastructure Protection, And Innovation. 117. Cong. [Hearing Transcript]. (2022). https://www.congress.gov/117/chrg/CHRG-117hhrg48856/CHRG-117hhrg48856.pdf Last retrieved: 02 July 2024

Voting Rights And Election Administration: Combatting Misinformation In The 2020 Election: Hearing Before The Subcommittee On Elections. 116. Cong. [Hearing Transcript]. (2020). https://www.govinfo.gov/content/pkg/CHRG-116hhrg42741/pdf/CHRG-116hhrg42741.pdf Last retrieved: 02 July 2024

White House Overreach On AI: Hearing Before The Subcommittee On Cybersecurity, Information Technology, And Government Innovation. 118.Cong. [Hearing Transcript]. (2024). https://www.congress.gov/118/chrg/CHRG-118hhrg55220/CHRG-118hhrg55220.pdf Last retrieved: 02 July 2024

White House Policy On AI: Hearing Before The Subcommittee On Cybersecurity, Information Technology, And Government Innovation. 118. Cong. [Hearing Transcript]. (2023). https://www.congress.gov/118/chrg/CHRG-118hhrg54392/CHRG-118hhrg54392.pdf Last retrieved: 02 July 2024

LGEAL TEXTS

Committee on Homeland Security and Governmental Affairs. (2022). *Deepfake Task Force Act: Report of the Committee on Homeland Security and Governmental Affairs, United States Senate, to accompany S. 2559 to establish the National Deepfake and Digital Provenance Task Force, and for other purposes: 117th Cong.*https://www.govinfo.gov/content/pkg/CRPT-117srpt114/pdf/CRPT-117srpt114.pdf Last retrieved: 02 July 2024

U.S. Congress. (2020). *Identifying Outputs of Generative Adversarial Networks Act: Public Law No. 116-258, 134 Stat. 1150*[Dataset]. https://www.govinfo.gov/content/pkg/PLAW-116publ258/pdf/PLAW116publ258.pdf Last retrieved: 02 July 2024

U.S. House of Representatives. (2021). *H.R.2395 To combat the spread of disinformation through restrictions on deep-fake video alteration technology.: 117. Cong* [Dataset]. https://www.congress.gov/116/bills/s4049/BILLS-116s4049es.pdf Last retrieved: 02 July 2024

U.S. House of Representatives. (2023a). *H.R. 5586: A bill to protect national security against the threats posed by deepfake technology and to provide legal recourse to victims of harmful deepfakes: 118th Cong.* [Dataset]. https://www.congress.gov/118/bills/hr5586/BILLS-118hr5586ih.pdf Last retrieved: 02 July 2024

U.S. House of Representatives. (2023b). *H.R. 5586: A bill to protect national security against the threats posed by deepfake technology and to provide legal recourse to victims of harmful deepfakes: 118th Cong.* [Dataset]. https://www.congress.gov/118/bills/hr5586/BILLS-118hr5586ih.pdf Last retrieved: 02 July 2024

U.S. House of Representatives. (2023c). *H.R. 5808: A bill to establish the Task Force on Artificial Intelligence in the Financial Services Sector to report to Congress on issues related to artificial intelligence in the financial services sector, and for other purposes: 118th Cong.* [Dataset]. https://www.congress.gov/118/bills/hr5808/BILLS118hr5808ih.pdf Last retrieved: 02 July 2024

U.S. House of Representatives. (2024). *H.R.7766: A bill to require the National Institute of Standards and Technology to establish task forces to facilitate and inform the development of technical standards and guidelines relating to the identification of content created by generative artificial intelligence, to ensure that audio or visual content created or substantially modified by generative artificial intelligence includes a disclosure acknowledging the generative artificial intelligence origin of such content, and for other purposes: 118th Cong.* [Dataset]. https://www.congress.gov/118/bills/hr7766/BILLS-118hr7766ih.pdf Last retrieved: 02 July 2024

U.S. Senate. (2020). *S.4049 An act to authorize appropriations for fiscal year 2021 for military activities of the Department of Defense, for military construction, and for defense activities of the Department of Energy, to prescribe military personnel strengths for such fiscal year, and for other purposes: 116th Cong.* [Dataset]. Last retrieved: 02 July 2024

U.S. Senate. (2021). *S. 2559: A bill to establish the National Deepfake and Digital Provenance Task Force, and for other purposes: 117th Cong.* [Dataset]. https://www.congress.gov/117/bills/s2559/BILLS-117s2559rs.pdf Last retrieved: 02 July 2024

U.S. Senate. (2023). *S. 2770 A bill to prohibit the distribution of materially deceptive AIgenerated audio or visual media relating to candidates for Federal office, and for other purposes: 118th Cong.* [Dataset]. https://www.congress.gov/118/bills/s2770/BILLS-118s2770is.pdf Last retrieved: 02 July 2024

7.2 CITATIONS Citations From: 4.2 Deepfakes as Benefits for Industries

1. “Now, I know that there are instances where there’s a parody and there’s humor, and I’ve seen AI with prominent politicians doing funny things, and it is funny, but it’s also quite obvious.” (Iyer, 2024)

2. “As I said before, our goal is not to make everything impossible, but our goal is to make malicious activity more difficult and more complicated, while allowing deepfakes for Hollywood movies or other types of entertainment applications to proceed.” (*Advanced Technology: Examining Threats to National Security: Hearing before the Subcommittee on Emerging Threats and Spending Oversight,* 2023, p.23)

3. “They are used to enhance video games and other forms of entertainment, and they are being used to advanced medical research as well, but deepfake technology can also be weaponized and cause great harm.” (Miller, 2023, p. 1)

4. “For example, AI technology and deep learning algorithms can help us detect cancers earlier and more quickly. Clinical trials are already underway and making major breakthroughs to diagnose cancers.” (*Americans At Risk: Manipulation And Deception In The Digital Age: Hearing Before The Subcommittee On Consumer Protection And Commerce,* 2020, p. 4)

Citations From: 4.3 Deepfakes as Threat to Democracy, Society and Individual

1. “The threat that deepfakes could pose if used in misinformation campaigns is well known and well-documented. The Congressional Research Service (CRS) has noted that ‘state adversaries or politically motivated individuals could release falsified videos of elected officials or other public figures making incendiary comments or behaving inappropriately‘” (Committee on Homeland Security and Governmental Affairs, 2022, p.2)

2. “As the world approaches another election cycle, Trump raises a pertinent concern about the malicious use of AI to influence voters. He warns that in an election year, AI could be used to maliciously generate images, voices and other forms of content to manipulate public opinion.” (Gil, 2024)

3. “In the interview, Trump called for action regarding AI- and AI-generated deepfakes, raising the concern that the technology could be used to start wars.” (Nelson, 2024a)

4. “Foreign governments and fraudsters have pursued political disinformation campaigns using new technology like deepfakes designed to sow civil unrest and disrupt democratic elections.” (*Fostering A Healthier Internet To Protect Consumers: Joint Hearing Before The*

Subcommittee On Communications And Technology And The Subcommittee On Consumer Protection And Commerce, 2020, p.12)

5. "The non-consensual sharing of synthetic intimate images, often called deepfake pornography, is an exponentially growing form of abuse that humiliates, degrades and threatens women and girls, causing real and lasting damage to their mental health and wellbeing." (*Congressman Joe Morelle Announces New Support For His Legislation To Stop Deepfake Pornography*, 2024)

Citations From: 4.4 Deepfake Challenges: Legal, Detection and Accessibility Issues

1. "The interviewees, including Trump and Gensler, emphasize the inadequacy of existing legal frameworks in addressing the risks posed by deepfakes, signaling the need for innovative and adaptive regulatory approaches." (Adebayo, 2024)

2. "The hearing concluded with a. call for action to address the challenges posed by deepfake technology, while also being mindful of protecting civil liberties and privacy" (Miller, 2023)

3. "What you say at an interview with you almost doesn't matter anymore. They can change things around, and nobody can tell the difference; even experts can't tell the difference. This is a tremendous problem in terms of security." (Nelson, 2024a)

4. "Of all the surfaces- video, image, audio- that AI can be used for voter suppression, audio is the biggest vulnerability. It is easy to clone a voice using AI, and it is difficult to identify." (Chopra, 2024).

5. "Recent technological advancements in artificial intelligence have opened the door for bad actors with very little technical knowledge to create deepfakes cheaply and easily." (Miller, 2024)

Citations From: 4.5 Solutions: A Multi-Faceted Approach to Combat Deepfakes

1. "Collaboration between Congress and technology companies is essential, and I am glad to see that that is happening to address the challenges posed by deepfakes. Tech companies should be responsible for developing and implementing the policies to detect and mitigate this type of content, including what we were hearing today, on their platforms and sharing, most importantly, what they learn with others. We have addressed this type of a problem with our cybersecurity, and we should be doing that same thing with our misinformation. More robust privacy and consent laws are needed to protect individuals from using their likeness and voice

in deepfake content without their permission, and continued research and development in AI deepfake technology are necessary, as is funding to counter deepfake misuse." (*Advances In Deepfake Technology: Hearing Before The Subcommittee On Cybersecurity, Information Technology, And Government Innovation*, 2023, p.8)

2. "There was a consensus on the urgent need for both public awareness and legislative action to combat the misuse of deepfake technology, with suggestions for incorporating digital literacy into education and developing technologies to detect and mark AI-generated content." (Miller, 2023)

3. "In February, the Biden administration said it would use watermarking and cryptography to fight political disinformation" (Nelson, 2024b)

4. "Specifically, the National Institute of Standards and Technology (NIST) U.S. AI Safety Institute will create guidelines, tools, benchmarks, and best practices to evaluate and mitigate risks using red-teaming and other methods. The Institute will develop technical guidance on issues such as authenticating content created by humans, watermarking AI-generated content, identifying and mitigating against harmful algorithmic discrimination and the creation of harmful synthetic content (such as deepfake image-based abuse), ensuring transparency, and enabling adoption of privacy-preserving AI, and would support the development of safe and trusted AI and support enforcement agencies' efforts to protect the American public rights and safety." (The White House, 2024c)