Olaf Scholz Deepfake:
How a Deepfake impacts Public Trust

Bachelor Thesis

Public Governance across Borders B.Sc.

1st Supervisor: Dr. Azadeh Akbari

2nd Supervisor: Dr. Ringo M.R.R Ossewaarde

University of Twente

Enschede

Christopher von Eynern

[redacted]

11.498 words

Ethical Approval: 240786

Presentation Date: 3rd of July 2024

Table of Contents

Abstract

In the digital age, deepfake technology significantly threatens the integrity of information and democratic discourse. This thesis examines the socio-political effects of deepfakes, focusing on cognitive and emotional responses and the broader implications for public trust in media and political institutions. Through the "AfD-Ban" deepfake video, this thesis provides a case-specific understanding of deepfakes' influence on public perceptions and trust. Utilizing an interpretive approach, experimental semi-structured interviews capture initial perceptions of the video's authenticity, responses upon learning the video's deepfake nature, and reflections on broader implications for public trust. Findings reveal that initial perceptions were influenced by digital literacy, pre-existing biases, and public trust itself. Higher digital literacy and critical thinking facilitated recognition of the deepfake, while lower digital literacy or stronger political biases increased the vulnerability to deception. Emotional responses ranged from indifference to fear. Deepfakes significantly impact public trust and democratic discourse, with concerns about the erosion of trust in media and political institutions, the amplification of misinformation, and the manipulation of public opinion. Multiple solutions are imaginable: enhancing media literacy, implementing technological safeguards, developing regulatory measures for transparency and accountability, promoting ethical standards in digital media, and fostering inclusive public discourse to counteract echo chambers and polarization.

1. Introduction

1.1 Background and State of the Art

In the digital age, technological advancements have dramatically transformed the way information is created, disseminated, and consumed (Castells, 2009). Among these transformations, deepfake technology stands out for its ability to create hyper-realistic digital manipulations of videos, audio recordings, or images using artificial intelligence. This technology can make it appear as if individuals are saying or doing things they never actually did, which poses a significant threat to the integrity of information. Deepfake videos are digital creations where a person's face or voice is artificially swapped with someone else's by using advanced artificial intelligence and machine learning technologies (Vasist & Krishnan, 2022). These videos can make it appear as if individuals are saying or doing things they never actually did, which triggers the dissolution of the line between reality and fiction. The spread of such technology, particularly within social networks, underscores the potential for misinformation and the influence of artificial intelligence in distorting public discourse (Gnatik, 2024; Hou, McGonigal, & Yu, 2024). Some social scientists name the current state a "post-truth era", further emphasizing the impact of misinformation on daily life (Harari, 2018). Therefore, deepfake technology has gained significant scientific and political attention for its potential to spread misinformation. Democracy thrives on a free, inclusive, and open discourse where the truthfulness of information is crucial (Habermas, 1989). The rapid increase in misinformation, exacerbated by deepfakes, is perceived as a significant threat to this democratic discourse (Gnatik, 2024; Hou, McGonigal, & Yu, 2024), which has led to a heightened sense of urgency to understand and mitigate the risks associated with deepfakes. Despite the substantial body of existing literature on the technical and ethical aspects of deepfake technology, there's a noticeable gap in understanding its socio-political effects from an individual´s perspective, and especially from a qualitatively experiential standpoint (Vasist & Krishnan, 2022). Recent studies have only started to investigate how deepfakes could mislead the public, diminish media credibility, and manipulate political narratives. However, the individual experiences of those who encounter deepfakes, including their cognitive and emotional reactions and the subsequent societal effects, remain largely undiscovered in the social sciences. While psychology scholars like Lundberg & Mozelius (2023) are indeed trying to find answers to the individual's behavior and feelings towards deepfakes, the second step of researching the societal consequences,

in the realms of politics which focuses on deepfakes´ implications for individuals and their resulting societal impacts, remains unenlightened. This thesis aims to address this gap by exploring the cognitive and emotional responses of individuals to deepfake content and the broader implications for public trust in media and political institutions. Through the examination of the case study of the "AfD-Ban" deepfake video featuring German Chancellor Olaf Scholz, this thesis seeks to provide a nuanced individual-focused understanding of how deepfakes influence public perceptions and trust. To understand people's perceptions and emotions is essential, as deepfakes' development continues to grow, posing challenges to public trust and democratic dialogue. Because of this Pei et al. (2024) emphasize the importance of addressing the human aspects of deepfakes to develop effective responses. Addressing this need, this thesis seeks to explore the individual and societal effects of deepfakes, using the “AfD-Ban” deepfake as a case study. Through experimental semi-structured interviews, this research aims to uncover how individuals perceive authenticity, manage their emotional and cognitive reactions, and perceive deepfakes' influence on their trust in media and democratic institutions. By examining the case study of the "AfD-Ban" deepfake video featuring German Chancellor Olaf Scholz, this study seeks to provide a nuanced understanding of how deepfakes influence public perceptions and trust. This research will contribute to the literature by highlighting the need for enhanced digital- and media literacy and regulatory frameworks to mitigate the socio-political risks posed by deepfakes, thereby adding knowledge to the individual focused lens of the current understanding of deepfake technology.

1.2 Research Questions

To fill this knowledge gap, the aim of this research is to explore the impact of deepfake technology on public trust. Therefore, the main research question is: "How do perceptions and reactions of individuals to the “AfD-Ban” deepfake video reflect concerns about the influence of deepfake technology on public trust?" This research question seeks to uncover the nuances of how deepfakes influence individual perceptions and trust in media and political institutions. Through the investigation of cognitive and emotional responses to the deepfake video, this thesis further aims to reveal the underlying factors that shape these perceptions. To further elucidate the main research question, the following sub-questions are formulated:

1. "How do individuals initially perceive the authenticity of the “AfD-ban” deepfake video, and what factors influence their judgment?" This sub-question aims to capture the cognitive abilities of the participants to recognize a deepfake. It explores the initial reactions and the factors, such as digital literacy and pre-existing biases, that influence these judgments.
2. "What are the cognitive and emotional responses after learning of the video’s deepfake nature?" This sub-question examines the reactions after the participants learn about the falsity of the video. It delves into the emotional and cognitive dissonance experienced upon discovering that the content was manipulated.
3. "How do individual reactions to the “AfD-Ban” video influence broader perceptions of public trust?" This sub-question investigates the potential consequences of the deepfake at a macro level, focusing on the impacts on public trust in media and democratic processes.

1.3 Research Approach

The approach used in this thesis is interpretive, which is suitable for exploring complex social phenomena through the subjective meanings and perspectives of the involved actors (Bhattacherjee, 2012). According to Bhattacherjee (2012), interpretive research is valuable for understanding social realities by considering the contexts in which they occur. Therefore, this approach allows for the discovery of new insights and theories through its flexible and context-sensitive analysis of data. Interpretive research is particularly well suited for this study because it facilitates a deep understanding of individual experiences and perceptions. This approach recognizes that social phenomena are constructed through the meanings that individuals attach to them (Given, 2008). By using semi-structured interviews, this study aims to capture the rich, qualitative data needed to understand the cognitive and emotional responses to deepfake content. The primary method employed in this thesis is qualitative content analysis, conducted through the data collection via experimental semi-structured interviews. This method is chosen for its ability to generate in-depth data that reflect the theoretical concepts of authenticity, trust, and the impact of misinformation. The experimental semi-structured interviews provide the flexibility to explore participants' perceptions and reactions comprehensively while allowing for the emergence of new themes during the conversation and upon learning of the video’s true nature (Given, 2008). Consequently, the data collection process requires conducting the experiments

with a purposively sampled and diverse group of individuals. To provide guidance for the reader, this thesis is organized into successive and interrelated chapters to present the findings in a structured manner:

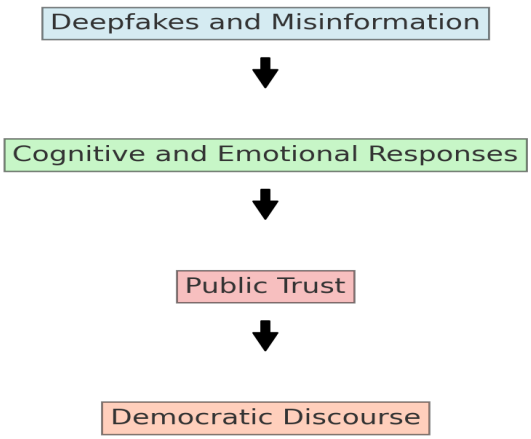
1. Introduction: Establishes the background, research questions, and approach.
2. Theory: Provides the theoretical framework, including key concepts from deepfake research and the understanding of "trust."
3. Methods: Describes the data collection and analysis methods, including the rationale for using semi-structured interviews and content analysis.
4. Analysis: Presents the findings from the interview experiments, addressing the research questions and connecting them to the theoretical framework.
5. Conclusion: Summarizes the research findings, discusses their implications, and suggests directions for future research and policy making.

In conclusion, this bachelor thesis seeks to contribute to the understanding of the socio-political effects of deepfake technology. Via the examination of the "AfD-Ban" deepfake video case study, it aims to provide valuable insights into how deepfakes influence individual perceptions and the broader implications on public trust. Furthermore, it seeks to provide insights into possible strategies and policy solutions to mitigate the risks posed by this emerging misinformation technology.

2. Theory

2.1 Introduction

The purpose of this chapter is to provide a comprehensive theoretical foundation for examining the impact of deepfake technology on public trust. This chapter aims to delineate and integrate key concepts and theoretical frameworks that will guide the analysis of the research data, ensuring the study is grounded in established theories while also identifying gaps that this research seeks to address. The chapter is structured to introduce the theoretical framework first, followed by detailed discussions of each key concept, and concludes with a synthesis of these concepts to set the stage for the research analysis. To aid in understanding the theoretical connections and provide a clear roadmap for this chapter, the following visualization outlines the linear theoretical framework used in this thesis:



2.2 Key Concept: Deepfakes and Misinformation

Deepfakes are hyper-realistic digital manipulations created using advanced artificial intelligence (AI) and machine learning (ML) technologies. These technologies can produce convincing fake videos and audio recordings, where individuals appear to say or do things they never actually did (Vasist & Krishnan, 2022). The term "deepfake" is derived from "deep learning," a subset of ML used to create these realistic forgeries. This key concept explores the technical and ethical dimensions of deepfakes as a form of misinformation. So far, research on deepfakes has primarily focused on their creation, detection, and ethical implications. Chesney and Citron (2019) discuss the rise of deepfake technology and its potential to deceive viewers, highlighting the challenges in distinguishing real from fake content. They argue that deepfakes could significantly impact public trust, particularly if used in political campaigns or to spread

disinformation. Diakopoulos and Johnson (2021) examine the mechanisms by which deepfakes are created, emphasizing the role of generative adversarial networks (GANs) in producing realistic forgeries. The ethical concerns surrounding deepfakes are significant. Vaccari and Chadwick (2020) argue that deepfakes can erode trust in visual and auditory evidence, fundamentally challenging the credibility of media. Furthermore, they point out the potential for deepfakes to be used maliciously, such as in political campaigns to spread disinformation and manipulate public opinion. These ethical concerns are compounded by the difficulty in detecting deepfakes, as the technology to create them continues to improve. While the technical aspects and ethical concerns of deepfakes are well-documented, there is a noticeable gap in understanding the individual and societal impacts of deepfakes from a qualitative, experiential perspective. This research aims to fill this gap by exploring how deepfakes influence individual perceptions and public trust. By focusing on the cognitive and emotional responses of individuals to deepfake content, this study seeks to provide a deeper understanding of the broader societal implications of this technology.

2.3 Key Concept: Cognitive and Emotional Responses

On the individual level, cognitive and emotional responses refer to the mental and emotional reactions individuals have when they encounter deepfake content. These responses can include cognitive dissonance, emotional distress, and changes in perception and attitude (Hameleers et al., 2020). Cognitive responses involve the mental processes of perception, memory, judgment, and reasoning, while emotional responses encompass feelings such as anger, fear, surprise, and betrayal. Vosoughi et al. (2018) discuss the cognitive biases, such as the illusory truth effect and confirmation bias, that can lead individuals to believe false information, including deepfakes. These biases are particularly potent when the content aligns with pre-existing beliefs, making deepfakes a powerful tool for misinformation. The illusory truth effect occurs when repeated exposure to false information makes it seem more credible, while confirmation bias leads individuals to favor information that confirms their existing beliefs. Emotional responses to deepfakes can be profound. Hameleers et al. (2020) found that emotionally charged misinformation is more likely to be believed and shared. This aligns with the findings of Brooks (2021), who noted that deepfakes can induce significant emotional reactions, such as anger, betrayal, and confusion, which can affect individuals' trust in media and political figures. For example, a deepfake video depicting a political figure making controversial statements

can provoke strong emotional reactions, leading to changes in public perception and trust. Existing research has identified the cognitive biases and emotional reactions to misinformation, but there is limited understanding of how these responses manifest in the context of deepfakes specifically. This thesis seeks to explore these responses in detail, particularly focusing on the cognitive and emotional impacts on individuals who encounter deepfake content. By examining the immediate and possible long-term cognitive and emotional responses to deepfakes, this study aims to provide insights into how such content can influence public trust and perception. To better separate the individual (micro) from the general (macro), this thesis partly makes use of an existing theoretical framework introduced by Vasist & Krishnan (2022). Vasist & Krishnan (2022) introduced this framework within their literature review on current deepfake research where they suggest this holistic framework for further research on deepfakes. This thesis will make use of this framework for the conception of the data collection and analysis. Specifically focusing on the introduced aspects of deepfake viewing, and deepfake consequences on micro- and macro-level (see *Figure 1*).

I. The viewing phase is particularly critical as it engages viewers cognitively and emotionally, factors which are relevant to the belief in the deepfake. Cognitive biases such as the illusory truth effect or confirmation biases can lead viewers to accept deepfakes as authentic, especially when the content aligns with their pre-existing beliefs (Vosoughi et al., 2018; Diakopoulos & Johnson, 2021). Additionally, participant´s emotional responses also play a significant role, because according to Hameleers et al. (2020) content that includes a strong emotional reaction is likely to be more convincing.

II. On a micro-level, the consequences on the individual are investigated. In this case, the individual consequences for the viewers (interviewees) are the micro-level. On the individual level, encountering deepfakes can lead to changes in the viewer´s attitudes and beliefs, and a deepened uncertainty about the truthfulness of media (Westerlund, 2019; Brooks, 2021).

III. At the macro-level, the impacts on institutional, market, and/or societal level is highlighted (Vasist & Krishnan, 2022). According to Ahmed (2021), deepfakes have the potential to erode institutional trust, amplify social polarization, and distort political processes.

They can challenge the credibility of foundational institutions, thereby weakening democratic governance. Deepfakes can also exacerbate social divides by reinforcing in-group biases and deepening societal fractures (Ahmed, 2021). Moreover, the legal consequences are significant because deepfakes test existing legal frameworks around intellectual property, consent, and defamation (Vaccari & Chadwick, 2018).

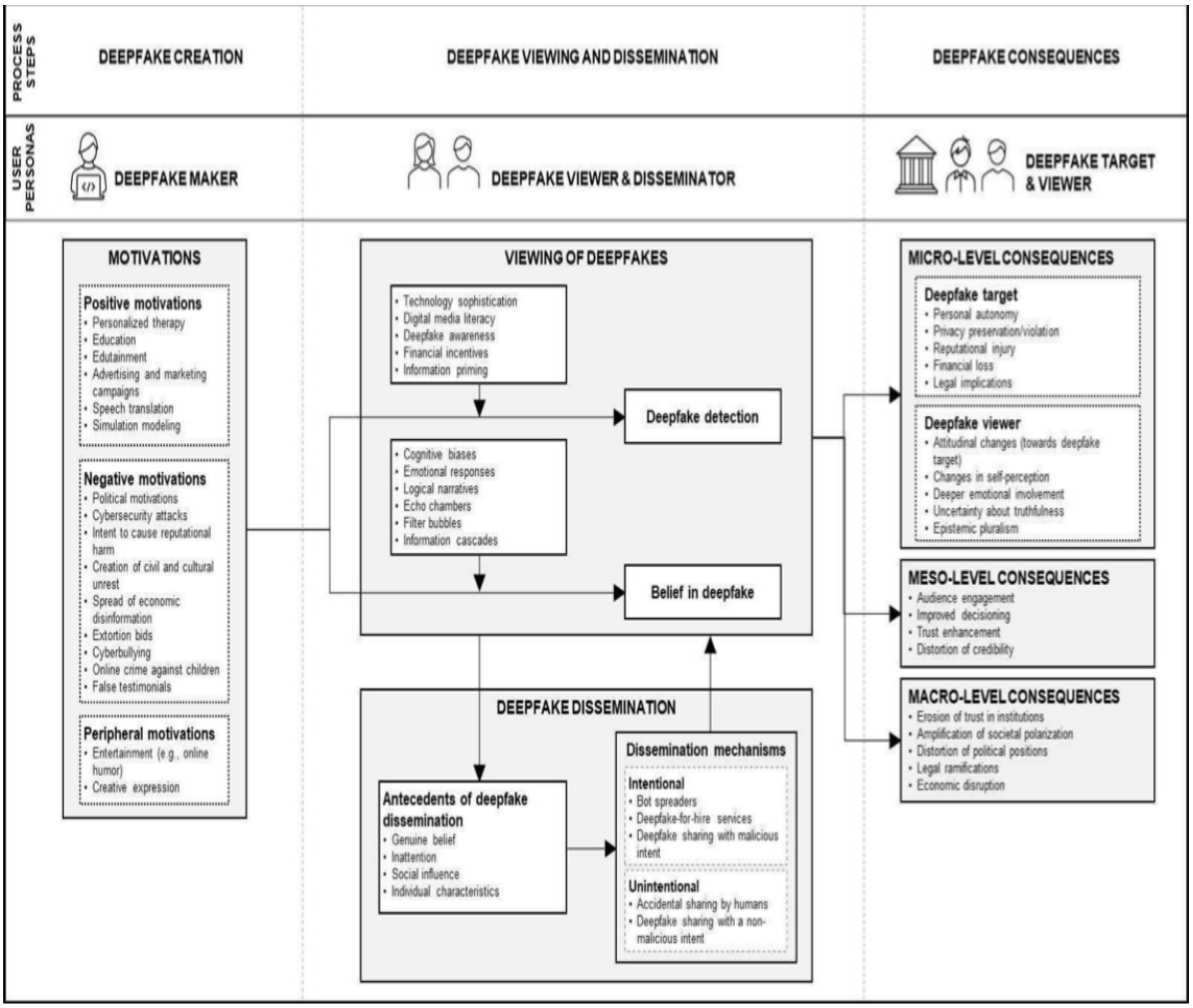


Figure 1: Vasist & Krishnan (2022)

2.4 Key Concept: Public Trust

This section provides the necessary definition of the main variable of investigation: Public Trust. For this thesis, public trust is defined as the confidence that individuals have in the integrity and reliability of media, political figures, and institutions (Rousseau et al., 1998). Trust is a crucial component of social capital, facilitating cooperation and social cohesion (Putnam, 2000). Trust can be broken down into several key dimensions:

Integrity: The belief that the other party adheres to a set of principles that the trustor finds acceptable.

Benevolence: The belief that the trustee has the trustor's best interests at heart.

Ability: The belief that the trustee possesses the necessary skills and competence to perform as expected.

In this regard, Mishler and Rose (2001) discuss the origins of political trust, highlighting the role of institutional performance and cultural factors. They argue that trust in institutions is essential for the functioning of democratic societies. Cappella and Jamieson (1997) emphasize the role of media in shaping public perceptions and trust, noting that the credibility of media sources is critical for maintaining public trust. The proliferation of deepfakes poses a significant threat to public trust. Chesney and Citron (2019) argue that deepfakes can undermine trust in media and political figures by creating a "post-truth" environment where objective facts are overshadowed by convincing falsehoods. Consequently, and further highlighting the need for combating misinformation, this erosion of trust can lead to political instability and reduced compliance with laws and regulations (Vaccari & Chadwick, 2020). When individuals lose trust in their information sources, it can create a climate of uncertainty and skepticism, further complicating the dissemination of accurate information. Based on this conceptualization of public trust, the methodology as well as the analysis can be constructed for a precise setting of interview questions and a theory-based analysis of the data. However, in order to put “public trust” into perspective, and completely answer the research question with the focus on the macro implications, a holistic setting is needed, in which the variables “public trust” and “the individual” can relate to broader implications. For this, the utilization of the public sphere and its emphases on democratic discourse becomes adequate.

2.5 Key Concept: Democratic Discourse

Democratic discourse refers to the free and open exchange of ideas and information that is essential for a functioning democracy. It relies on the availability of accurate information and the ability of individuals to critically engage with this information (Habermas, 1989). The concept of the public sphere, as introduced by Habermas, is central to understanding democratic discourse. The public sphere is a space where individuals can come together to discuss, and debate matters of public interest. Habermas (1989) traces the origins of the public sphere to the 18th-century European bourgeois society, where rational-critical debate among private individuals helped to shape the public opinion. Calhoun (1992) and Fraser (1990) extend this concept by discussing the evolution and inclusivity of the public sphere, which highlighted the role of media even more in facilitating public discourse. The rise of digital communication technologies has transformed the public sphere, creating new challenges for democratic discourse (Habermas, 2012). Papacharissi (2010) discusses the potential of digital public spheres to enhance democratic engagement by providing platforms for marginalized voices. However, Sunstein (2001) highlights the risks of echo chambers and polarization, where individuals are exposed primarily to information that reinforces their existing beliefs. These echo chambers then limit exposure to diverse perspectives, which undermines the deliberative quality of public discourse. The impact of deepfakes on democratic discourse is profound. Fuchs (2014) and Marwick & Boyd (2011) discuss how misinformation, in general, and online harassment can hinder the inclusivity and quality of public discourse. Deepfakes, by creating realistic but false content, pose a significant threat to the integrity of democratic dialogue. For example, a deepfake video that falsely depicts a political candidate engaging in unethical behavior could influence public opinion and disrupt the democratic process. While there is plenty of research exploring how deepfakes influence democratic discourse, particularly in terms of their potential to amplify misinformation and specifically polarize public opinion. This research aims to examine the broader societal implications of deepfake-related trust implications on democratic engagement and discourse. By investigating the effects of deepfakes on the public sphere, this study ultimately goes one step further in investigating not only the micro-level consequences of deepfakes, but rather the implication of the micro-level consequences for the macro-level consequences within the public sphere. However, because of the limitations of this bachelor thesis, only the important aspect of “public

trust" as a crucial part of the public sphere, for equal and believable transmission of information within democratic societies, can be investigated.

2.6 Conclusion

This chapter has provided a comprehensive overview of the key theoretical concepts that underpin this research. The four key concepts—deepfakes and misinformation, cognitive and emotional responses, public trust, and democratic discourse—are essential for understanding the impact of deepfake technology on public trust. Therefore, based on all theoretical insights discussed, this research expects to find that:

1. Deepfakes significantly influence individual perceptions and can lead to cognitive biases and emotional reactions, such as confusion, anger, and betrayal.
2. Exposure to deepfakes erodes public trust in media and political figures, contributing to a "post-truth" environment.
3. The impact of deepfakes on public trust has broader implications for democratic discourse, potentially amplifying misinformation and polarizing public opinion. By investigating these theoretical expectations, this research aims to contribute to a deeper understanding of the socio-political effects of deepfake technology and to formulate possible strategies to mitigate its risks.

Finally, this theoretical framework will serve as the foundation for the empirical analysis in the subsequent chapters, guiding the investigation of how deepfake technology affects individual perceptions (Micro-Level), public trust, and democratic discourse (Macro-Level).

3. Methods

3.1 Introduction

This chapter outlines the methodology used to explore the impacts of the "AfD-Ban" deepfake on public trust at both micro and macro levels. The purpose of this chapter is to provide a detailed description and justification of the research design, case study selection, data collection methods, and data analysis techniques used for this thesis. By elaborating on these methodological choices, this chapter seeks to establish the necessary reliability and validity of the research findings. This chapter is structured as follows: It begins with a comprehensive description of the case study, explaining its selection and relevance. This is followed by a description of the data collection method, including the sampling strategy and interview procedure. After that, the chapter then delves into the data analysis methods, in order to provide a rationale for the chosen approach and to describe the coding and analysis processes. Finally, the chapter concludes with a summary of the research activities and their alignment with the research objectives.

3.2 Description of the Case

The case selected for this research is the "AfD-Ban" deepfake video, created and published on Instagram by the German activist group "Zentrum für Politische Schönheit" (ZPS) on November 27, 2023. The video features a digitally manipulated version of German Chancellor Olaf Scholz, who is holding a speech about initiating a legal process to ban the far-right and partially extremist political party Alternative für Deutschland (AfD). The video gained attention on social media and mainstream news outlets, sparking widespread debate about its authenticity and implications leading to a ban of the original video. The video used for this thesis is a legally accessible and edited version of the original. It is publicly available on the ZPS YouTube channel. The video's content slightly differs from the original one, featuring a longer reasoning of the banning procedure and putting more emphasis on the historical parallels between the rise of the NSDAP during the Weimar Republic Era and the rise of the AfD. To put the video into perspective and provide context, it is important to mention that the public discourse about banning the AfD is still ongoing. Especially after the European Elections on June 2024, which just happened while writing this thesis, the AfD, despite having to deal with several corruption and treason scandals during the campaign,

gained the second largest number of votes right behind the conservative party, CDU. Overall, this case was chosen for several reasons:

1. Relevance: The deepfake involves high-profile political figures and addresses a controversial political topic, making it highly relevant for studying the impacts of deepfakes on public trust.
2. Sophistication: The use of advanced deepfake technology in creating the video provides a contemporary and impactful example for analysis.
3. Public Reception: The video's widespread dissemination and public and media attention make it a rich context for examining individual and societal reactions.

While the public reception after the publication by ZPS was immense in media and politics, it was not possible to find scientific work on this case featuring the implications of deepfakes in general, nor were any in-depth analyses on how the video impacted individuals found. That said, it is possible to assume that this case is relatively unexplored. Finally, due to the lack of resources available in the context of a bachelor thesis, this research will not cover the investigation into the origins or motivations behind the dissemination of this deepfake. Therefore, further exploration into the usage of deepfakes by civic activist groups, like ZPS, even though scientifically relevant, falls outside the scope of this thesis, especially considering the prevalent focus on political deepfakes and their implications on individuals.

Figure 2: Screenshot of the deepfake video. Source: ZPS (2024)

3.3 Method of Data Collection

Data for this research were collected through experimental semi-structured interviews with a purposively sampled group of eight individuals. This qualitative approach was chosen to explore in-depth perceptions and experiences rather than achieve statistical generalization. The sample was designed to represent a diverse cross-section of the population, including variations in age, education, gender, political affiliation, and digital literacy.

Sampling Strategy

To ensure transparency for this thesis's reliability specifically, it needs to be mentioned that the selection process involved the researcher in contacting friends and family to find suitable participants fitting the estimated profiles. However, due to a lack of time resources and adequate networking, participants 2, 5, and 7 already knew the researcher which potentially influenced their responses thus making the method less reliable and potentially diminishing external validity (Smith & Noble, 2014). In general, however, the Participants were selected based on specific criteria to ensure diversity and relevance to the research questions. All Participants were selected based on Germany's current demographic development (Statistisches Bundesamt, 2022), the percentual estimated voting share within Germany's Party system (INSA, 2024), and Germany's educational share (Statistisches Bundesamt, 2022). The selection criteria included:

Age: Participants ranged from 18 to over 65 years old to capture different generational perspectives.

Education: A mix of participants with varying educational backgrounds to explore how education influences the perception of deepfakes.

Gender: A balanced representation of male and female participants.

Political Affiliation: Participants with affiliations to different political parties to understand how political bias may affect the interpretation of deepfakes.

Digital Literacy: Participants with varying levels of digital literacy to examine the role of technological awareness in deepfake detection. However, due to the high complexity of measuring the spectrum of digital literacy with all its nuances, a complete grasp of each participant's digital literacy skills would not be possible in a scope of such bachelor thesis. Therefore, the criteria "using social media" and "consuming online information" will deal as indices for the participants possible digital literacy skills.

Participant Profiles

Group	Participant	Gender	Political Affiliation	Education Level	Social Media	Online Media
plus 65 years	1	Male	CDU supporter	No high school	None	None
plus 65 years	2	Female	CDU supporter	No high school	None	Active
40-64 years	3	Female	Green supporter	Academic	Active	Active
40-64 years	4	Female	CDU supporter	High school	Active	Active
40-64 years	5	Male	Other	Academic	Active	Active
18-39 years	6	Male	AfD supporter	No high school	Active	Active
18-39 years	7	Male	SPD supporter	High school	Active	Active
18-39 years	8	Female	SPD supporter	High school	None	Active

Interview Experiment Procedure

The interviews were conducted in a semi-structured experimental format, which allowed for flexibility to explore topics in depth while maintaining consistency across interviews. The semi-structured approach is ideal for this research as it facilitates the collection of rich, qualitative data and allows for the exploration of complex and nuanced responses for the in-depth analysis (Given, 2008). The experimental character stems from the fact that participants were initially shown the deepfake video without being informed of its manipulated nature to elicit authentic reactions and measure their cognitive and emotional responses to successfully conduct the followed micro-level analysis. Following the viewing, a series of questions were asked to allocate their perceptions of the video's authenticity, more of their cognitive and emotional responses, and their views on its implications for public trust. After the initial reactions were recorded, participants were informed about the video's deepfake nature, and follow-up questions were asked to explore changes in their perceptions and trust levels on the micro-level, as well as their broader perceptions from the individual to the public sphere (macro-level). It is important to note, that ethical approval was obtained prior to conducting the interviews. Participants were informed about the purpose of the study, assured of their anonymity, and given the option to withdraw at any time. Consent was obtained for recording the interviews, and all data were anonymized during to ensure confidentiality.

3.4 Method of Data Analysis

Thematic Analysis

Thematic analysis was chosen as the primary method for data analysis due to its flexibility and effectiveness in identifying, analyzing, and reporting patterns (themes) within qualitative data (Braun & Clarke, 2006). This approach allows for the detailed examination of participants' experiences and perceptions, aligning well with the research objectives. The thematic analysis approach is particularly suited to this research because it enables the identification of themes related to trust, authenticity perception, and emotional responses, which are central to understanding the impacts of deepfakes (Given, 2008). It also allows for the exploration of how these themes intersect and influence each other, providing a comprehensive understanding of the data (Braun & Clarke, 2006). Both content analysis and interviews operate under the assumption that data is contextually bound (Given, 2008). Understanding the socio-political, historical, and cultural contexts of the data is imperative for the accurate interpretation of the participants' actively and passively declared opinions, emotions, and cognitive responses. This contextual sensitivity adds depth to the analysis but also introduces complexity, necessitating careful consideration of relevant factors to maintain analytical validity (Given, 2008). Furthermore, the identification of implicit messages within texts or spoken words during interviews have the potential for this analysis to reveal underlying ideologies or biases of the researcher or the participants. This aspect of analysis challenges the researcher to navigate interpretive biases while striving for analytical transparency (Given, 2008). Documenting analytical decisions and providing clear justifications are essential strategies for enhancing the credibility of the research findings. Furthermore, the systematic approach is vital in both content analysis and interviews, this, however, requires the data to be broken down into manageable units for the categorization (for instance into the micro, and macro level consequences) and analysis. However, while this systematicity contributes to consistency and rigor, it may also restrict the flexibility needed to explore possible unexpected findings (Given, 2008). That is where coding comes into play, together with Atlas.ti as a tool, the textual data will be categorized before conducting the content analysis, with the contextually right theoretical background, the interpretive nature of this thesis will offer sufficient validity and reliability to not be labeled as a so-called “everything goes” approach.

Nevertheless, this research approach comes with limitations which, firstly within the scope of a bachelor thesis can't be resolved and secondly are natural to a qualitative interview-based approach within the realm of so-called "interpretative"-science. Specifically due to the, to a bachelor's degree limited availability of financial-, time-, and personal academic expertise- resources, only a small sample of interviewees can be investigated. While I wish I could implement more participants with a more diverse background to the sample, it is unrealistic regarding the guidelines of such a bachelor thesis. Additionally, the scope of the research can only be as wide as it is possible to answer the questions within the scope. That said, I wish I could implement the whole theoretical framework of Vasist & Krishnan (2020) into my research which would lead me to also investigate the intention, creation, and dissemination of the "AfD-ban" deepfake and its authors ZPS, as well as more investigation on the possibly more diverse and greater impacts of the case on society and politics, policies, and polities. This thesis can only assess these gaps by highlighting them and suggesting future research in multidisciplinary directions on these topics in the conclusion section.

Use of Atlas.ti Software and Coding

Atlas.ti software was used to facilitate the coding and analysis process. Atlas.ti is a powerful tool for qualitative data analysis, offering features such as code management, memo writing, and data visualization. Its use is justified by its ability to handle large volumes of data efficiently and to support rigorous and systematic analysis (Friese, 2014). The coding process overall involved five steps:

1. Familiarization with the Data: Transcripts were read multiple times to gain an in-depth understanding of the content and to possibly discover certain underlying schemes early.
2. Generation of Initial Codes: Initial codes were created based on key themes derived from the research questions and theoretical framework. These included codes for authenticity perception, emotional response, and trust impacts.
3. Reviewing and Refining Codes: Codes were reviewed and refined to ensure they accurately captured the data's meaning and relevance to the research questions.
4. Development of a Coding Scheme: A coding scheme was developed to systematically categorize the data. The scheme included main themes and sub-themes, each with specific codes to ensure consistency in the analysis.

The Coding Scheme

Code	Group
Anger	Cognitive and Emotional Responses
Authenticity Perception: Unsure about Authenticity	Cognitive and Emotional Responses
Authenticity Perception: Believed Authentic	Cognitive and Emotional Responses
Authenticity Perception: Recognized Unauthenticity	Cognitive and Emotional Responses
Authenticiy Perception: Recognized Deepfake	Cognitive and Emotional Responses
Betrayal	Cognitive and Emotional Responses
Change in Political Opinion	Cognitive and Emotional Responses
Changes in Media Trust	Cognitive and Emotional Responses
Changes in Political Trust	Cognitive and Emotional Responses
Cognitive Responses	Cognitive and Emotional Responses
Concerned with misinformation lvl 1-5	Concerned with misinformation Level
Confusion	Cognitive and Emotional Responses
Decreased Media Trust	Cognitive and Emotional Responses
Decreased Political Trust	Cognitive and Emotional Responses
Distrust	Cognitive and Emotional Responses
Elements influencing perception: Content Cues	Cognitive and Emotional Responses
Elements influencing perception: Visual/Audio Cues	Cognitive and Emotional Responses
Emotional Reactions After Deepfake Reveal	Cognitive and Emotional Responses
Emotional Reactions Before Deepfake Reveal	Cognitive and Emotional Responses
Erosion of Public Trust	Cognitive and Emotional Responses
High Threat	Cognitive and Emotional Responses
Increased Media Trust	Cognitive and Emotional Responses
Increased Political Trust	Cognitive and Emotional Responses
Indifference	Cognitive and Emotional Responses
Insecurity	Cognitive and Emotional Responses
Knowledge lvl Deepfakes 1-5	Deepfake Knowledge Level
Low Threat	Cognitive and Emotional Responses
Lvl of change online information perception 1-5	Change of Online Info Perception Level
Lvl of deepfake´s influence on consumed information 1-5	Deepfake´s Influence on Consumed Info Level
Lvl of deepfake´s influence on trust in media 1-5	Deepfake Influence on Media Trust Level
Lvl of deepfake´s influence on trust in politics 1-5	Deepfake Influence on Political Trust Level
Lvl of potential consequences for democracy 1-5	Potential Consequences for Democracy Level
Lvl of probability that other deepfakes influence the public 1-5	Probability Level other Deepfakes influencing Public
Neutral	Cognitive and Emotional Responses
No Change Media Trust	Cognitive and Emotional Responses
No Change Political Trust	Cognitive and Emotional Responses
No Threat	Cognitive and Emotional Responses
Perceived level of safety from misinformation 1-5	Perceived Level of Safety from Misinformation
Perceived Threat to Democratic Process	Cognitive and Emotional Responses
Political Manipulation	Cognitive and Emotional Responses
Prior Media Trust lvl 1-5	Prior Media Trust Level
Public Awareness	Broader Perceptions of Deepfake Technology
Regulatory Measures	Broader Perceptions of Deepfake Technology
Risks Associated with Deepfakes	Broader Perceptions of Deepfake Technology
Social Polarization	Broader Perceptions of Deepfake Technology
Subject will be more sceptical: negative	Cognitive and Emotional Responses
Subject will be more sceptical: positive	Cognitive and Emotional Responses
Technological Solutions	Broader Perceptions of Deepfake Technology
Thought Process After Reveal	Cognitive and Emotional Responses
Thought Process Before Reveal	Cognitive and Emotional Responses
Trust	Cognitive and Emotional Responses
Trust in AfD lvl 1-5	Trust in AfD Level
Trust in Government lvl 1-5	Trust in Government Level
Trust in politicians lvl 1-5	Trust in politicians Level

Step-by-Step Analysis

1. Initial Coding: Applying initial codes to data segments that aligned with the research questions and theoretical framework.
2. Theme Development: Grouping related codes into broader themes that captured the overarching patterns in the data.
3. Theme Review: Reviewing themes to ensure they were coherent, distinct, and relevant to the research objectives.
4. Defining and Naming Themes: Defining each theme in detail and assigning descriptive names to encapsulate their essence.
5. Report Writing: Synthesizing the themes into a coherent argumentation that addressed the research questions and highlighted key findings.

3.5 Conclusion

This methods chapter has outlined the systematic approach taken to investigate the impacts of the "AfD-Ban" deepfake video. The research activities comprised the Case Selection, where the "AfD-Ban" deepfake video was chosen for its political relevance and impact. This chapter was followed by the explanation of this thesis's Data Collection which consisted of experimental semi-structured interviews with a purposively sampled group of participants, for ensuring diversity in age, education, gender, political affiliation, and digital literacy. Finally, the Data Analysis chapter showcased how and why the thematic analysis was employed, and how Atlas.ti was used for coding.

4. Analysis

4.1 Introduction

This chapter provides an in-depth analysis of the participants' responses to the "AfD-Ban" deepfake video, interpreting these findings through the theoretical frameworks of Vasist & Krishnan (2022) and Jürgen Habermas's Public Sphere Theory (1989). The analysis is structured around the three sub-questions of this thesis: the initial perception of authenticity (4.2), cognitive and emotional responses upon learning the deepfake nature (4.3), and the broader implications on public trust (4.4) followed by a discussion of the findings (4.5). Through a detailed examination of the individual reactions, the first section aims to provide an in-depth understanding of the micro- and macro-level impacts of deepfakes on public trust, to achieve a holistic discussion of the relevant research findings thus answering the research questions.

4.2 Initial perception of authenticity

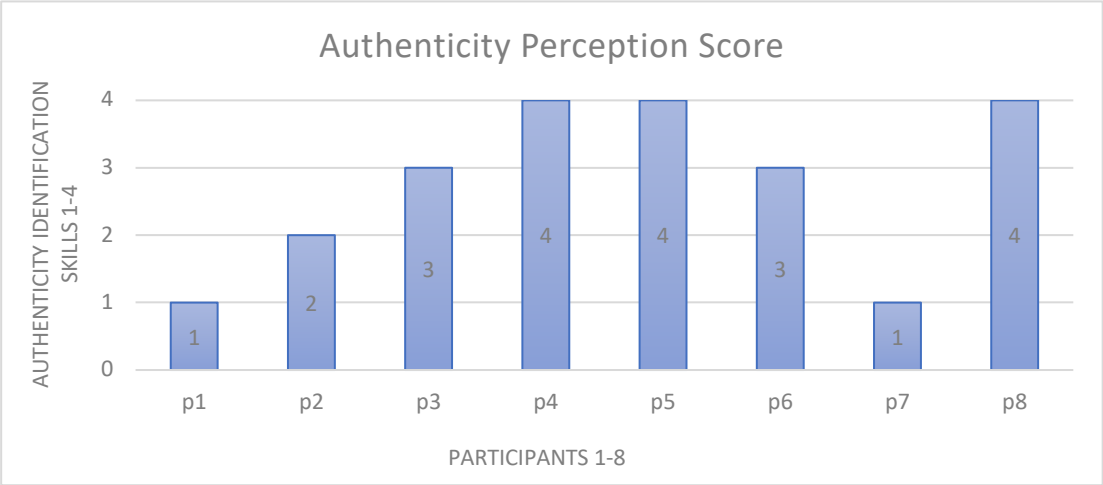


Chart 1: This score represents the participant´s ability in recognizing the deepfake nature of the video. The y-axis indicates the kind of authenticity perception of the participant. "1": Believed Authentic. "2": Unsure about Authenticity. "3" Recognized Unauthenticity. "4"Recognized Deepfake.

Cognitive Biases and Initial Reactions

Starting with the participant observations in their ability to identify the deepfake video´s true nature, this sub-chapter will present the initial reactions relevant for the micro-level analysis.

(P1), an elderly male over 65 who identifies as a CDU supporter with no high school education and minimal digital literacy, initially perceived the video as authentic. His high trust in mainstream media and political figures influenced this perception, aligning with the illusory truth effect, where repeated exposure increases perceived truthfulness (Vosoughi et al., 2018). Relying heavily on traditional media sources likely contributed to his acceptance of the video's authenticity. His experience highlights how deepfakes exploit trust in media institutions, particularly among those with limited digital literacy.

(P2), an elderly female over 65 and a CDU supporter with no high school education, was unsure about the video's authenticity. Despite similar educational background to P1, her active online engagement exposed her to discussions about digital manipulations. Her growing skepticism demonstrates how digital awareness can mitigate cognitive biases. P2's case illustrates that digital literacy is not solely dependent on formal education but also on active engagement with digital content. Her uncertainty about the video's authenticity suggests a growing awareness among older adults about digital deception, which will be further examined in chapter 4.4, discussing micro- and macro-level consequences.

(P3), a female aged between 40 and 64, a Green Party supporter, and an academic with high digital literacy, immediately identified the video as fake due to content inconsistencies. Her academic background and critical thinking skills enabled her to critically engage with the content. P3's experience underscores the necessity of teaching critical analysis skills, not just technical aspects of digital literacy, to better equip individuals to identify deepfakes.

(P4), a female aged between 40 and 64, a CDU supporter, and a high school graduate with active social media use, doubted the video's authenticity from the outset. Her inherent mistrust in political communication and skepticism towards online content influenced by her experiences and social media usage led her to immediately doubt the video's authenticity. P4's case highlights the dual-edged nature of social media: it can be a source of misinformation but also provides a platform for diverse viewpoints fostering critical thinking.

(P5), a male aged between 40 and 64, an "Other Party" supporter with an academic background in social sciences, quickly identified the video as a deepfake. His active social media presence, digital literacy skills, and academic training enabled him to recognize technical inconsistencies. P5's case emphasizes the importance of technical knowledge and digital literacy in detecting misinformation.

(P6), a young male aged between 18 and 39, an AfD supporter with no high school education but active on social media, initially accepted the video due to political bias and lack of formal education. However, his active social media use led him to question the video's authenticity. P6's journey from acceptance to skepticism illustrates the interplay between political bias, education, and digital engagement.

(P7), a young male aged between 18 and 39, an SPD supporter with high school education and active social media use, was misled by the video despite his skepticism and exposure to diverse perspectives. His case highlights that critical evaluation is not only an individual task but also a communal process. P7's case suggests fostering environments for collective discussion can enhance media literacy, though even with skepticism and regular social media interaction, one can still be misled by a deepfake.

(P8), a young female aged between 18 and 39, an SPD supporter with high school education and no social media use but active online media consumption, directly recognized the video as a deepfake. Her critical thinking skills developed through education in media-related fields and active engagement with online news sources aided her recognition. P8's case emphasizes the role of education and active media consumption in fostering resilience against digital misinformation.

Factors Influencing Initial Perception

The initial perceptions of the participants were shaped by several key factors. Participants like P1 and P7, who had high trust in so-called mainstream media, were more likely to perceive the video as authentic. This trust often stems from long-term consumption of traditional news sources, which they consider reliable (Buturoiu et al., 2023). Participants with higher digital literacy, such as P6 and P8, were more adept at recognizing the deepfake because of technical cues, such as a not perfect visualization of mouth movement to the spoken text. Their familiarity with digital manipulation techniques combined with critical thinking skills, and in the case of P8's general distrust in politics-related media, enabled them to scrutinize the video's authenticity correctly. Another factor was the awareness of deepfake technology itself, which influenced the skepticism of participants like P5 and P8. Their prior knowledge provided them with the context to question the video's authenticity. Finally, P4's and P6's mistrust in political communication and P2's skepticism towards online content shaped their initial reactions. These biases, formed from past experiences, played a significant role in their initial skepticism.

4.3 Cognitive and Emotional Responses

Upon discovering the video's deepfake nature, participants experienced a range of cognitive and emotional responses. These responses are critical for understanding the micro-level impacts of deepfakes on individuals, which is the crucial step for answering the second sub-question.

Emotional Responses

(P1) exhibited a strong sense of indifference and skepticism towards the content of the video. He described the video as "flatly informative" and "parroted," indicating that it did not make a significant impression on him. Notably, P1 did not report feelings of betrayal, anger, or confusion upon learning that the video was a deepfake. Instead, he maintained his emotional equilibrium throughout the revelation.

(P2) initially expressed significant fear and concern after watching the video. She mentioned feeling anxious about the potential for such content to destabilize democracy and restrict freedoms, particularly worried about societal divisions and the spread of hate. Her emotions were heightened by the content, leading to visible distress

during the interview. Upon learning that the video was a deepfake, it confirmed her fears about the prevalence of misleading information online.

(P3) felt validated in her skepticism. Her critical stance was confirmed, and although she found the video "shocking" and "extreme," her reaction was more intellectual than emotional. She was concerned about the video's historical parallels and extreme tone.

(P4) maintained a critical stance, experiencing minimal emotional distress. She quickly recognized the video as a deepfake, expressing disbelief at its dramatic tone and content, and frustration with the portrayal of Olaf Scholz. Her pragmatic approach focused on the technology's implications rather than personal emotional responses.

(P5) immediately recognized the video as fake, describing it as "fake news" and expressing disbelief that a legitimate political figure would make such statements. He expressed frustration with the content, noting that it seemed unrealistic and out of character for a politician like Olaf Scholz to make such dramatic claims.

(P6) experienced mixed emotions, including doubt and concern. His critical thinking skills led to a nuanced cognitive and emotional response. P6´s response highlighted his concern for the broader societal implications of deepfakes. He expressed frustration with what he perceived as incitement in the video, viewing it as a misleading portrayal aimed at discrediting the AfD by focusing on isolated incidents rather than the party as a whole.

(P7) felt shocked and deceived, reflecting significant emotional turmoil due to his initial trust in the video's authenticity. P7's reaction included a sense of betrayal and a loss of trust in his usual information sources. His response highlights the vulnerability of individuals even with perceived higher positions in the digital literacy spectrum to manipulation. While he was not able to identify the video as fake, he announced that he was extra cautious of the video’s authenticity due to the experiment situation.

(P8) felt a mix of validation and concern. Her prior knowledge of digital media helped her process the emotional impact of the deepfake revelation. P8's reaction included a recognition of the need for increased vigilance in digital media consumption. She felt uneasy with the violent imagery and the extreme comparisons made in the video, which she found exaggerated and manipulative. Despite this, P8 did not experience significant emotional distress and maintained a level-headed approach to evaluating the video's content. Upon learning that the video was a deepfake, P8 expressed a sense of validation for her skepticism. P8's initial unease and suspicion were confirmed, but this did not lead to significant emotional upheaval.

Cognitive Responses

(P1) demonstrated a high level of critical analysis when discussing the video. He recognized it as a mix of previously known information and perceived it as not impactful. He mentioned that it did not influence his pre-existing opinions about Olaf Scholz or the AfD. Although he was not deeply familiar with deepfake technology, he acknowledged its potential dangers. After the revelation, he recognized the authenticity in the portrayal but maintained his skepticism about the video's impact. P1's trust in media and information remained largely unchanged, scoring it a "two" on a scale where "five" indicates a significant impact.

Cognitively, **(P2)** was initially unsure about the video's authenticity, questioning its validity even as she watched it. Her skepticism was informed by previous knowledge of misinformation, which helped her remain critical of the video's content. The revelation that the video was a deepfake heightened her awareness of the dangers of such technology. This experience further eroded her trust in media and online information, reinforcing her pre-existing skepticism. The deepfake video did not significantly alter P2's existing political opinions. She maintained her distrust towards the AfD and continued to view Olaf Scholz in a generally positive light. However, she expressed a commitment to being more vigilant and critical of the information she encounters, both online and offline, intending to cross-reference information more rigorously before accepting it as true. Ultimately, P2's fundamental views remained unchanged, though her vigilance towards media consumption has been notably heightened.

(P3) critically analyzed the video's content, noting the historical comparisons and the implications of allowing a party like the AfD to gain traction. She found the portrayal of the AfD's impact on society both realistic and concerning. She mentioned the speech seemed unreal and noticed synchronization issues with the mouth movements. P3 had a limited awareness of deepfake technology, rating her familiarity as a two on a scale of one to five. Despite this, her skepticism about the video's authenticity was influenced by her general knowledge and intuition. The deepfake video reinforced P3's pre-existing views against the AfD and her support for liberal parties. She was already skeptical of the AfD and found the video's portrayal consistent with her beliefs about the party's potential dangers.

As in the case of P3, **(P4)** critically analyzed the video's content, noting specific elements that indicated its inauthenticity. She pointed out the unrealistic nature of the themes discussed, such as the five-year anniversary of the Walter Lübke assassination,

which she found implausible and outdated. P4 had a pre-existing awareness of deepfake technology, rating her familiarity as a three on a scale of one to five. This prior knowledge helped her quickly identify the video as a deepfake. Despite recognizing the deepfake, P4's trust in mainstream media remained unaffected. She continues to rely on multiple sources and uses her critical thinking skills to filter information. The deepfake video did not alter P4's political opinions. She maintains a low opinion of Olaf Scholz and distrusts the AfD, viewing them as too radical and untrustworthy. P4's experience with the deepfake reinforced her practice of cross-referencing information from various sources.

Cognitively, **(P5)** critically analyzed the video, identifying specific elements that signaled its inauthenticity. He mentioned the implausibility of the content, such as the proposed actions against the AfD, and noted that such actions would be inconsistent with democratic principles. P5 was already familiar with deepfake technology, rating his knowledge as a four out of five. He quickly identified the video as a deepfake due to subtle visual cues and inconsistencies in the speech synchronization. Despite recognizing the video as a deepfake, P5's trust in mainstream media and democratic institutions remained unaffected. He maintained his critical stance and continued to rely on multiple sources to verify information. The deepfake video did not alter P5's political opinions.

(P6) critically analyzed the video, identifying specific elements that raised doubts about its authenticity. He noted that the dramatic tone and extreme statements were out of character for a politician like Olaf Scholz. P6 was familiar with deepfake technology, rating his knowledge as a two or three out of five. He quickly identified the video as potentially being a deepfake due to its unnatural presentation and inconsistencies. Despite recognizing the video as a deepfake, P6's trust in mainstream media remained low. He rated his trust in mainstream media as a two out of five, citing past disappointments and perceived bias. The deepfake video did not significantly alter P6's political opinions. He remained skeptical of mainstream political parties and maintained a cautious stance towards all political information. P6's experience with the deepfake reinforced his practice of cross-referencing information and remaining skeptical of online content.

(P7) critically analyzed the video's content, noting that the dramatic tone and extreme statements seemed out of character for Olaf Scholz. He found the portrayal of the AfD and the dramatic call to action unconvincing. Despite his skepticism about the video's content, P7 did not initially doubt its authenticity and took the message at face value.

He rated his trust in mainstream media relatively high, at four out of five, but acknowledged the need to be more critical. Upon learning that the video was a deepfake, P7 experienced a sense of surprise and concern. He realized that the emotional manipulation had affected him more than he initially thought. This revelation led to a heightened awareness of the potential dangers of deepfake technology and the need for greater vigilance in evaluating online content. P7 acknowledged that deepfake technology could have serious implications for democracy by misleading the public and influencing political outcomes. He stressed the importance of public education and technological solutions to combat misinformation. The deepfake video did not significantly alter P7´s political opinions. However, the experience reinforced his practice of cross-referencing information and being cautious of emotional and sensationalist content.

(P8) critically analyzed the video's content and presentation. She doubted the authenticity of the video, particularly the idea that Olaf Scholz would establish a website for citizens to denounce others, which she found unrealistic and legally questionable. She reasoned that such a significant initiative would have been widely reported if it were true. P8 was familiar with deepfake technology, rating her knowledge as a three out of five. She noted that the video's facial movements seemed unnatural and that the voice did not perfectly match Olaf Scholz's known voice. The deepfake video did not significantly alter P8's political opinions.

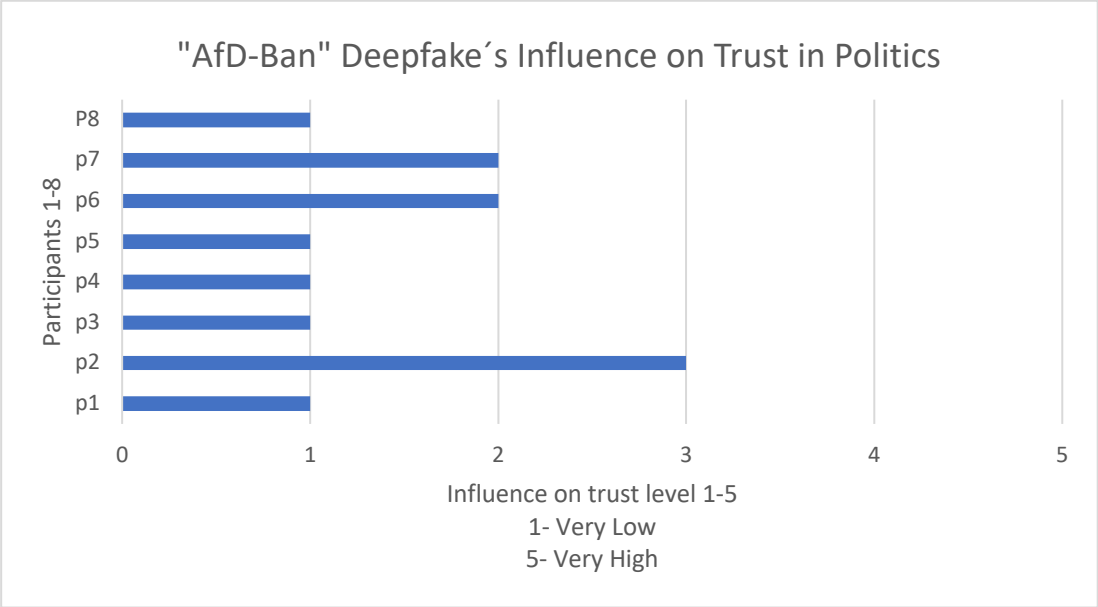
Key Findings and Implications for Theory and Research Question

The findings on the cognitive and emotional responses of participants to the deepfake video reveal a spectrum of reactions that underscore the micro-level impacts of deepfakes on individuals, which are crucial for understanding their broader implications on public trust and democratic discourse (Vasist & Krischnan, 2022). Participants' emotional responses ranged from indifference and skepticism to significant fear, concern, and shock. Cognitively, participants demonstrated varying levels of critical analysis. These findings align with theoretical frameworks on misinformation, reflecting how emotionally charged deepfakes can provoke strong reactions and impact public trust. Emotional responses such as fear, skepticism, and validation highlight the influence of deepfakes on perceptions and trust, while cognitive responses emphasize the role of critical thinking and media literacy in combating misinformation (Saleh & Taieb, 2023). In conclusion, the emotional and cognitive responses to deepfakes demonstrate the significant impact of such

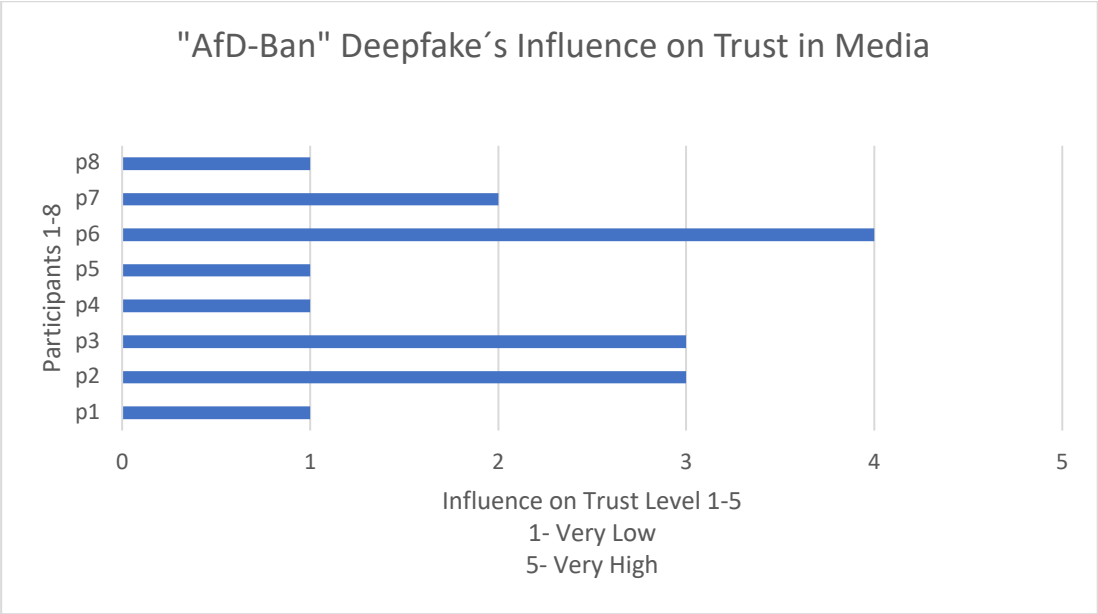
technology on the micro-level, underscoring the need for enhanced media literacy, public education, technological, and regulatory measures to mitigate the effects of deepfake technology on democratic discourse and societal trust.

4.4 Broader Implications on Public Trust

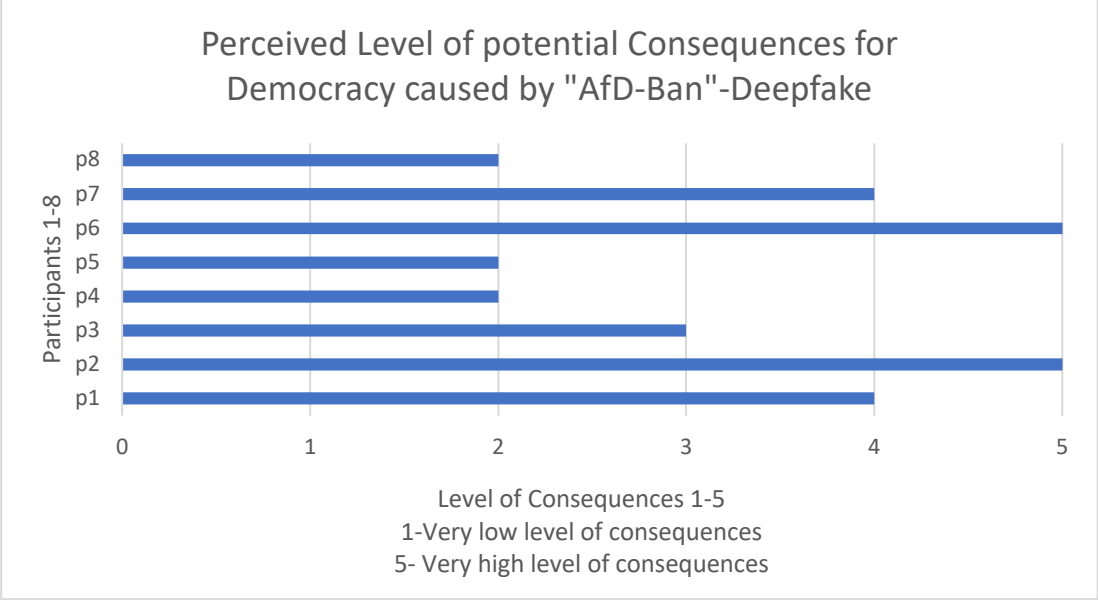
The broader implications of deepfakes on public trust were shown in the participants' reflections on their experiences. These implications are crucial for understanding the macro-level impacts of deepfakes on society. The participants expressed significant concerns about the erosion of trust in media and political institutions due to deepfake technology. Furthermore, they discussed the potential impact of deepfakes on democratic engagement, highlighting the risk of voter manipulation and erosion of democratic processes.



Graph 2: This graph illustrates the perceived impact on participants' trust levels in politics and polities after the deepfake revelation, highlighting an influence on trust experienced by 37,5% of participants.



Graph 3: This graph illustrates the perceived impact on participants' trust levels in media after the deepfake revelation, highlighting an influence on trust experienced by 50% of participants.



Graph 4: This graph highlights the perceived level of potential consequences for Democracy caused by the deepfake video. All participants stated that the „AfD-Ban "-Video has the potential of causing consequences for the democracy.

(P1) exhibited strong skepticism towards the deepfake video but maintained his moderate trust in traditional media. His established skepticism towards political figures, due to their perceived inability to take responsibility, was not significantly altered by the deepfake. This indicates once more that while deepfakes challenge trust, pre-existing skepticism and critical thinking can buffer their impact.

(P2) `s trust in political figures and online content was shaken. The deepfake revelation made her question the authenticity of political communication and increased her overall skepticism. She acknowledged the potential for deepfakes to spread misinformation and negatively influence public opinion. P2 is deeply concerned about

the broader implications of deepfake technology on democracy. She believes such technology can mislead the public, potentially influencing voting behaviors and increasing societal divisions. Despite the emotional impact.

Although **(P3)** felt validated in her skepticism, P3 expressed concerns about the broader implications for public trust. Her critical perspective highlighted the potential for deepfakes to diminish confidence in legitimate media sources and political discourse. P3 acknowledged that deepfake technology could have serious implications for democracy by misleading the public. She emphasized the importance of education and awareness to combat misinformation and prevent manipulation.

(P4) `s inherent mistrust in political communication was reinforced. She saw the deepfake as a confirmation of her suspicions about the manipulative nature of political content, which could further polarize public opinion. She remains vigilant and skeptical towards online content but believes in the importance of education and political awareness to combat misinformation. P4 is concerned about the broader implications of deepfake technology on democracy. She acknowledges that well-made deepfakes have the potential to mislead the public and influence political outcomes, especially among those less familiar with the technology.

(P5) ´s academic background in social sciences made him acutely aware of the risks posed by deepfakes. P5's trust in media was already tentative, and the deepfake validated his belief in the need for robust verification mechanisms and more effective regulatory measures aiming to effectively hold tech companies accountable for their actions.

(P6) `s cognitive and emotional responses underscored the vulnerability of public trust in an era of sophisticated digital manipulations. While P6 advocated for improved media literacy and technological safeguards to mitigate risks of political manipulation, it is necessary to mention, that P6 was the only participant who perceived deepfakes as a sort of external threat which is used by elite actors to politically manipulate the own population. This perception and the inherent political bias, displays the “victim”-narrative typically utilized by right-wing populist political actors, such as AfD, to actively enhance public mistrust in political institutions and to foster illegitimacy of the political enemy (Lengfeld, 2017). This, combined with his skepticism towards mainstream media showcases the importance of public awareness through effective educational programs aiming at enhancing critical thinking and media literacy.

In the case of **(P7)**, the deepfake decreased his overall trust, leading him to question the reliability of digital content in general. P7’s initial acceptance of the video and

subsequent feeling of shock highlighted the emotional and cognitive challenges posed by deepfakes, and their broader implications on trust in media.

While **(P8)** managed to identify the deepfake, she expressed multifaceted concerns about the broader societal impacts of deepfakes and misinformation in general. Her awareness of deepfake technology and its societal risks underscored the need for public education to prevent misinformation and foster an informed citizenry for an inclusive public sphere (Habermas, 1968).

4.5 Discussion

Micro-Level Impacts

The micro-level impacts of deepfake technology on individuals' perceptions and emotional responses are profound and multifaceted. Participants in this study exhibited a range of cognitive and emotional reactions upon discovering the video's deepfake nature. These reactions included skepticism, validation, concern, and emotional distress, underscoring the potential for deepfakes to provoke cognitive dissonance and emotional turmoil.

- **(P1)** maintained moderate trust in traditional media despite his skepticism towards political figures. His pre-existing skepticism and critical thinking appeared to buffer the impact of the deepfake. This aligns with Vosoughi et al. (2018), who argue that cognitive biases, such as the illusory truth effect, can be mitigated by prior critical thinking skills.
- **(P2),** however, experienced a significant shake in her trust in political communication, leading to increased overall skepticism. Her response highlights the vulnerability of individuals with limited digital literacy to deepfake-induced misinformation.
- **(P3)** and **(P5)**, both possessing higher digital literacy and academic backgrounds, were able to identify the deepfake more readily and expressed concerns about its broader implications for public trust. Their reactions underscore the importance of media literacy in mitigating the cognitive and emotional impacts of deepfakes (Hameleers et al., 2020).
- Conversely, **(P6)**, with lower educational attainment but active social media use, exhibited a subtle response combining skepticism with concern for societal manipulation. His reaction illustrates the complex interplay of cognitive and emotional factors in shaping responses to deepfakes.

The varied responses among participants highlight the significant role of individual backgrounds, digital literacy, and pre-existing biases in shaping perceptions of deepfake authenticity. These findings suggest that while deepfakes can exploit existing biases and vulnerabilities, critical thinking and media literacy can serve as protective factors.

Macro-Level Impacts

The broader implications of deepfakes on public trust and democratic discourse are significant, particularly within the framework of Habermas's public sphere theory. The erosion of trust in media and political institutions due to deepfake technology poses serious risks to the integrity of democratic processes and societal cohesion. Participants expressed significant concerns about the erosion of trust in media and political figures.

- **(P1)** and **(P4),** who exhibited inherent mistrust in political communication, saw the deepfake as confirmation of their suspicions about the manipulative nature of political content. This reinforcement of pre-existing skepticism can further polarize public opinion and deepen societal divides (Habermas, 1989).
- The erosion of trust in media sources, as highlighted by **(P2)** and **(P3)**, underscores the potential for deepfakes to create a climate of uncertainty and skepticism, complicating the dissemination of accurate information. Furthermore, the participants highlighted the potential for deepfakes to amplify misinformation and diminish confidence in legitimate media sources. This aligns with the findings of Hameleers et al. (2020), who noted that emotionally charged misinformation is more likely to be believed and shared. The erosion of trust can create a climate of uncertainty and skepticism, further complicating the dissemination of accurate information.
- **(P7)** and **(P8)** emphasized the need for public education to foster an informed citizenry, which echoes Habermas's (1989) emphasis on the importance of rational-critical debate in the public sphere. Finally, the potential for deepfakes to manipulate public opinion and influence voting behaviors poses a significant threat to democratic engagement.
- Participants like **(P5)** and **(P6)**, who recognized the deepfake's manipulative potential, underscored the importance of technological safeguards and regulatory measures to protect democratic processes, even though they were arguing from opposite ideological motivations. Ultimately, this possible risk of

voter manipulation and erosion of democratic discourse highlights the need for robust verification mechanisms and more effective regulatory measures to hold tech companies accountable (Vaccari & Chadwick, 2020).

In conclusion, the discussion of these macro-level impacts within the framework of Habermas's public sphere theory emphasizes the importance of informed and rational discourse for the sake of maintaining democratic legitimacy. As seen in the experiments, The Olaf Scholz deepfake has the potential to challenge the inclusivity and quality of public discourse by creating realistic but false content that can mislead the public and distort political narratives (Sunstein, 2001). The erosion of trust in media and political figures undermines the foundational principles of the public sphere, where accurate information and critical engagement are essential for democratic governance (Friedland & Kunelius, 2023).

Connecting Micro- and Macro-Level findings for Recommendations to Enhance Public Trust

The findings from this research highlight the need for strategies to mitigate the socio-political risks of deepfakes. To enhance public trust in media and political institutions for a flourishing public sphere it requires a multi-faceted approach that includes:

1. Enhancing media literacy and public education. On the micro-level, participants like (P2), who exhibited limited digital literacy, experienced significant increases in skepticism and emotional distress when exposed to deepfakes. This underscores the vulnerability of individuals without critical digital skills. On the macro-level, the erosion of trust in media and political figures, as seen in (P1) and (P4), highlights the broader societal risks. Public awareness campaigns and educational programs must focus on teaching individuals to recognize and critically evaluate digital content, particularly deepfakes and misinformation. This includes promoting an understanding of the technical aspects of deepfakes and the importance of verifying information from multiple sources (Pei et al., 2024).
2. Implementing robust technological safeguards is essential. Participants with higher digital literacy, such as (P3) and (P5), were able to identify deepfakes more readily and expressed concerns about their broader implications. On a macro level, the potential for deepfakes to amplify misinformation and diminish confidence in legitimate media sources was highlighted by (P2) and

(P3). Developing and deploying advanced AI tools to identify deepfake content and integrating these tools into social media platforms and news outlets is critical. Collaborations between tech companies, researchers, and policymakers can help create effective detection mechanisms and ensure their widespread adoption (Chesney & Citron, 2019).

3. Enforcing regulatory measures is also necessary. Participants like (P6), who showed concern for societal manipulation, illustrate the complex interplay of cognitive and emotional factors. Macro-level impacts, such as the need for technological safeguards and regulatory measures emphasized by (P5) and (P6) to protect democratic processes, further justify this recommendation. Policymakers should develop and enforce regulations that require transparency in the creation and dissemination of digital content. This includes labeling deepfake videos and penalizing the intentional spread of false information while protecting individuals' privacy and ensuring that regulatory frameworks are adaptable to evolving technological landscapes (Diakopoulos & Johnson, 2021).
4. Promoting ethical standards in digital media is vital. Micro-level findings show that participants' varied responses underscore the importance of ethical media production in maintaining public trust. The erosion of trust due to deepfake-induced misinformation complicates the dissemination of accurate information. Media organizations, tech companies, and content creators should adhere to ethical guidelines that prioritize accuracy, transparency, and accountability. Promoting ethical journalism and responsible digital content creation can help rebuild trust in media institutions (Vaccari & Chadwick, 2020).
5. Finally, and all including “must-have” is the fostering of a truly inclusive public discourse. The emotional distress and cognitive dissonance experienced by participants highlight the importance of inclusive discourse on a micro level. On the macro level, the reinforcement of pre-existing skepticism and polarization, as seen in participants like (P1) and (P4), can undermine democratic engagement. Ensuring the public sphere remains open to diverse perspectives and supports rational-critical debate is a key factor, especially in an era where populist actors exploit post-truth dynamics to promote their own exclusive realities, to undermine rational discourse as already indicated by Bruns (2023). This includes supporting independent journalism, facilitating

community discussions, and promoting civic education that emphasizes informed and inclusive public discourse (Papacharissi, 2010).

By addressing the cognitive and emotional responses of individuals to deepfakes on the micro level and exploring the broader societal implications on the macro level, this research provides insights into the challenges posed by deepfakes. Implementing strategies such as enhancing media literacy, deploying technological safeguards, enforcing regulatory measures, promoting ethical standards, and fostering inclusive public discourse are essential steps to maintain and improve public trust and protecting and improving the integrity of democratic processes in the digital age. In conclusion, the discussion of the findings reflects the massive impacts of deepfake technology on public trust and democratic engagement. By addressing the cognitive and emotional responses of individuals to deepfakes on the micro-level and exploring the broader societal implications on the macro-level, this thesis provides further insights into the challenges posed by deepfakes and can validate the necessity for finding strategies needed to mitigate the risks of deepfakes.

5. Conclusion

5.1 Addressing the Research Questions

This thesis aimed to explore the impact of deepfake technology on public trust through a case study of the "AfD-ban" deepfake video featuring German Chancellor Olaf Scholz. The study focused on three research questions to understand how deepfakes influence individual perceptions and trust in media and political institutions.

Main Research Question: "How do perceptions and reactions of individuals to the "AfD-ban" deepfake video reflect concerns about the influence of deepfake technology on public trust?"

Participants' perceptions and reactions to the deepfake video reflected significant concerns about the erosion of trust in media and political institutions. Those with higher digital literacy and critical thinking skills were more likely to recognize the deepfake, while individuals with lower digital literacy or stronger political biases were more likely to be misled. Emotional responses ranged from indifference and skepticism to significant fear and concern, which highlights the potential for deepfakes to provoke cognitive dissonance and emotional turmoil, thereby affecting public trust. Overall, the findings indicate that deepfakes can exploit existing biases and vulnerabilities, undermining confidence in legitimate media sources and thus political discourse.

Sub-question 1: "How do individuals initially perceive the authenticity of the "AfD-ban" deepfake video, and what factors influence their judgment?"

Initial perceptions of the video's authenticity varied among participants. Factors influencing these judgments included digital literacy, pre-existing biases, and trust in media. Those with higher digital literacy, such as participants P3, P5, and P8, were more adept at identifying the deepfake due to their familiarity with digital manipulation techniques and critical thinking skills. In contrast, participants with lower digital literacy or stronger political biases, such as P1 and P6, were more likely to perceive the video as authentic initially. This highlights the role of education and awareness in shaping individuals' abilities to critically evaluate digital content.

Sub-question 2: "What are the cognitive and emotional responses after learning of the video's deepfake nature?"

Upon learning about the video's deepfake nature, participants experienced a range of cognitive and emotional responses. These included skepticism, validation, concern, and emotional distress. For instance, P2 felt significant fear and concern initially but experienced relief upon discovering the deepfake nature. P3 and P5 felt validated in their skepticism, while P7 experienced shock and a sense of betrayal. These responses underscore the potential for deepfakes to provoke cognitive dissonance and emotional turmoil, highlighting the need for enhanced media literacy to mitigate these effects.

Sub-question 3: "How do individual reactions to the "AfD-ban" video influence broader perceptions of public trust?"

Individual reactions to the deepfake video significantly influenced broader perceptions of public trust. Participants expressed concerns about the erosion of trust in media and political institutions. For example, P4's inherent mistrust in political communication was reinforced, while P7's initial trust in digital content was significantly eroded. Participants like P8, who recognized the deepfake, expressed concerns about the broader societal impacts, emphasizing the need for public education and technological safeguards. These findings align with theoretical frameworks on misinformation, reflecting how emotionally charged deepfakes can diminish confidence in legitimate media sources and political discourse.

5.2 Broader Implications for Public Trust and Democratic Discourse

The broader implications of deepfakes on public trust and democratic discourse are significant. The study found that deepfakes can erode trust in media and political figures, amplify misinformation, and potentially manipulate public opinion and voting behaviors. This erosion of trust poses a serious risk to the integrity of democratic processes and societal cohesion. Participants expressed significant concerns about the erosion of trust in media and political institutions due to deepfake technology. This erosion can create a climate of uncertainty and skepticism, complicating the dissemination of accurate information. Moreover, the potential for deepfakes to manipulate public opinion and influence voting behaviors poses a significant threat to democratic engagement. Therefore, to mitigate the socio-political risks posed by deepfakes, this research underscores the urgent need for strategies to enhance public trust in media and political institutions via enhancing media literacy, implementing technological safeguards, enforcing regulatory measures, promoting ethical standards, and fostering inclusive public discourse.

5.4 Limitations and Future Research Directions

This research is subject to several limitations. The possibly greatest limitation and weakness of this thesis is its strong national and Eurocentric focus. Of course, it is a case study that gained relevance in the national context, and the public sphere in this case is also broadly tied to the nation-state. However, deepfake technology is by no means a national phenomenon. I would like to take this opportunity to highlight that the risks of deepfakes are of a supranational nature and, as with so many crises, threats, and transformations in general, the weakest are the most affected by the impacts. Therefore, it is important that science responds accordingly and focuses particularly on regions that are most affected by the negative impacts of (technological) transformations. Coming to the practical limitations, due to the scope and resource constraints of a bachelor thesis, the sample size is small and does not represent the broader population in its diversity. Additionally, the study focuses on a specific deepfake case, limiting the generalizability of the findings. The qualitative nature of the research, while providing in-depth insights, does not capture the full spectrum of individual experiences and societal impacts. Further research with larger and more diverse samples, as well as longitudinal studies, are necessary to comprehensively understand the long-term effects of deepfake technology on public trust.

Nevertheless, this thesis contributes to the existing literature by providing an in-depth, individual-focused analysis of the cognitive and emotional impacts of one deepfake case on public trust. By examining the case study of the "AfD-Ban" deepfake video, the research highlights the need for enhanced digital and media literacy and regulatory frameworks to mitigate the socio-political risks posed by deepfakes. Future research should explore the long-term effects of deepfakes on public trust and democratic engagement. Investigating the effectiveness of media literacy programs and technological safeguards in different cultural and political contexts can provide further insights into combating the challenges posed by deepfakes. Additionally, examining the role of social media platforms and regulatory measures in preventing the spread of deepfakes can inform policy development and implementation. While this thesis shall and cannot be seen as a truly gap-filling scientific work, it must be understood as a thesis where the method pursued an experimental approach because future research could build up on this, with more expertise, better sample, and various more resources. In conclusion, this thesis showed how the Olaf Scholz deepfake reflected broader concerns of how deepfake technology presents a profound risk to public trust and inclusive democratic discourse. By addressing the cognitive and emotional responses of individuals and exploring the broader societal implications, this research provides insights into the strategies needed to mitigate the risks of deepfakes and protect the integrity of democratic processes in the digital age.

References

Use of Artificial Intelligence: While it may sound ironic utilizing AI while writing a thesis that criticizes it, it is important to mention, that for the sake of better readability and adequate grammar, some written parts were improved by AI programs such as scribbr and ChatGPT. The researcher is not a native English speaker. However, all ideas and the content stems from the researchers own mind and research. The AI was not used inappropriately, no content was provided or written by AI.

AHMED, S. (2021). Don't believe your eyes: How deepfakes make fake news look real. The Conversation.

BHATTACHERJEE, A. (2012). Social Science Research: Principles, Methods, and Practices. Open Access Textbooks.

BRAUN, V., & CLARKE, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77-101.

BROOKS, D. (2021). The Second Mountain: The Quest for a Moral Life. Random House Trade Paperbacks.

BRUNS, A. (2023). From “the” public sphere to a network of publics: Towards an empirically founded model of contemporary public communication spaces. Communication Theory, 33(2-3), 70–81. **https://doi.org/10.1093/ct/qtad007**

BUTUROIU, R., CORBU, N., & BOTAN, M. (2023). News Consumption Patterns Then and Now: From Traditional Media Repertoires to New Ways of Consuming News. https://doi.org/10.1007/978-3-031-41954-6_5

CALHOUN, C. (1992). Habermas and the Public Sphere. MIT Press.

CAPPELLA, J. N., & JAMIESON, K. H. (1997). Spiral of Cynicism: The Press and the Public Good. Oxford University Press.

CASTELLS, M. (2009). Communication Power. Oxford University Press.

CHESNEY, R., & CITRON, D. K. (2019). Deepfakes and the New Disinformation War: The Coming Age of Post-Truth Geopolitics. Foreign Affairs, 98(1), 147-155.

DIAKOPOULOS, N., & JOHNSON, B. (2021). Anticipating and Addressing the Ethical Implications of AI-Generated News. Digital Journalism, 9(2), 236-255.

FRASER, N. (1990). Rethinking the Public Sphere: A Contribution to the Critique of Actually Existing Democracy. Social Text, 25/26, 56-80.

FRIESE, S. (2014). Qualitative Data Analysis with ATLAS.ti. SAGE Publications.

FRIEDLAND, L. A., & KUNELIUS, R. (2023). The public sphere and contemporary lifeworld: Reconstruction in the context of systemic crises. Communication Theory, 33(2-3), 153–163. **https://doi.org/10.1093/ct/qtad010**

FUCHS, C. (2014). Social Media: A Critical Introduction. SAGE Publications.

GIVEN, L. M. (2008). The Sage Encyclopedia of Qualitative Research Methods. SAGE Publications.

GNATIK, E. N. (2024). Digital transformation of socio-technological reality: Problems and risks. RUDN Journal of Philosophy. Retrieved from **https://journals.rudn.ru/philosophy/article/view/38430**

HABERMAS, J. (1989). The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society. MIT Press.

HABERMAS, J. (2012). The Crisis of the European Union: A Response. Polity.

HAMELEERS, M., BOS, L., & DE VREESE, C. H. (2020). The appeal of media populism: The role of news exposure and trust in public opinion formation. European Political Science Review, 12(1), 45-62.

HARARI, Y. N. (2018). Are we living in a post-truth era? Yes, but that's because we're a post-truth species. Retrieved from **https://ideas.ted.com/are-we-living-in-a-post-truth-era-yes-but-thats-because-were-a-post-truth-species/**

HOU, Y., MCGONIGAL, J., & YU, L. (2024). Deepfake Detection: Advancements and Future Directions. Journal of Artificial Intelligence Research, 61, 1-22.

INSA. (2024). Wahlumfrage 04.06.2024. Wahlrecht.de. **https://www.wahlrecht.de/umfragen/insa.htm**

LENGFELD, H. (2017). Die „Alternative für Deutschland ": Eine Partei für Modernisierungsverlierer? Kölner Zeitschrift Für Soziologie, 38, 379–399.

LUNDBERG, E., & MOZELIUS, P. (2023). How Deep is the Fake? The Potential Effects of Deepfakes on News Media and Entertainment. Researchgate.net. Retrieved from **https://www.diva-portal.org/smash/record.jsf?pid=diva2:1818361**

MARWICK, A., & BOYD, D. (2011). The Drama! Teen Conflict, Gossip, and Bullying in Networked Publics. Proceedings of the ACM Conference on Computer Supported Cooperative Work.

MISHLER, W., & ROSE, R. (2001). What Are the Origins of Political Trust? Testing Institutional and Cultural Theories in Post-Communist Societies. Comparative Political Studies, 34(1), 30-62.

PAPACHARISSI, Z. (2010). A Private Sphere: Democracy in a Digital Age. Polity.

PEI, G., ZHANG, J., HU, M., ZHAI, G., WANG, C., ZHANG, Z., & LI, B. (2024). Deepfake generation and detection: A benchmark and survey. Retrieved from **https://arxiv.org/abs/2403.17881**

PUTNAM, R. D. (2000). Bowling Alone: The Collapse and Revival of American Community. Simon & Schuster.

ROUSSEAU, D. M., SITKIN, S. B., BURT, R. S., & CAMERER, C. (1998). Not So Different After All: A Cross-Discipline View of Trust. Academy of Management Review, 23(3), 393-404.

SALEH, H., & TAIEB, L. (2023). Impact of deepfake technology on social media: Detection, misinformation, and societal implications. Journal of Media and Communication Studies, 12(3), 45-58. Retrieved from **https://www.researchgate.net/publication/374761496_Impact_of_Deepfake_Technology_on_Social_Media_Detection_Misinformation_and_Societal_Implications**

STATISTISCHES BUNDESAMT. (2022). Bevölkerung Deutschlands nach Altersgruppen. Retrieved from **https://www.destatis.de**

SUNSTEIN, C. R. (2001). Echo Chambers: Bush v. Gore, Impeachment, and Beyond. Princeton University Press.

VASIST, P., & KRISHNAN, S. (2022). Deepfakes: An integrative review of the literature and an agenda for future research. Communications of the Association for Information Systems, 51, Article 26. **https://doi.org/10.17705/1CAIS.05126**

VACCARI, C., & CHADWICK, A. (2018). Deepfakes and Disinformation: Exploring the Impact on Democracy. Digital Journalism, 6(8), 1041-1058.

VOSOUGHI, S., ROY, D., & ARAL, S. (2018). The Spread of True and False News Online. Science, 359(6380), 1146-1151.

WESTERLUND, M. (2019). The Emergence of Deepfake Technology: A Review. Technology Innovation Management Review, 9(11), 39-52.

ZENTRUM FÜR POLITISCHE SCHÖNHEIT (2024). Scholz greift durch: Bundeskanzler verbietet AfD – und sich selbst. Second version of deepfake. Youtube 15.03.2024 from: https://youtu.be/GChN6HDvKSw?si=Vn4MCYE9si2aq127

Appendix

Code Groups

Code Group	Size	Justification
Broader Perceptions of Deepfake Technology	7	Reflects participants' general understanding and opinions about deepfake technology as a whole. This aligns with the theoretical concept of deepfakes and misinformation discussed by Vasist & Krishnan (2022).
Change of Online Info Perception	5	Captures how participants' perceptions of online information change after encountering deepfakes. This relates to the cognitive biases and emotional responses outlined by Vosoughi et al. (2018).
Perception and Emotional Responses	12	Addresses the cognitive and emotional reactions of participants upon learning the video is a deepfake. This is directly connected to the key concept of cognitive and emotional responses discussed by Hameleers et al. (2020).
Macro Level Deepfake Influence	6	Considers the broader societal impacts of deepfakes on public trust and democratic processes. This links to the macro-level consequences discussed by Vasist & Krishnan (2022) and Ahmed (2021b).
Deepfake Influence on Media Trust Level	8	Measures the impact of deepfakes on participants' trust in media sources. This is justified by the concept of public trust and its erosion due to deepfakes, as discussed by Chesney and Citron (2019).
Deepfake Influence on Political Trust Level	5	Assesses the influence of deepfakes on trust in political figures and institutions. This aligns with Mishler and Rose (2001) who discuss the origins of political trust and its importance in democratic societies.
Deepfake Knowledge	5	Evaluates participants' prior knowledge and awareness of deepfake technology. This is necessary to understand how pre-existing knowledge affects perception and detection of deepfakes, as discussed in the methodology section.
Deepfake's Influence on Consumed Info Level	6	Explores how deepfakes affect the credibility of information participants consume. This is linked to the discussion on how misinformation influences public perception and trust (Chesney and Citron, 2019).

Code Group	Size	Justification
Broader Perceptions of Deepfake Technology	7	Reflects participants' general understanding and opinions about deepfake technology as a whole. This aligns with the theoretical concept of deepfakes and misinformation discussed by Vasist & Krishnan (2022).
Change of Online Info Perception	5	Captures how participants' perceptions of online information change after encountering deepfakes. This relates to the cognitive biases and emotional responses outlined by Vosoughi et al. (2018).
Perception and Emotional Responses	12	Addresses the cognitive and emotional reactions of participants upon learning the video is a deepfake. This is directly connected to the key concept of cognitive and emotional responses discussed by Hameleers et al. (2020).
Macro Level Deepfake Influence	6	Considers the broader societal impacts of deepfakes on public trust and democratic processes. This links to the macro-level consequences discussed by Vasist & Krishnan (2022) and Ahmed (2021b).
Deepfake Influence on Media Trust Level	8	Measures the impact of deepfakes on participants' trust in media sources. This is justified by the concept of public trust and its erosion due to deepfakes, as discussed by Chesney and Citron (2019).
Deepfake Influence on Political Trust Level	5	Assesses the influence of deepfakes on trust in political figures and institutions. This aligns with Mishler and Rose (2001) who discuss the origins of political trust and its importance in democratic societies.
Deepfake Knowledge	5	Evaluates participants' prior knowledge and awareness of deepfake technology. This is necessary to understand how pre-existing knowledge affects perception and detection of deepfakes, as discussed in the methodology section.
Deepfake's Influence on Consumed Info Level	6	Explores how deepfakes affect the credibility of information participants consume. This is linked to the discussion on how misinformation influences public perception and trust (Chesney and Citron, 2019).

Interview Question	Justification
Are you active on social media? Where? (Ja/Nein)	Establishes the participant's engagement with digital content, essential for understanding their exposure to potential deepfakes and their digital literacy level.
How much do you trust the news and information from so-called mainstream media? (1-5)	Assesses the participant's baseline trust in mainstream media, crucial for understanding the impact of the deepfake on their trust levels. This ties into Habermas's Public Sphere Theory, emphasizing the role of media credibility in democratic discourse.
Follow-up: Can you provide examples of media sources you find trustworthy or untrustworthy? Why do you not trust them?	Provides specific examples to gauge the level of trust in different media sources.
How much trust do you have in politicians to act in the best interest of citizens? (1-5)	Gauges the participant's general trust in political figures, providing a baseline to compare any changes post-exposure to the deepfake.
Follow-up: Are there specific politicians or political parties you trust more or less? Which ones and why?	Identifies specific political figures or parties that the participant trusts more or less, providing context for their trust levels.
What do you understand by misinformation and fake news, and how concerned are you about their spread? (1-5)	Aims to understand the participant's awareness and concerns about misinformation, central to exploring their responses to the deepfake video.
Follow-up: How do you typically verify the accuracy of the information you receive?	Explores the participant's methods for verifying information, indicating their level of media literacy.
Follow-up: How safe do you feel from fake news, and why? (1-5)	Assesses the participant's perceived safety from fake news, relevant to their trust in information sources.
How much do you trust the government and specifically Olaf Scholz to act in the best interest of citizens? (1-5)	Focuses on the trust in the specific political figure featured in the deepfake, providing a direct link to how the deepfake might influence this trust.
Follow-up: Do you believe the government/Olaf Scholz acts truthfully?	Further explores the trust in Olaf Scholz's truthfulness.

How much do you trust the party 'AfD' to act in the best interest of citizens? (1-5)	Explores the participant's trust in the AfD, providing context for how the deepfake might influence opinions about this party.
Follow-up: Do you believe the AfD acts truthfully?	Further examines trust in the AfD's truthfulness.
Follow-up: Should the AfD be banned, and why/why not?	Investigates the participant's opinion on the banning of the AfD, providing deeper insight into their political views.
Can you describe your initial reaction to the video?	Captures the participant's immediate cognitive and emotional responses, addressing the sub-question about initial perceptions and emotional reactions.
What do you think of Olaf Scholz's statements in the video?	Explores the participant's cognitive interpretation of the content before knowing it is a deepfake.
What do you think about the AfD as presented in the video?	Assesses the impact of the video's content on the participant's perception of the AfD.
How has the video influenced or confirmed your opinion about the AfD and Olaf Scholz?	Directly addresses how the deepfake might affect existing opinions, relevant to understanding the influence on public trust.
How did you feel when you first saw the video and thought it was real?	Captures the participant's emotional response before knowing the video is fake, crucial for understanding the initial impact.
Follow-up: Can you describe your emotions in detail?	Provides deeper insight into the participant's emotional state.
How did your feelings about the video change after learning it is a deepfake?	Explores the shift in emotional response, addressing how the revelation of the deepfake affects trust and emotional state.
Follow-up: Did you feel emotions like betrayal, anger, confusion, or uncertainty? Can you describe these emotions?	Examines specific emotional reactions to the discovery that the video is a deepfake.
Were there specific elements that made you believe it was authentic or made you doubt its authenticity?	Addresses factors influencing the perception of authenticity, directly related to the sub-question on judgment factors.
How familiar were you with deepfake technology before this interview? (1-5)	Assesses the participant's prior knowledge, providing context for their ability to recognize the deepfake.

Follow-up: Did your previous knowledge affect your perception of the video's authenticity?	Explores how prior knowledge of deepfakes influenced the participant's judgment.
What thoughts went through your mind when you found out the video is a deepfake?	Explores cognitive reactions and potential changes in trust, addressing the impact on information credibility.
Follow-up: Did the video affect your trust in the information you consume? If so, how? (1-5)	Assesses the impact on trust in information sources.
Do you think this deepfake affects your trust in media? If so, how? (1-5)	Specifically addresses the impact on media trust, directly linking to the research question about the influence on public trust.
Follow-up: Could this example affect your media consumption? If so, how?	Explores potential changes in media consumption habits.
Do you think this deepfake can affect your trust in political figures or institutions? If so, how? (1-5)	Explores the broader impact on political trust, relevant to understanding the implications for democratic processes.
Follow-up: Are there specific aspects of your trust in politics that are affected?	Provides specific details on how political trust is impacted.
Do you think this deepfake has the potential to have consequences for democracy? Why or why not? Which? (No consequences to very significant consequences 1-5)	Assesses the perceived broader societal and political consequences, addressing the macro-level impacts discussed in the theoretical framework.
Follow-up: What measures do you think should be taken to combat deepfakes?	Investigates opinions on potential regulatory measures.
How likely do you think it is that other deepfakes could influence public opinion and/or political outcomes? And why? (1-5)	Gauges the perceived risk and potential influence of deepfakes, linking to concerns about misinformation and its societal impact.
After learning about deepfakes, has your perception of online information changed? How? (How much changed? 1-5)	Assesses changes in perception and skepticism, relevant to the overall impact on trust and information consumption.
Follow-up: Will you be more skeptical of political content you consume in the future? (Yes/No)	Explores future skepticism towards political content.

Relative Occurrence of Codes

	P1 Gr=25	P2 Gr=24	P3 Gr=28	P4 Gr=23	P5 Gr=19	P6 Gr=26	P7 Gr=29	P8 Gr=21
○ Trust in politicians lvl 5	1	0	0	0	0	0	0	1
○ Trust in politicians lvl 4	0	1	0	0	1	0	1	0
○ Trust in politicians lvl 3	0	0	0	0	0	0	0	0
○ Trust in politicians lvl 2	0	0	1	1	0	1	0	0
○ Trust in politicians lvl 1	0	0	0	0	0	0	0	0
○ Trust in Government lvl 5	0	1	0	0	1	0	0	1
○ Trust in Government lvl 4	0	0	1	0	0	0	1	0
○ Trust in Government lvl 3	0	0	0	1	0	0	0	0
○ Trust in Government lvl 2	1	0	0	0	0	1	0	0
○ Trust in Government lvl 1	0	0	0	0	0	0	0	0
○ Trust in AfD lvl 5 Gr=0	0	0	0	0	0	0	0	0
○ Trust in AfD lvl 4 Gr=1	0	0	0	0	0	1	0	0
○ Trust in AfD lvl 3 Gr=0	0	0	0	0	0	0	0	0
○ Trust in AfD lvl 2 Gr=1	0	0	0	0	0	0	0	1
○ Trust in AfD lvl 1 Gr=6	1	1	1	1	1	0	1	0
○ Trust Gr=0	0	0	0	0	0	0	0	0
○ Thought Process Before Reveal Gr=0	0	0	0	0	0	0	0	0
○ Thought Process After Reveal	0	0	0	0	1	0	0	0
○ Technological Solutions	1	0	0	0	0	0	1	0
○ Subject will be more sceptical: positive	0	1	1	0	0	1	1	1

○ **Subject will be more sceptical: negativ**	1	0	0	1	1	0	0	0
○ **Social Polarization** Gr=1	1	0	0	0	0	0	0	0
○ **Risks Associated with Deepfakes** Gr=4	2	0	0	1	0	0	1	0
○ **Regulatory Measures**	0	1	1	0	0	0	1	1
○ **Public Awareness** Gr=6	1	0	0	2	0	1	1	1
○ **Prior Media Trust lvl 5**	0	0	0	0	0	0	0	1
○ **Prior Media Trust lvl 4**	0	1	1	1	1	0	1	0
○ **Prior Media Trust lvl 3**	1	0	0	0	0	0	0	0
○ **Prior Media Trust lvl 2**	0	0	0	0	0	1	0	0
○ **Prior Media Trust lvl 1**	0	0	0	0	0	0	0	0
○ **Political Manipulation**	1	0	0	0	1	1	2	0
○ **Perceived Threat to Democratic Process**	0	1	0	0	0	0	0	0
○ **Perceived level of safety from misinformation 5**	0	0	0	0	1	0	0	0
○ **Perceived level of safety from misinformation 4**	1	0	1	0	0	0	1	0
○ **Perceived level of safety from misinformation 3**	0	0	0	1	0	0	0	1
○ **Perceived level of safety from misinformation 2**	0	0	0	0	0	0	0	0
○ **Perceived level of safety from misinformation 1**	0	1	0	0	0	1	0	0
○ **No Threat** Gr=0	0	0	0	0	0	0	0	0
○ **No Change Political Trust**	0	0	0	0	0	0	0	0

○ **No Change Media Trust**	0	0	0	1	0	0	0	0
○ **Neutral** Gr=1	0	0	0	1	0	0	0	0
○ **Lvl of probability that other deepfakes influence the public**	1	1	1	1	0	1	0	1
○ **Lvl of probability that other deepfakes influence the public**	0	0	0	0	0	0	1	0
○ **Lvl of probability that other deepfakes influence the public**	0	0	0	0	1	0	0	0
○ **Lvl of probability that other deepfakes influence the public**	0	0	0	0	0	0	0	0
○ **Lvl of probability that other deepfakes influence the public**	0	0	0	0	0	0	0	0
○ **Lvl of potential consequences for democracy 5**	0	1	0	0	0	1	0	0
○ **Lvl of potential consequences for democracy 4**	1	0	0	0	0	0	1	0
○ **Lvl of potential consequences for democracy 3**	0	0	1	0	0	0	0	0
○ **Lvl of potential consequences for democracy 2**	0	0	0	1	1	0	0	1
○ **Lvl of potential consequences for democracy 1**	0	0	0	0	0	0	0	0
○ **Lvl of deepfake´s influence on trust in politics 5**	0	0	0	0	0	0	0	0
○ **Lvl of deepfake´s influence on trust in politics 4**	0	0	0	0	0	0	0	0
○ **Lvl of deepfake´s influence on trust in politics 3**	0	1	0	0	0	0	0	0
○ **Lvl of deepfake´s influence on trust in politics 2**	0	0	0	0	0	1	1	0

○ **Lvl of change online information perception 1**	1	0	0	1	1	0	0	0
○ **Low Threat** Gr=0	0	0	0	0	0	0	0	0
○ **Knowledge lvl Deepfakes 5**	0	0	0	0	0	0	0	0
○ **Knowledge lvl Deepfakes 4**	0	0	0	0	1	0	0	0
○ **Knowledge lvl Deepfakes 3**	0	1	0	1	0	0	1	1
○ **Knowledge lvl Deepfakes 2**	1	0	1	0	0	1	0	0
○ **Knowledge lvl Deepfakes 1**	0	0	0	0	0	0	0	0
○ **insecurity** Gr=2	0	1	1	0	0	0	0	0
○ **Indifference** Gr=2	0	0	0	0	1	0	1	0
○ **Increased Political Trust**	0	0	0	0	0	0	0	0
○ **Increased Media Trust**	0	0	0	0	0	0	0	0
○ **High Threat** Gr=1	1	0	0	0	0	0	0	0
○ **Erosion of Public Trust**	0	0	0	0	0	0	1	0
○ **Emotional Reactions Before Deepfake Reveal**	1	2	2	0	0	2	0	0
○ **Emotional Reactions After Deepfake Reveal**	1	0	1	2	0	0	1	1
○ **Elements influencing perception:**	0	1	0	0	0	1	0	1
○ **Elements influencing perception: Content**	0	1	1	1	1	2	0	2
○ **Distrust** Gr=1	0	1	0	0	0	0	0	0
○ **Decreased Political Trust**	0	1	0	0	0	0	1	0
○ **Decreased Media Trust**	0	0	0	0	0	0	1	0

○ **Lvl of deepfake´s influence on trust in politics 1**	1	0	1	1	1	0	0	1
○ **Lvl of deepfake´s influence on trust in media 5**	0	0	0	0	0	0	0	0
○ **Lvl of deepfake´s influence on trust in media 4**	0	0	0	0	0	1	0	0
○ **Lvl of deepfake´s influence on trust in media 3**	0	1	1	0	0	0	0	0
○ **Lvl of deepfake´s influence on trust in media 2**	0	0	0	0	0	0	1	0
○ **Lvl of deepfake´s influence on trust in media 1**	1	0	0	1	1	0	0	1
○ **Lvl of deepfake´s influence on consumed**	0	0	0	0	0	0	0	0
○ **Lvl of deepfake´s influence on consumed**	0	0	0	0	0	0	0	0
○ **Lvl of deepfake´s influence on consumed**	0	0	0	0	0	1	1	0
○ **Lvl of deepfake´s influence on consumed**	1	0	1	0	0	0	0	0
○ **Lvl of deepfake´s influence on consumed**	0	1	0	1	1	0	0	1
○ **Lvl of change online information perception 5**	0	0	0	0	0	0	0	0
○ **Lvl of change online information perception 4**	0	1	0	0	0	0	0	0
○ **Lvl of change online information perception 3**	0	0	1	0	0	1	1	0
○ **Lvl of change online information perception 2**	0	0	0	0	0	0	0	1

○ **Confusion** Gr=2	0	0	1	0	0	0	1	0
○ **Concerned with misinformation lvl 5** Gr=2	0	2	0	0	0	0	0	0
○ **Concerned with misinformation lvl 4** Gr=3	0	0	0	1	1	1	0	0
○ **Concerned with misinformation lvl 3** Gr=4	1	0	1	0	0	0	1	1
○ **Concerned with misinformation lvl 2** Gr=0	0	0	0	0	0	0	0	0
○ **Concerned with misinformation lvl 1** Gr=0	0	0	0	0	0	0	0	0
○ **Cognitive Responses**	4	0	3	2	1	5	3	2
○ **Changes in Political Trust**	0	1	0	0	0	0	1	0
○ **Changes in Media Trust**	0	1	0	0	0	0	1	0
○ **Change im Political Opinion**	0	0	2	0	0	0	1	0
○ **Betrayal** Gr=0	0	0	0	0	0	0	0	0
○ **Authenticiy Perception: Recognized**	0	0	0	1	1	0	0	1
○ **Authenticity Perception: Recognized**	0	0	1	0	0	1	0	0
○ **Authenticity Perception: Believed Authentic**	1	0	0	0	0	0	1	0
○ **Authenticity Perception: Unsure about Authenticity**	0	1	0	0	0	0	0	0
○ **Anger** Gr=1	0	0	0	0	0	0	1	0

*Trancripts can be found in the separate appendix