Master Degree Project

UNIVERSITY OF SKÖVDE

The Cybersecurity Threat of Deepfake

Master Degree Project in Informatics with a specialization in Privacy, Information and Cyber Security

Second Cycle 15 credits

Spring term 2024

Student: Johan Brandqvist

Supervisor: Martin Lundgren

Examiner: Marcus Nohlberg

ABSTRACT

The rapid advancement of deepfake technology, utilizing Artificial Intelligence (AI) to create convincing, but manipulated audio and video content, presents significant challenges to cybersecurity, privacy, and information integrity. This study explores the complex cybersecurity threats posed by deepfakes and evaluates effective strategies, to prepare organizations and individuals for these risks. Employing a qualitative research approach, semi-structured interviews with cybersecurity- and AI experts were conducted to gain insights into the current threat landscape, the technological evolution of deepfakes, and strategies for their detection and prevention.

The findings reveal that while deepfakes offer opportunities in various sectors, they predominantly also pose threats such as misinformation, identity theft, and fraud. This study highlights the dual-use nature of deepfake technology, where improvements in creation and detection are continually evolving in a technological arms race. Ethical and societal implications are examined, emphasizing the need for enhanced public awareness and comprehensive regulatory frameworks to manage these challenges.

The conclusions drawn from this research underscore the urgency of developing robust, AI-driven detection tools, advocating for a balanced approach that considers both technological advancements and the ethical dimensions of these innovations. Recommendations for policymakers and cybersecurity professionals include investing in detection technologies, promoting digital literacy, and fostering international collaboration to establish standards for ethical AI use. This thesis contributes to the broader discourse on AI ethics and cybersecurity, providing a foundation for future research and policy development in the era of digital manipulation.

Keywords: **Deepfake**, **Cybersecurity, Artificial Intelligence, Experts**

Acknowledgments

I want to send my deepest gratitude to my significant other, who, (un)fortunately had to have a lot of alone time during this project. A special thanks to my supervisor Martin, who was always reachable as a solid sounding board, and to my examiner Marcus, who provided not only a foundation to build upon but also invaluable feedback. Thank you, your support has been immense.

Table of Contents

1. Introduction

The arrival of deepfake technology has introduced a new era in digital media, where the distinction between real and fake is increasingly blurred. Deepfakes, derived from the concepts of Deep Learning (DL) and "fake," employ Artificial Intelligence (AI) and Machine Learning (ML) techniques to create convincingly altered videos and audio recordings. This technology enables the manipulation of media so effectively that it can portray individuals saying or doing things they never actually did, posing challenges to information integrity and personal privacy (Kaushal et al., 2022). Historically, the technology underlying deepfakes has evolved from simple image manipulation to complex video alterations powered by Generative Adversarial Networks (GAN). The term "deepfake" was popularized around 2017 when an anonymous user began sharing realistic video manipulations on the internet. This sparked both public fascination and concern as these manipulations quickly showed potential for misuse in creating political misinformation, celebrity pornographic videos, and other malicious content. The sophistication of deepfakes means they can be indistinguishable from genuine content to the untrained eye, raising concerns about their potential to deceive individuals, manipulate public opinion, and disrupt democratic processes. One publicized example of deepfake technology in action was a video of former U.S. President Barack Obama, created by Oscar-winning BuzzFeed director Jordan Peele in 2018. In this video, Peele used AI to manipulate a video of Obama, making it appear as though the former president was voicing Peele's words (Vaccari & Chadwick, 2020). This demonstration showed the potential for deepfakes to spread misinformation and fake news, contributing to a growing awareness and concern over the technology. Another notable instance involved a deepfake of Facebook CEO Mark Zuckerberg, wherein he seemingly made statements about the power over users' data, complimenting Spectre, the fiction criminal network from James Bond (Westerlund, 2019). This video further underscores the technology's potential for creating propaganda and misinformation.

In the chapters that follow this introduction, there will first of all be a presentation of the research question, with problem definition, motivation, limitations, and expected result. Then there will be an exploration of deepfake technology and its implications for cybersecurity, starting with the background chapter, which lays the groundwork by exploring and explaining the evolution and technical foundations of deepfakes and their implications. Following this, the methodology chapter details the approach, centered around semi-structured interviews with experts. The implementation chapter goes further into how this methodology is put into practice, including the interview process and the employment of the laddering technique. Findings from these interviews are then presented in the analysis and results chapter, offering insights into the current and potential future cybersecurity challenges posed by deepfakes. The discussion chapter contemplates these findings, considering their broader significance and suggesting pathways for future research and strategy development. This leads to the conclusion chapter, which generates the study's key insights, emphasizing the urgency of addressing deepfake-related security threats.

1.1. Research Question and Aim

This study aims to assess the impact and mitigation strategies of deepfake technology on cybersecurity, how to face the threat, and what to expect by

interviewing professionals at the forefront of research. For the RQ itself, the following was examined:

"How should organizations and individuals prepare for the cybersecurity challenges posed by deepfake technology?"

1.2. Problem Definition

The arrival of deepfake technology has introduced capabilities for generating artificial content that closely mimics reality, challenging traditional beliefs of trust and authenticity in digital media—this technology's rapid evolution and accessibility present cybersecurity challenges for organizations and individuals alike (Tolosana et al., 2020). The ability to create convincingly manipulated videos and images has implications for privacy, information integrity, and cybersecurity, raising concerns about the spread of misinformation, impersonation, and fraudulent activities. As deepfakes become more indistinguishable from genuine content, the need to identify, prepare for, and mitigate these cybersecurity threats becomes vital. This study aims to explore the impact of deepfake technology on cybersecurity, focusing on privacy invasion and the integrity of information, and seeks to identify effective strategies for organizations and individuals to counteract these challenges.

1.3. Motivation

While deepfakes have been acknowledged as a threat, there is a lack of comprehensive research analyzing their direct impact on cybersecurity and effective countermeasures. This interview study aims to fill this gap by focusing on the cybersecurity threats posed by deepfakes and exploring potential countermeasures. Continued, the motivation for this study stems from the evolution of deepfake technology and its growing implications for cybersecurity, aligning with the thoughts of Chen et al. (2021) and Zhu et al. (2020), which mentions that despite advancements in detection technology, the quality of video forgery still worryingly improves, at a rapid pace. As digital content becomes easier to manipulate with high realism, the potential for misuse grows, posing threats to personal privacy, information integrity, and public trust. This study is driven by the need to understand and develop effective strategies to mitigate these risks. The invasive nature of deepfakes in political, social, and personal realms highlights a gap in current research—particularly in understanding the depth of cybersecurity threats they pose and how they can be effectively countered. The motivation is further reinforced by the potential societal impact of unchecked deepfake technology, which could undermine democratic institutions and processes by spreading disinformation and eroding public trust in media.

1.4. Limitations

This study faces some limitations that may affect the scope and depth of the findings. Firstly, the qualitative nature of the study, while rich in detailed insights, limits the generalizability of the results. The findings are based on the experiences and opinions of a select group of experts, which may not represent all possible perspectives within the field of cybersecurity. Secondly, the pace of technological development in AI and deepfake generation may outpace the applicability of the current detection and mitigation strategies discussed in this study. Additionally, the reliance on semi-structured interviews, though valuable for in-depth exploration, is subject to the interpretations and biases of both the interviewer

and the participants. Lastly, the specific focus on cybersecurity aspects of deepfakes might overlook broader societal, psychological, and political implications.

1.5. Delimitations

This study deliberately narrows its focus to specific areas to maintain clarity and depth in its investigation of the cybersecurity threats posed by deepfake technology. The research exclusively targets professionals with expertise in cybersecurity, AI, and digital content management. By selecting this particular demographic, the study aims to gather insights from those who are directly engaged with or affected by deepfake technology in their professional capacities, while intentionally excluding perspectives from laypersons or users who might interact with deepfakes in a non-professional context.

Geographically, while deepfake technology has global implications, this study limits its participant pool primarily to professionals accessible via the professional network LinkedIn, complemented by an interview with a representative of a Swedish authority. This focus is designed to provide a manageable and relevant sample within the scope of a single study, though it may not fully capture the diverse international strategies and impacts of deepfakes since no particular geographical filter was taken into consideration when reaching out to interviewees.

Methodologically, this study limits itself to semi-structured interviews and does not incorporate quantitative measures or experimental techniques. This approach allows for an exploration of expert opinions and perceptions but inherently limits the study's capacity to provide statistical generalizations about the effectiveness of different deepfake countermeasures.

By setting these boundaries, the study ensures a focused exploration within the defined areas. While these choices deliberately limit the breadth of contexts and perspectives that might otherwise inform a broader understanding of deepfake-related cybersecurity challenges, they allow for a concentrated examination of the issues at hand, aiming to provide nuanced insights into an evolving threat landscape.

1.6. Expected Result

The expected result of this study is to provide an overview of the current threat landscape posed by deepfakes and to identify effective countermeasures and strategies for mitigating these risks. It is anticipated that the study will clarify the complex interplay between technological advances in deepfake creation and the evolving strategies for their detection and mitigation. Through expert interviews, the study aims to reveal foundational beliefs and strategies that underpin expert responses to deepfake threats. Ultimately, this research expects to contribute insights into the effectiveness of current strategies, propose recommendations for strengthening cybersecurity against deepfake challenges, and contribute to battling deepfakes by preparing organizations and individuals for what is about to happen. These results are intended to inform both policymakers and cybersecurity professionals, guiding future efforts to secure digital media and communications.

2. Background

In the digital age, the increase of deepfake technology presents a dual-edged sword, offering advancements in creative fields while posing threats to information integrity, personal security, and cybersecurity. This chapter lays the groundwork for understanding the intricacies of deepfakes, from their technical underpinnings to the evolving landscape of cybersecurity threats they generate, including ethical dilemmas anticipated in the cybersecurity field. Through an exploration of preliminaries, this chapter delves into the mechanisms and methodologies enabling the creation and detection of deepfakes. The "Threat Landscape" section examines the broader implications of these technologies, setting the stage for a focused discussion on "Current Threats and Implications" where the current and potential future challenges of deepfakes in cybersecurity are analyzed. After this, the techniques used in creating deepfakes are discussed, to further explain the foundations of deepfake technology. The last section is related research, where this study aims to align and extend by hopefully contributing new insights into the evolving field of deepfake technology in a preventative way. This approach aims to equip readers with a nuanced understanding of deepfake technology's impact on society, highlighting the urgency of developing countermeasures.

2.1. Deepfake Technology

The emergence of deepfake technology has posed challenges to the integrity of digital content where the anticipation of emerging ethical issues in cybersecurity, including nuances of deepfake technology, underscores the complexity of maintaining digital integrity and privacy in a rapidly evolving cyber threat landscape (Pawlicka et al., 2023). Utilizing sophisticated AI and ML algorithms, deepfakes can generate realistic images, audio, and videos, making it increasingly difficult to distinguish between genuine and manipulated content. Frolov et al. (2022) contribute to this discourse by emphasizing the speed at which deepfake technology is advancing, alongside the development of increasingly sophisticated methods for their creation and propagation. This evolution raises concerns for information security, as it facilitates the spread of misinformation and presents unique challenges in the authentication of digital content. The work of Frolov et al. (2022) broadens the understanding of deepfake technology's impact, not just in terms of potential misuse, but also regarding the need for effective and innovative detection methods to counteract these emerging threats.

The literature review by Rana et al. (2022) categorizes detection methods into four main approaches: DL-based, classical ML-based, statistical, and blockchain-based solutions. DL-based methods are highlighted for their effectiveness in identifying subtle inconsistencies in deepfake-generated content, leveraging the power of neural networks to analyze visual and auditory cues that distinguish genuine from manipulated media. Classical ML techniques, though not as advanced as their DL counterparts, provide insights into the texture, frequency, and pattern differences between real and fake content. Statistical methods focus on identifying anomalies in the data that would not occur naturally, while blockchain-based solutions offer a novel approach to ensuring the authenticity of digital content through secure and permanent record-keeping. The article's analysis of 112 studies underscores the dynamic and complex nature of deepfake detection research. It reveals the strengths and weaknesses of each method, suggesting that a hybrid approach combining the precision of DL with the reliability of

blockchain technology might offer the most robust defense against deepfake manipulations. Furthermore, as deepfake technology continues to evolve, so too must the strategies for its detection. The review from Rana et al. (2022) emphasizes the importance of continuous research, interdisciplinary collaboration, and the development of innovative technologies to safeguard the authenticity of digital media. The fight against deepfake content is not only a technical challenge but also an act to protect democratic processes, personal privacy, and the trustworthiness of information in the digital world.

Additionally, Rana et al. (2024) continue to highlight the role of continuous innovation and interdisciplinary collaboration in enhancing the resilience of digital media against deepfake manipulations. By integrating blockchain and steganalysis, their proposed solution sets a new benchmark for the detection and prevention of deepfakes, highlighting the potential for these technologies to secure digital content authenticity and integrity.

Continued, the insights from Di Dario et al. (2023) into the practical challenges of implementing security testing in large IT organizations explain the complexity of establishing effective defense mechanisms against cyber threats, including deepfakes. This context underscores the need for developing standardized methodologies and enhanced communication strategies within the cybersecurity community to address these threats.

2.2. Deep Learning, Machine Learning, and Generative Adversarial Networks

Techniques such as Deep Learning (DL), Machine Learning (ML), and Generative Adversarial Networks (GAN) are all vital in deep fake technology. Traditional ML techniques have been critical in early efforts to detect manipulations in digital media and identify inconsistencies in images. These methods, while effective against certain types of manipulations, often struggle with the complexity and realism of deepfake images generated by advanced DL models, indicating the need for more sophisticated approaches. DL represents an advancement in the capacity to analyze and understand complex data patterns, as highlighted by Remya Revi et al. (2021). DL, particularly through the use of Convolutional Neural Networks (CNN), has become the cornerstone of modern deepfake detection methodologies. These networks automatically learn hierarchical features from data, making them good at identifying subtle cues and alterations in images that might indicate manipulation. The article reviews various DL-based techniques for deepfake detection, showcasing the evolution from reliance on handcrafted features to automated feature extraction and classification directly from raw data.

GANs, represent both the source of the challenge and a pathway to a solution in the context of deepfake generation and detection. GANs, with their unique architecture comprising a generator and a discriminator network, have revolutionized the ability to create realistic fake images. Remya Revi et al. (2021) explain how the adversarial training process of GANs produces images that are increasingly difficult to distinguish from real ones, underscoring the need for effective detection mechanisms. It further explores the intricacies of different GAN architectures and their implications for the development of detection methods.

Remya Revi et al. (2021) illustrate the connected roles of ML, DL, and GANs within the environment of deepfake creation and detection. While ML laid the groundwork for digital image analysis, DL, through the use of neural networks

has enhanced the ability to detect manipulated content. GANs, meanwhile, embody the evolving challenge of artificial image generation, pushing the boundaries of what is possible in digital image creation and necessitating continual advancements in DL-based detection strategies. Remya Revi et al. (2021) position ML and DL as critical components in the ongoing effort to combat the challenges posed by GAN-generated deepfakes. It highlights the progression from traditional ML techniques to advanced DL models in detecting sophisticated manipulations, underscoring the dynamic and adversarial nature of technological advancements in digital image creation and verification.

GANs, as detailed by Creswell et al. (2018) present a framework in the domain of DL for generating complex, high-dimensional data distributions. Established in 2014, GANs operate on an innovative principle where two neural networks, namely the generator and the discriminator, engage in a competitive training process. This dynamic is often analogized with an art forger (generator) attempting to create convincing forgeries, while an art expert (discriminator) tries to distinguish between the authentic and the forged.

The generator network aims to generate data that mimics the real data as closely as possible, without having direct access to the actual data instances. It learns to produce realistic data through feedback received indirectly from the discriminator. On the other hand, the discriminator network is trained to classify data as either real (originating from the actual dataset) or fake (generated by the generator). This setup encourages the generator to improve its data generation capabilities progressively. Both networks are usually composed of convolutional and/or fully connected layers, making them differentiable, which is crucial for the propagation of errors used during training. The generator maps from an underlying space, often characterized by a random noise distribution, to the data space, thereby indirectly modeling the data distribution. An important aspect of GANs is their applicability across a broad spectrum of tasks including image synthesis, semantic image editing, style transfer, and image super-resolution. This versatility is attributed to the deep representations that GANs learn, which can capture complex patterns and details of the data distribution without requiring extensively annotated training data. However, training GANs presents several challenges, notably the issues of mode collapse where the generator produces limited diversity in outputs and training instability, often manifested as a failure to achieve merging between the generator and discriminator. Despite these challenges, GANs remain an influential model in the AI and DL community, with ongoing research focused on addressing these challenges and expanding their application domain (Creswell et al., 2018).

2.3. Threat Landscape

As we navigate the future of cybersecurity, ethical considerations become increasingly complex, reflecting a broad spectrum of concerns from privacy breaches to the ethical implications of AI in cybersecurity practices (Pawlicka et al., 2023). The threat landscape is not only defined by the technological capabilities of deepfake creators but also by the ethical challenges posed by the deployment of such technologies.

2.4. Current Threats and Implications

The survey by Mirsky & Lee (2021) insights into the misuse of deepfakes highlight concerns for security, privacy, and misinformation. It underscores the potential for deepfakes to undermine public trust, manipulate political discourse,

and infringe on individual rights. Furthermore, technologies like AI and the Internet of Things (IoT), alongside ethical considerations such as privacy, consent, and the risk of biases in AI-driven systems, present challenges in maintaining cybersecurity while upholding ethical standards (Pawlicka et al., 2023). This discussion is paramount for framing the real-world significance of deepfake technology, stressing the need for security measures and ethical guidelines to mitigate these threats and safeguard societal norms.

The investigation from Malik et al. (2022), explore that deepfake technology also sheds light on the complex threats posed by the misuse of this technology. Highlighting the potential for generating misleading information, privacy invasion, and security breaches, the study calls attention to the need for robust detection mechanisms. The societal and ethical implications discussed underscore the importance of advancing deepfake detection capabilities to safeguard against the malign use of AI in fabricating realistic digital content.

Shoaib et al. (2023) delve into the interplay between deepfake technology and the creation of misinformation and disinformation, underscoring a societal challenge. The study illustrates how deepfakes amplify the risks associated with misinformation, affecting not only personal reputations but also democratic processes and public trust. It emphasizes the role of frontier and generative AI technologies in intensifying these challenges, highlighting the need for a multidimensional approach to address the spread of AI-generated false information. Continued, Shoaib et al. (2023) explore the ethical dilemmas posed by the misuse of AI in creating convincing yet fraudulent content. They advocate for the development of ethical guidelines and frameworks that govern the use of AI in content creation, suggesting a balance between innovation and the protection of individual rights and societal values.

The interview study by Di Dario et al. (2023) emphasizes the role of continuous innovation and interdisciplinary collaboration in enhancing digital security measures, a principle directly applicable to the fight against deepfake content. By integrating advanced security testing frameworks and fostering open dialogue between cybersecurity professionals, we can develop more robust strategies for detecting and mitigating deepfake-related risks.

2.5. Research background

The study by Goh et al. (2022) emerges as a resourceful piece of research, it delves into how individuals discern the authenticity of videos in the age of deepfake technology. By conducting semi-structured interviews, the study reveals seven main strategies users deploy to identify deepfakes. This research underscores the gap between technological advancements in deepfake detection and the human capacity to manually verify video authenticity. It suggests a need for educational initiatives and tools to enhance public awareness and capabilities in recognizing deepfakes, thereby contributing to a more secure digital environment. The study by Goh et al. (2022) is practical for understanding the human element in cybersecurity defense mechanisms against deepfake threats, making it an important reference for discussions on the impact of deepfake technology on cybersecurity.

The insights from Di Dario et al. (2023) provide a perspective on the operational challenges and strategies employed within the IT industry for security testing. Their interview study, conducted through structured and semi-structured interviews with security experts, sheds light on the realities of implementing security measures in large IT organizations. This reflects the broader cybersecurity

ecosystem's readiness to address emerging threats such as deepfakes. Specifically, they highlight the variation in security testing practices and the importance of adapting these practices to keep pace with technological advancements and evolving threat landscapes. The findings underscore the need for continuous research, interdisciplinary collaboration, and the development of innovative technologies—principles that are equally applicable to the fight against deepfake content. This aligns with the research aim of this study, to explore effective countermeasures against deepfake technology, emphasizing the need for a multilayered approach that includes not only technical solutions but also organizational readiness and collaborative efforts across the cybersecurity community.

Robila (2023) highlights the potential of experiential learning in computer science education, suggesting that hands-on experience with AI and security testing can enhance understanding and innovation in combating cybersecurity threats like deepfakes. This approach aligns with the aim of this study, to understand deepfake technology's cybersecurity implications, emphasizing the importance of integrating practical, experiential learning into cybersecurity training and education to prepare the next generation of experts equipped to tackle such threats. The insights from Robila (2023) into the effectiveness of experiential learning in computer science education could be useful in devising new strategies for public awareness and professional training in cybersecurity. By incorporating hands-on, practical experiences into learning pathways, individuals and organizations can better prepare for and respond to the unique challenges presented by deepfake technology.

The work of Pawlicka et al. (2023) offers an exploration into the future ethical challenges in cybersecurity. The study highlights the significance of both "strong" and "weak" signals in the ethical landscape, revealing an array of anticipated ethical dilemmas in cybersecurity ranging from privacy breaches to the misuse of AI. Their investigation into the "black swans" of cybersecurity ethics emphasizes the unpredictable, yet impactful, ethical issues that may arise as technology continues to advance. Their insights into the necessity of ethical vigilance and the development of forward-thinking ethical guidelines in cybersecurity practices provide a perspective for framing the broader implications of technological advancements like deepfakes on societal norms and values.

The survey by Mirsky & Lee (2021) offers a foundation for understanding deepfake technology's evolution, showcasing advancements in generative DL that enable realistic media combinations. This article outlines the technical methodologies behind deepfake creation and detection, illuminating the arms race between generating deepfakes and developing detection mechanisms. It provides context for this study, emphasizing the importance of continuous innovation in detection strategies to combat increasingly sophisticated deepfakes.

Malik et al. (2022) categorize the creation and detection methodologies of deepfakes, presenting an analysis of the latest advances in DL technologies. The study underscores the evolution of deepfake generation techniques, the development of detection models, and the diverse datasets central to training these models. This overview lays a foundation for understanding the technical intricacies and methodological approaches in combating deepfake technologies.

Kaushal et al. (2022) did a study comparing different detection tools from ten different articles, where some articles got a very high detection rate of identifying deepfake, for their specific tool. However, the study's analysis underscores the advancements and varied effectiveness of deepfake detection methods.

3. Methodology

This study employs a qualitative research approach through semi-structured interviews to explore six experts' opinions on deepfake-related cybersecurity threats. It was first imagined as a literature review, but it felt as if there was not enough data to conduct one yet due to the relatively new subject. A discussion with teachers at the University of Skövde led to the idea of doing an interview study instead. An interview study felt well-suited for this work as it allowed for a deep exploration of experiences and opinions from experts. To ensure the collection of rich, qualitative data, this study employs both audio and text-based interviews, catering to the preferences and availability of the participants. This approach draws inspiration from Guion et al. (2011), which outlines an approach to in-depth interviewing. The approach emphasizes the use of open-ended questions to produce detailed insights from participants, a semi-structured format that combines pre-planned questions with the flexibility to adapt to the conversation's flow, and the importance of active listening and interpretation to understand the respondent's feelings and perspectives. In line with Guion et al. (2011), an interview guide was developed to feature questions that encourage expansive responses to ensure conversations extend beyond simple affirmatives or negatives. The semi-structured nature of the interviews allows for conversational depth, making it possible to explore unforeseen areas that may arise during interviews.

As mentioned in the previous paragraph, this study utilizes a carefully developed interview guide, drawing from the literature. The guide features a series of open-ended questions designed to explore the impact, challenges, and countermeasures associated with deepfake technology in cybersecurity. Further, the study employs semi-structured interviews to navigate the nuanced landscape of deepfake technology's cybersecurity implications. According to Rowley (2012), semi-structured interviews provide a flexible yet structured environment for discussion, allowing for in-depth exploration of participants' experiences and perspectives while ensuring that all relevant topics are addressed. This balance is crucial for delving into the interplay between technological advancements and cybersecurity vulnerabilities. Unstructured interviews were considered at first, however, insufficient knowledge about the subject in combination with a lack of experience and knowledge from the author of this study makes semi-structured interviews suitable. Additionally, the choice of semi-structured interviews is justified by the need to explore complex and evolving insights from experts. This method allows for a deeper understanding of the challenges and potential countermeasures, aligning with the study's exploratory nature. The interviewing part has been refined to include an explanation of the laddering technique, emphasizing its role in exploring the beliefs of cybersecurity experts.

Adopting the strategies outlined by Guion et al. (2011), the interview process is carefully planned to ensure the effective collection and analysis of qualitative data. This involves preparing an interview guide that fosters in-depth discussions, utilizing a semi-structured format to accommodate the dynamic flow of conversation, and practicing active listening to capture the full depth of interviewees' insights. Rowley (2012) further advises on ensuring interview questions are free from jargon, not leading, and clearly understood by interviewees, thereby improving the quality of the data collected. The methodology advocated by Guion et al. (2011) with its emphasis on thorough preparation and active listening, aligns with and is complemented by the structured framework for

developing a semi-structured interview guide proposed by Kallio et al. (2016). The five-phase approach, proposed by Kallio et al. (2016) and used in this study includes:

1. Identifying prerequisites for using semi-structured interviews, ensuring this method aligns with our research objectives
2. Retrieving and utilizing previous knowledge to inform the development of our interview guide
3. Formulating the preliminary interview guide based on this knowledge
4. Pilot testing the guide with a small sample to refine questions and approach
5. Presenting the finalized interview guide, thereby enhancing the structure and depth of the interviews

This integration of methodologies underlines the approach's thoroughness, combining the strategies from Guion et al. (2011) with the process of creating and refining the interview guide by Kallio et al. (2016). In line with insights from Walle (2015), particular attention will be paid to the design of the interview process to minimize coverage, sampling, and nonresponse errors to enhance the credibility and reliability of the gathered data.

Further, by following the seven stages of conducting in-depth interviews as identified by Guion et al. (2011) —thematizing, designing, interviewing, transcribing, analyzing, verifying, and reporting— the aim is to achieve an understanding of the cybersecurity threats posed by deepfakes, the effectiveness of current countermeasures, and how organizations and individuals can prepare for the future. This structured yet flexible approach allows this study to probe deeply into the subject matter, ensuring that the study captures a broad range of perspectives and contributes meaningfully to the field.

To further ensure the robustness and rigor of the qualitative data collection process, principles adapted from Wohlin et al. (2012) will be employed, particularly in the systematic planning and execution of interviews. This involves developing a clear, structured approach to selecting interview subjects, designing interview questions, and conducting interviews in a manner that maximizes the reliability and validity of the information obtained. Following the suggestions from Wohlin et al. (2012), while the interviews are semi-structured, thorough attention will be paid to the sequence of questions and the adaptability of the interview protocol to ensure it captures the understanding of deepfake technology's impact on cybersecurity. The integration of Kallio et al. (2016) five-phase approach enhances this process by ensuring the interview guide is not only thoroughly developed but also tested and documented, facilitating transparency and replicability in this research.

3.1. Semi-Structured Interviews

The interview guide, as discussed in previous sections, is structured around key themes such as the technical evolution of deepfakes, cybersecurity threats they pose, mitigation strategies, and ethical dilemmas. These key themes are based on previous research and will be discussed further, later in this study. The questions are designed to prompt expansive responses, allowing experts to share their insights and experiences more freely. The goal of using semi-structured interviews in this study was to systematically examine the perspectives of deepfake experts while still allowing for the exploration of unforeseen topics. This

approach enabled the study to address specific research questions - such as the technical challenges in detecting deepfakes, the effectiveness of current countermeasures, and the ethical dilemmas posed by this technology - within a flexible conversational framework. The semi-structured format ensured that all interviews covered these essential topics, while also permitting experts to elaborate on areas based on their unique experiences and knowledge (Walle, 2015). This method was chosen to balance the need for comprehensive, comparable data with the opportunity to gain detailed insights into the evolving field of deepfake technology. While the interview guide serves as a foundational tool for ensuring comprehensive coverage of all relevant topics, interviewers maintain the flexibility to delve deeper into specific areas as conversations unfold, allowing for the exploration of emergent themes and insights.

Ethical considerations are paramount in conducting research interviews. (Rowley (2012) emphasizes the importance of informed consent, ensuring participants are fully aware of the interview's nature, its purpose, and the use of its data. Anonymity and confidentiality must be upheld to protect participants' privacy, requiring careful handling and storage of interview data. This study adheres to these principles, ensuring ethical integrity and respect for participants' rights and well-being throughout the research process.

3.2. Data Analysis

Following the guidance of Rowley (2012) on qualitative data analysis, this study will employ thematic analysis to interpret interview data. This involves coding the data to identify key themes and patterns related to deepfake technology's cybersecurity impacts. The analysis will be iterative, moving back and forth between the dataset and the emerging analysis to refine themes and understandings. This approach allows for an understanding of the complex issues surrounding deepfakes and cybersecurity, grounded in the perspectives of those directly engaged in the field.

As mentioned above, thematic analysis will be conducted on the transcriptions of the audio and text-based interviews. This method involves identifying, analyzing, and reporting patterns (themes) within the data. It starts with a close reading of the data to generate initial codes to see if sub-themes arise within the semi-predefined themes that arise when creating the interview guide, as discussed in previous sections, later in Chapter 4, and then shown in full, in Appendix A – Interview Guide. These codes will be grouped into potential themes, which will then be reviewed and refined to ensure they accurately represent the data. This approach allows for the structured and detailed exploration of the dataset, providing an understanding of the participants' perspectives and experiences.

3.3. Validity

To enhance the validity and credibility of this study, the methodology incorporates several key practices based on the recommendations of Rowley (2012), as presented in 3.3.1 through 3.3.4.

3.3.1. Design of Non-Leading, Open-Ended Questions

The study captures the genuine expertise and experiences of participants by crafting the interview guide with open-ended questions. These questions are structured to facilitate expansive responses without directing participants toward predetermined answers. The focus is on probing the evolution of deepfake technology, its cybersecurity threats, and potential countermeasures, enabling

experts to share insights grounded in their professional experiences and knowledge.

3.3.2. Pilot Testing of the Interview Guide

Before the principal data collection phase, the interview guide is subjected to pilot testing. This entails conducting initial interviews with a selected subset of participants to evaluate the clarity and interpretability of the questions. Feedback from the pilot testing is employed to refine the questions, ensuring they effectively elicit insightful responses and accurately align with the research objectives. This step is critical for validating the interview protocol and enhancing the data collection process's reliability. This study's pilot test consisted of one peer student, which proved helpful in clarifying the questions, making them harder to misunderstand.

3.3.3. Employment of Active, Reflective Interview Techniques

This study adopts an approach of active listening and adaptability during interviews, for delving deeper into related topics and clarifying uncertainties. This methodology facilitates a dynamic flow of conversation, allowing for the exploration of emergent areas of interest. Moreover, it entails the interviewer's reflection on personal biases and assumptions in real-time, thus maintaining an open and unbiased dialogue.

3.3.4. Engagement in Reflective Practice Across the Research Process

Beyond the interviewing phase, reflective practice is integrated throughout the research process, encompassing data analysis. By continuously engaging in a reflective examination of the research methodology, inherent biases, and analytical decisions, this study aspires to ensure that the findings authentically represent the collected data. Engaging in reflective practice bolsters the study's credibility, necessitating a thorough examination and justification of each step in the research process, from data collection to the presentation of findings.

3.3.5. Internal Validity

To enhance internal validity, strategies recommended by Wohlin et al. (2012) for ensuring the consistency and reliability of qualitative research will be implemented. This includes thorough documentation of the research process, from the selection of interviewees to the conduct of interviews, ensuring that the study's findings are firmly grounded in the data collected.

3.3.6. External Validity

Wohlin et al. (2012) emphasize the importance of generalizability in research findings. Although challenging in qualitative research, strategies to enhance external validity will include selecting a diverse range of interviewees from different sectors within the cybersecurity field. This diversity ensures that the insights and conclusions drawn from the study have broader applicability and relevance to the field of cybersecurity at large.

3.3.7. Validity of Conclusions

In line with the guidance of Wohlin et al. (2012), the validity of conclusions in this qualitative study will be bolstered through triangulation. Multiple data sources, including interviews with both industry experts and academic researchers, ensure that conclusions are not only based on a single source or perspective but are corroborated by multiple pieces of evidence, enhancing the study's overall validity.

4. Implementation

The interview guide is based on the background of this study, as discussed in Chapter 2. The background is, as mentioned, grounded in the literature of scientific studies regarding deepfake technology and its implications and so, the interview guide tries to capture those problems with these interview questions. The complete interview guide can be found in Appendix A – Interview Guide, however, chosen categories and example questions are showcased below.

Technical advancements and their implications – Referencing the studies of Frolov et al. (2022), Zhu et al. (2020), and Tolosana et al. (2020), this section will investigate how recent technical advancements contribute to the evolving landscape of deepfakes technology. Two examples of questions regarding this are "What are the key technical advancements that have made deepfakes more convincing and difficult to detect?" and "In your opinion, how has deepfake technology evolved over the past few years?"

Threats to cybersecurity and integrity from deepfakes – Inspired by Tolosana et al. (2020) and Chen et al. (2021), discussions will focus on identifying the many cybersecurity threats posed by deepfakes. For this category, two examples are "From your experience, what are the most concerning cybersecurity threats posed by deepfakes?" and "How do you foresee the evolution of deepfake technology impacting cybersecurity in the next five years?"

Tools for detecting deepfakes – Drawing on Kaushal et al. (2022), this part will delve into current and emerging technologies designed to detect and counteract deepfakes with questions such as "What countermeasures are currently employed to mitigate the risks associated with deepfakes?"

Ethical dilemmas – Guided by Pawlicka et al. (2023), this part will explore the ethical considerations surrounding the mitigation of deepfakes where an example of a question is "What ethical considerations should guide the development of deepfake detection and mitigation strategies?"

Future of deepfake technology evolution – With insights from Rana et al. (2022, 2024), the guide will explore anticipated developments in deepfake technology and its potential impact on cybersecurity. Example questions for this are "In your opinion, what is the most concerning aspect of deepfake technology?" and "What steps should organizations and individuals take now to prepare for future challenges posed by deepfakes?"

4.1. Laddering Technique

To explore intricate subjects, such as the cybersecurity implications of deepfake technology, this study employs the laddering interview technique to uncover the values that influence experts' perceptions and decision-making processes. As explained by Trocchia et al. (2007), laddering is a qualitative interviewing method that investigates deeper into the respondent's reasons behind specific choices or beliefs, ultimately linking these choices to underlying values. To adapt the laddering technique for interviewing cybersecurity experts, the following structured approach will be practiced:

1. **Initial Probing:** Begin with open-ended questions about the expert's experiences and views on deepfake technology and its challenges.
2. **Sequential Depth:** Utilize a series of 'why' questions to delve deeper into the reasons behind their views, aiming to connect specific observations or strategies to broader consequences and, ultimately, to

underlying values. A possible question could be "Why do you believe that specific countermeasure is more effective than others?" This aims to link their observations or choices to broader implications and core values.

3. **Identification of Values:** Drawing upon Rokeach's Values Inventory, as used by Trocchia et al. (2007), the interview process will guide experts to articulate values that underpin their perspectives on cybersecurity and ethical considerations in technology.

4.1.1. Ethical Considerations

Before conducting interviews, it will be ensured that the ethical standards are upheld by obtaining informed consent from all participants, guaranteeing confidentiality, and treating all responses with the utmost respect and dignity. This aligns with the ethical considerations emphasized by Trocchia et al. (2007) in their study.

4.1.2. Analyzing Laddering Interviews

Analysis of the laddering interviews will involve identifying the means (specific observations or strategies), consequences (impact on cybersecurity), and ends (underlying values) chain for each expert. This analysis will not only provide insights into the cognitive structures of experts regarding deepfakes but also reveal the values that drive their professional judgments and decisions.

4.1.3. Implementing Laddering Interviews

This method ensures that this study not only captures the technical and strategical insights from cybersecurity experts but also sheds light on the values driving their perspectives and actions. Inspired by Trocchia et al. (2007), the approach enriches the research methodology by providing a deeper understanding of the interplay between technology, cybersecurity, and human values. The laddering technique, if used correctly, can be utilized to get deeper answers to the research question of this study. Continued, further laddering technique tips on how to apply 'why' questions were developed from reading the work of Brandi Sørensen & Askegaard (2007). The remainder of this subsection will showcase the laddering technique, where the core aim is to explore the research question further.

Initial question: "In your opinion, how has deepfake technology evolved over the past few years?"

Why question: "Why do you believe these particular changes or advancements in deepfake technology are significant?"

Initial question: "From your experience, what are the most concerning cybersecurity threats posed by deepfakes?"

Why question: "Why do you consider these threats more concerning than others?"

Initial question: "What countermeasures are currently employed to mitigate the risks associated with deepfakes?"

Why question: "Why do you think these strategies have been adopted, and what makes them effective or ineffective?"

Initial question: "How do you foresee the evolution of deepfake technology impacting cybersecurity in the next five years?"

Why question: "Why do you anticipate these specific impacts, and what underlying factors are driving these changes?"

Initial question: "What ethical considerations should guide the development of deepfake detection and mitigation strategies?"

Why question: "Why are these particular ethical considerations important, and how do they influence the development of countermeasures?"

4.2. Participant Selection

For this study, the selection of interview participants was planned to ensure that the insights and data gathered were both relevant and authoritative. Participants were chosen based on their alignment with several specific criteria that signify their expertise and involvement in the field related to the research question.

The initial phase of participant recruitment involved reaching out to relevant authorities in Sweden to seek their involvement in the study. However, this approach yielded only one positive response. Consequently, the scope of recruitment was expanded to include professionals on the social media platform LinkedIn, where a broader array of potential participants could be targeted. On LinkedIn, participants were chosen based on their alignment with several specific criteria that signify their expertise and involvement in fields appropriate to the research on deepfake technology and its cybersecurity implications, presented below.

Relevant industry or field: Participants were selected from industries and fields directly related to the focus areas of this study. This includes professionals actively working in cybersecurity, AI, digital content management, and the ethics of cybersecurity technologies. Their direct involvement in these areas ensures they bring practical and current insights into the evolving landscape of deepfake technology.

Job title and role: The study targeted individuals holding preferred roles within their organizations that provide them with expert knowledge and responsibilities relevant to the study's core questions. Examples of such titles include cybersecurity analyst, AI researcher, digital content strategist, and other positions that deal directly with issues of digital authenticity and security.

Experience level: A preference was given to recruiting participants who were at senior or mid-career levels, in this study, meaning at least five years of experience in the relevant field. Their experience not only enriches the quality of the insights but also ensures a deeper understanding of both the technological and strategic dimensions of combating deepfakes.

Skills and endorsements: Skills appropriate to the study, such as expertise in AI, ML, data protection, and deepfake were considered essential. Participants were expected to have demonstrated competencies and endorsements in these areas, confirming their capability to provide detailed and technically accurate information.

Publications and contributions: Potential participants were also vetted for their contributions to the body of knowledge surrounding deepfake technology and related fields. Those who have published relevant articles, papers, or blog posts, or who have participated in speaking engagements at conferences, were particularly sought after. These contributions indicate an active engagement with the subject matter and a deeper-than-average understanding of the issues at stake.

Recruitment strategy: Messages asking about participation in the study were sent out to potential interviewees who met the above criteria. To aid in

identifying suitable candidates, a key search for “deepfake AI scams” was conducted, which helped in pinpointing individuals who are not only knowledgeable but also currently engaged with the specific challenges posed by deepfake technologies.

4.3. Dataset

This section outlines the composition and scope of the primary data collected for this study. The dataset contains qualitative data derived from semi-structured interviews conducted both in audio and text formats with professionals who met the specific criteria relevant to cybersecurity and deepfake technology, as discussed in the previous section. The core of the dataset consists of transcriptions of semi-structured interviews conducted with participants from various roles within the fields of AI, cybersecurity, and digital content management. These interviews are designed to gather in-depth insights into the participants’ experiences, perceptions, and expert opinions on the challenges posed by deepfakes and the strategies employed to mitigate these threats. To provide context for the insights gathered, it is important to note the diverse backgrounds of the interviewees:

Industry involvement: Participants include professionals from cybersecurity firms, AI development companies, digital content firms, and a representative of a Swedish authority on private data protection. This variety ensures a direct connection to the latest developments and challenges in the field.

Expertise and roles: The dataset includes data from a variety of expert roles such as cybersecurity analysts, AI researchers, digital content strategists, and governance, risk, & compliance advisors. Each participant brings a unique perspective on the technical and strategic dimensions of deepfake challenges, enriched by their firsthand experience in the industry.

Contributions to the field: Many participants have actively contributed to the discourse on deepfakes and cybersecurity through publications, conference presentations, and various projects. Their veteran insights enrich the data further, providing a well-rounded view of the current landscape and future directions.

A total of six interviews were fully completed, partial interviews were deleted due to not providing a holistic view from the interviewee, and partial interviews were text-based interviews where the interviewee stopped responding. Each interview was conducted either as an audio recording or via text-based platforms, depending on the preference, availability, and timezone of the participant. Audio interviews were recorded with consent and transcribed accurately. Text-based interviews were conducted using the chat available on LinkedIn after asking participants for their preferred chat platform. The chat allowed for a detailed and thoughtful exchange of ideas. The transcriptions and text outputs serve as the primary data sources for thematic analysis, allowing for an examination of the nuances in participants’ responses.

The qualitative data from the interviews will be analyzed and presented using thematic analysis techniques in Chapter 5. This involved coding the data into themes that emerged gradually from the discussions. These themes helped identify patterns and derive insights related to the research question. To reiterate the ethical handling of the data, this study emphasizes confidentiality, the anonymity of the participants, and the storage of sensitive information. Participants were

informed about the use of the data collected, which will be transcribed and saved until the end of this study, and then deleted.

4.4. Implementation of Thematic Analysis

Upon conducting interviews using the laddering technique, it is anticipated to uncover a lot of values and ethical considerations that inform experts' approaches to cybersecurity threats posed by deepfakes. A thematic analysis will be applied to this data, with a special focus on how organizations and individuals can prepare for the future that might come with the evolution of deepfake technology. To properly conduct thematic analysis, a six-phase process will be applied, following Braun & Clarke (2006):

1. **Familiarization with the data:** Reading through the transcriptions multiple times to gain a deep understanding of the content.
2. **Generating initial codes:** Systematically coding the data to identify significant statements related to cybersecurity threats posed by deepfakes.
3. **Searching for themes:** Organizing codes into potential themes and gathering all data relevant to each potential theme.
4. **Reviewing themes:** Checking if the themes work with the coded extracts and the entire dataset.
5. **Defining and naming themes:** Ongoing analysis to refine the specifics of each theme and the overall story the analysis tells.
6. **Producing the report:** Relating the analysis to the research questions and literature, providing clear examples of how the data supports the identified themes.

5. Results

This chapter presents the findings from the semi-structured interviews conducted with experts in cybersecurity, AI, and digital content management. These findings stem directly from the thematic analysis of the interviews, as outlined in the methodology section. The analysis has been structured to reflect the complexity and depth of the responses, providing insights into the dual-use nature of deepfake technology, its associated threats, and the ongoing efforts to develop effective countermeasures.

To maintain the confidentiality and anonymity of the participants, each has been assigned a pseudonym from the Greek alphabet, such as Alpha, Beta, and so on. The assignment of these pseudonyms follows no particular order but is aligned with the sequence of thematic analysis conducted by the author. This method ensures that the focus remains on the content of the contributions rather than on the identities of the contributors, adhering to ethical standards of research. Table 1 below, showcases each participant with their professional title, as references to show who supports each statement.

Pseudonym	**Title**
Alpha	AI Specialist
Beta	IT Security and AI Researcher
Gamma	Cybersecurity Specialist
Delta	AI Speaker
Epsilon	Executive Cybersecurity Advisor
Zeta	Cybersecurity Analyst

Table 1. Interview Participants (Author's own)

The results are organized into six key themes that emerged during the analysis. Each theme is treated as a sub-chapter and discussed in detail, reflecting the perspectives of the experts involved. To bring these themes to life and illustrate the real-world implications of the findings, selected quotes from the interviews are included. These quotes are chosen to highlight the consensus or diversity of opinion among the experts and to provide a direct insight into their reasoning and experiences.

5.1. Evolution and Accessibility of Deepfake Technology

This section examines how the accessibility and sophistication of deepfake technology have evolved, presenting both challenges and opportunities. Participants shared their observations on how these changes impact cybersecurity strategies and preparedness. All of the participants noted a significant improvement in the accessibility and quality of deepfake technology. Further, emphasizing its dual-use nature—while it offers innovative opportunities in media and entertainment, it also poses severe threats to misinformation and identity fraud. Alpha highlighted the dramatic shift in accessibility, stating: *"From being inaccessible, complicated, and not very good, to being accessible to everyone in a very simple way."* Beta mirrored this sentiment, reflecting on the technology's progression: *"from something quite inaccessible and rudimentary to something very advanced and accessible to almost anyone without special skills."* Zeta added to this perspective by noting the significant leap in technology post-2022, stating: *"The leap that AI has taken has led to this tech becoming more accessible to*

hackers," emphasizing the dual-edged nature of technological democratization with the public release of ChatGPT. Zeta continues on the matter, observing its increasing misuse alongside growing public awareness:

> *"[...] and like any other tech that gets cheaper and more accessible as time goes by, bad state actors, hackers, etc. have managed to get their hands on the tech that creates deepfakes. It is scary now... but, at the same time, the common man today is much more clued into such scams as compared to prior Nov 2022."*

Gamma reinforced these views by mentioning: *"Deepfake technology has evolved significantly over the past few years, both in terms of its accessibility and its sophistication,"* highlighting the ease with which these tools can now be used by the general public. These perspectives underscore the transformation of deepfake technology from a complicated expertise to a widely accessible tool, amplifying its potential for misuse in various contexts.

5.2. Cybersecurity Threats

Focusing on the direct threats posed by deepfakes, this theme explores the various ways in which deepfakes can undermine security protocols, manipulate information, and perpetrate fraud. This discussion ties directly to the research question by highlighting the need for organizations to stay informed about technological advancements to anticipate and mitigate associated risks.

Participants expressed concerns over deepfakes' potential to compromise personal and organizational security. The insights gathered here respond to the research question by identifying specific threats. Experts shared expressed concerns over how deepfakes can be used to perpetrate identity theft, financial fraud, and misinformation campaigns. One notable point was the difficulty in detecting these manipulations, which are becoming increasingly sophisticated. Gamma emphasized the manipulation capabilities of deepfakes, noting they can: *"impersonate individuals in videos or audio recordings to manipulate victims into revealing sensitive information."* Epsilon highlighted a specific corporate risk, pointing out that: *"deepfakes are used to impersonate target employees for a variety of motives, including espionage, data theft, financial benefits, and credential breaches."* Adding to the reflection on the practical consequences, Zeta emphasized its impact on personal and professional security as well:

> *"Deepfakes can be used to hoodwink you into handing over pretty much all your personal and professional data. It can be misused in finance; there are many cases where directors of companies and CEOs have been impersonated and company executives asked to transfer money."*

Zeta contributed further by detailing a real-world scenario involving deepfake phone calls, illustrating the direct and immediate threat of voice cloning used by scammers to extort money or steal personal data: *"I recently wrote an advisory on deepfake phone calls where voices of relatives, friends, and colleagues are cloned within a matter of minutes by these scamsters to extort money or steal personal data."* Beta also expressed concerns about the broader impacts of deepfakes, explaining: *"The real danger I foresee isn't the direct threats themselves, which are manageable, but rather how manipulated information gets*

'laundered' through reputable news channels, gaining an air of legitimacy that can have profound societal impacts." These insights collectively illustrate the diverse threats posed by deepfakes to personal and organizational security.

5.3. Countermeasures and Detection

In response to threats, this section delves into the current and developing technologies aimed at detecting and countering deepfakes. This theme is important for answering the research question as it discusses practical tools and strategies that organizations can employ to enhance their preparedness against deepfake-related security threats, such as those mentioned in the previous theme.

The discussions consistently pointed to the need for robust detection technologies and strategic countermeasures to combat deepfake threats. Gamma mentioned: *"We are currently developing more sophisticated ML models that can identify deepfake anomalies better than the previous generation."* Delta advocated for a comprehensive approach, suggesting: *"Never trust one single sense… but the combination of all the senses, we don't have a deep fake cologne yet."* Zeta highlighted the importance of vigilance, stating: *"The most important thing anybody can do is NOT to take any image, video at face value. Those days are gone… treat almost everything visual with suspicion."* This approach is echoed by Alpha and Beta, who both see the necessity for significant investment in AI and ML capabilities to develop detection technologies robust enough to counter deepfakes. Gamma emphasized the development of algorithms capable of detecting video frame anomalies invisible to human eyes, as a: *"critical advancement in the technological battle against deepfakes."* Alpha stated: *"Organizations need to invest in AI and machine learning capabilities to develop detection technologies robust enough to identify deepfakes,"* highlighting the ongoing arms race in technology development. Alpha, Beta, Gamma, Delta, Epsilon, and Zeta all agreed that organizations and authorities need to invest in research and development of advanced detection technologies that can and need to keep pace with the evolution of deepfake technologies, a race that according to Beta: *"we might not be able to win."* Epsilon further stressed the importance of public awareness and education, advocating for initiatives that inform the public about the nature and risks of deepfakes, which are essential for their early detection and mitigation.

5.4. Ethical and Regulatory Considerations

This theme addresses the ethical dilemmas and regulatory challenges that arise with the advancement of deepfake technology. The exploration of this theme offers insights into how ethical considerations and regulations can guide the development of countermeasures, directly tying back to the research question by suggesting frameworks for organizational and individual preparedness.

The lag between technological advancements and the development of corresponding legal frameworks was a recurrent theme. Beta expressed a common concern, saying: *"We are quite concerned about the regulatory aspects because the technology is moving faster than the law"* and continues, *"There's a critical need for international standards to manage this technology effectively."* This sentiment was supported by Epsilon and Gamma, who emphasized the importance of developing ethical guidelines and privacy considerations in the deployment of countermeasures with Gamma adding: *"Ethical considerations are crucial in guiding the development of deepfake detection and mitigation strategies to ensure that these technologies are used responsibly and do not*

inadvertently harm individuals or society.“ Beta also highlighted the responsibility of developers, calling for a code of ethics to guide the development and use of deepfake technology responsibly. Additionally, Alpha called for stronger regulations, *“There needs to be legislation that specifically addresses the creation and distribution of deepfakes, akin to laws against digital fraud.”* Zeta further highlighted the importance of ethics and awareness:

> *“Every individual and employee must be given at least a basic course in AI ethics. Also, they need to be sensitized to the harmful fallout of deepfakes in society, including the fact that reputation, as well as money, is at stake here.”*

The narrative here is focused on the need for comprehensive policies and ethical frameworks to effectively manage the use and mitigation of deepfakes.

5.5. Societal Implications

This section considers the broader societal implications of deepfakes, including their impact on public trust and democratic processes. Insights from participants reveal the potential for deepfakes to distort public discourse and influence political landscapes. Addressing the research question, this theme underscores the importance of societal awareness and educational initiatives as part of a comprehensive strategy to prepare for and mitigate the risks posed by deepfakes.

The societal implications of deepfakes were discussed in terms of both direct and indirect threats. Alpha brought attention to a societal risk, stating: *“There is a bigger threat than cybersecurity, which is that we can’t determine what is true or false... we will choose to believe whatever we believe out of our comfort zone.”* Beta highlighted the subtle ways deepfakes can manipulate public opinion: *“[...] shaping the opinions or behavior of groups through subtly skewed information.”* Gamma further adds to this discussion by saying:

> *“The most concerning aspect of deepfake technology, in my opinion, lies in its potential to undermine the foundation of truth in communication and media. This loss of trust is particularly alarming because it affects multiple dimensions of society, including politics, law enforcement, personal security, and journalism.”*

These perspectives provide an understanding of how deepfakes could impact societal structures and public discourse.

5.6. Future Outlook and Preparedness

This theme explores the anticipated developments in deepfake technology and assesses the readiness of organizations to face future challenges. It delves into the strategic planning and innovations required to keep pace with technological advancements. Insights from participants emphasize the arms race between deepfake creation and detection technologies. These discussions are directly tied to the research question by emphasizing the need for proactive strategies that anticipate future trends in deepfake technology.

While Delta provided an optimistic view of the potential benefits of deepfakes if used ethically, suggesting they could: *“[...] give everybody a half a day back... as a stepping stone for our next maybe even evolutionary step,”* However,

experts also shared more realistic thoughts on the evolution, with Delta saying: *"[...] this technology is becoming scary. And in the next one to three years, it's gonna get really scary as in live deep fakes."* Directly supported by Alpha: *"I believe that in a year, the technology will be so good and fast that you can just prompt Netflix to create what kind of movie you want to watch."* Alpha and Delta further, mentioned their belief in decentralized blockchain solutions as future technology to detect deepfake. Gamma and Epsilon focused on the challenges, with Gamma forecasting: *"As deepfake technology becomes more sophisticated and accessible, we can expect an increase in the frequency and quality of deepfake incidents."* All interviewees agreed that deepfake technology will continue to evolve. Gamma projected further, *"Deepfake technology will only get more sophisticated, and it is a race against time for detection technologies to keep up."* Delta emphasized the necessity for proactive measures, *"We can't just react to threats as they come; we need adaptive strategies that evolve as quickly as the technologies do."* Alpha provided a concerning thought, mentioning:

> *"It is certainly a challenge, and we have had challenges before that we, as a humanity, have been able to figure out and solve, but I mean, this is a crisis.. Where I don't think we can keep up, it's going too fast."*

A common opinion among experts was that training and awareness are crucial. Gamma and Delta summarize it well with Gamma saying:

> *"Investing in education and awareness is vital. The more the public knows about deepfakes, how they are made, and their potential impact, the less likely they are to be fooled by them. Educational initiatives should not only focus on helping people identify deepfakes but also on understanding the implications of AI and digital media manipulation on society."*

Delta adding to public knowledge with:

> *"[...] it's all about awareness right now, because you get it, you need to educate everybody in a way that they know, oh, can I trust my eyes now? So that they become aware of what they get fed, what they see is not automatically the truth, or automatically something they should believe. And it's my hope that it will make us more human, let's go have a cup of coffee."*

These two previous quotes were supported by Alpha, Beta, Epsilon, and Zeta. Further, these views enrich the dialogue about the future of deepfake technology, from its threats to its potential applications. In preparing for the challenges posed by deepfakes, organizations and individuals are urged to adopt proactive and informed strategies. This includes the development of advanced detection systems, accurate verification protocols, and comprehensive educational programs. The collective capacity to mitigate the risks associated with deepfakes can be enhanced by fostering collaboration among various stakeholders and advocating for legal frameworks that keep pace with technological advances.

6. Discussion

In discussing the findings, it is explored how the values and ethical considerations uncovered through the laddering technique inform current and future cybersecurity strategies against deepfakes. This will include a reflection on the effectiveness of the laddering method in revealing these insights, contributing to the broader dialogue on qualitative methods in cybersecurity research.

6.1. Results Concerning Previous Research

The results of this study can be compared with the findings of previous research to assess how they confirm, challenge, or expand the current understanding of deepfake technology's impact on cybersecurity. The discussion in this section further extends, by providing fresh, qualitative data from the field, emphasizing the speed at which these changes are occurring and the practical implications for cybersecurity. While the results from my study might not come as a shock or feel innovative, it expands on several earlier studies where, while not only extending previous research, it further adds reliability with scientific support from qualified experts, that work in close relation to deepfake technology.

My study's findings about the evolution and accessibility of deepfake technology reflect the concerns highlighted by Tolosana et al. (2020), who noted that the technology is rapidly evolving, with increasing realism and accessibility. This got further amplified when two independent experts foretold that, since deepfake technology evolves at such speed, we could have on-demand and instant deepfake creation from just a prompt, in just one year.

Rana et al. (2024) mention the promising use of blockchain technology in detecting deepfakes, which experts highlighted as a likely solution in this study too, further backing up that implication.

The expert's emphasis on collaboration to keep pace with technological advancements aligns with the insights of Di Dario et al. (2023) where they highlight the importance of varying methods to match the pace.

In the interviews conducted in this study, there was an emphasis on public education as a necessity. This further enhances the highlighting of the study by Goh et al. (2022).

The study by Pawlicka et al. (2023) highlighted the complexities concerning ethical dilemmas regarding deepfake technology, something that the experts in my study also emphasised, with pleas for some kind of regulations, to hinder the creation of malicious deepfakes.

6.2. Methods, Implementation, and Results

The methodological approach of this study was primarily based on semi-structured interviews with the laddering technique. This combination was chosen to delve into the perspectives of cybersecurity experts on the evolving threats posed by deepfake technology. Semi-structured interviews allowed for flexibility in the discussion, enabling interviewees to express their thoughts extensively while still covering the predetermined set of topics crucial to the research questions. This format facilitated an in-depth exploration of complex issues, allowing the interviewer to probe further based on the responses received, which is essential in capturing the subtleties of expert opinions on as intricate a subject as deepfake technology.

The decision to incorporate text-based interviews alongside traditional video and audio formats played a crucial role in ensuring sufficient participation. Given the geographical challenges and potential participants 'ghosting' even after initial positive engagements, text-based interviews provided a flexible alternative for many experts who might have faced scheduling conflicts or preferred a non-verbal communication format. This method was important in achieving the necessary number of interviews to support the analysis.

However, while text-based interviews increased participation, they also introduced specific challenges, particularly concerning the laddering technique. The asynchronous nature of text communication meant that responses were likely more considered and less spontaneous than those in real-time conversations. Participants had more time to reflect on and craft their answers, which could lead to responses that were more safe or polished but potentially less candid. This delay might have reduced the immediacy and spontaneity typically advantageous in laddering interviews, where immediate reactions can often reveal underlying values and motivations more effectively. Further, the text-based interviews introduced a privacy concern that only in hindsight, was highlighted. The fact that the interviews in text format were conducted in LinkedIn's chat made it impossible to maintain the certainty of anonymity due to not holding any power over LinkedIn's handling of chat data. Despite these limitations, text-based interviews were invaluable for broadening the study's participant base and ensuring a diverse range of inputs. For future studies, blending both interview formats while optimizing the use of laddering in live interviews could enhance data richness and authenticity, ensuring a balance between depth of insight and breadth of participation. The text-based interviews should also be held in chat applications that adhere to anonymity, to make sure anonymity can be upheld.

The laddering technique, employed to gain insights into the values and beliefs underpinning experts' opinions, proved valuable. It helped reveal deeper reasoning behind their strategic choices and perceptions of deepfake threats. By encouraging interviewees to elaborate on their initial responses, this technique provided a clearer picture of the reasoning and ethical outlines guiding their professional judgments and decisions. This methodological approach was instrumental in uncovering not just what the experts believe but why they hold these beliefs, offering a richer and more detailed dataset for analysis.

However, the interview pilot test only consisting of one peer student should be bigger in coming studies, it was a helpful pilot nonetheless. Further, the study faced challenges in terms of participant engagement as earlier mentioned, which could have influenced the comprehensiveness of the findings. Despite initial positive responses, a series of unfortunate events led to some potential interviewees not following through with their commitments. This issue resulted in fewer completed interviews than planned. The impact of this was two-fold: it limited the diversity of insights and potentially introduced a selection bias, as the views of those who did not participate might differ from those who did.

Reflecting on these experiences, the inclusion of more participants would likely have benefitted the study by broadening the range of insights and enhancing the conclusions. To mitigate similar issues in future research, alternative strategies such as over-scheduling participants, offering incentives for participation, or employing follow-up reminders closer to scheduled interview times might be considered. Furthermore, supplementing qualitative data with quantitative

methods, such as surveys, could provide additional layers of data for cross-validation and help overcome any biases that might arise from a limited sample size.

In conclusion, while the chosen methods were effective in exploring deep expert knowledge and producing rich, qualitative data, the challenges encountered highlight the need for participant engagement strategies and perhaps a mixed-methods approach in future studies to enhance data diversity and reliability.

6.3. Ethical and Societal Aspects

The ethical implications of deepfake technology are extreme and complicated, including issues of privacy, consent, and the integrity of information. Deepfakes pose challenges as they can be used to create convincing and unauthorized representations of individuals, thereby breaching the rights to privacy and leading to potential misuse in the form of misinformation and deception. The participants in this study frequently highlighted concerns about the misuse of deepfakes to manipulate public opinion, perpetrate fraud, and undermine trust in digital media.

Deepfakes directly challenge the ethical boundaries of consent and privacy. The ability to manipulate images and videos to such an extent that they can portray individuals saying or doing things they never did raises concerns about consent. None of the illustrated individuals have agreed to be represented in these contexts, which not only misrepresents their actions but can also have lasting impacts on their personal and professional lives. As such, the creation and distribution of deepfakes without consent violate basic ethical norms and privacy rights, warranting regulatory responses to deter misuse.

The capacity of deepfakes to spread misinformation is perhaps their most dangerous attribute. By creating false representations of events or statements, deepfakes can erode public trust in media and institutions. This can be particularly problematic in democratic societies where an informed population is important. Participants noted instances where deepfakes had been used to manipulate political discourse, which could potentially influence election outcomes or breed social unrest. The societal implications of such actions can be enormous, necessitating a combined effort to manage and mitigate these risks.

The findings of this study suggest that current legal and regulatory frameworks are inadequate to address the challenges posed by deepfakes. There is a clear need for updated policies that specifically address the creation, distribution, and use of deepfakes, with a focus on protecting privacy, preventing harm, and preserving integrity. Participants advocated for international cooperation to establish standards and regulations that could effectively manage the global nature of digital media and cyber threats. Proposed regulatory frameworks could include penalties for malicious use of deepfake technologies and mandatory use of watermarking technologies by creators of artificial media to ensure transparency and traceability. Furthermore, there could be a requirement for platforms that host user-generated content to implement more robust content verification processes to detect and mitigate the spread of deepfakes.

Beyond regulation, there is a need for ethical guidelines that govern the use of AI in the creation of digital content. These guidelines should emphasize respect for individual rights, consent, and the avoidance of harm. Educational initiatives that increase public awareness about the nature and risks of deepfakes are also critical. Such initiatives could help cultivate an understanding public that is

better equipped to question and critically evaluate the authenticity of digital content.

Reflecting on the ethical considerations raised by interview participants aligns with broader societal concerns about the integrity of information and personal security. Policymakers, legislators, technologists, and the public must collaborate to forge pathways that safeguard individual rights while employing the benefits of digital innovation.

6.4. Potential Harms and Ethical Considerations

Research into detecting and analyzing deepfakes, while beneficial, carries the risk of misuse. Improved techniques could wrongly assist malicious actors in creating more convincing deepfakes. To prevent such outcomes, this study advocates for responsible research propagation practices and collaboration with ethical organizations to ensure the findings are used to strengthen defenses against deepfake misuse rather than to facilitate their creation. By understanding the capabilities and limitations of deepfake technology, society can be better prepared to question and critically evaluate the authenticity of digital media.

Studying technologies associated with misinformation risks can contribute to a negative outlook on technological advancements. This study aims to balance this by highlighting the proactive measures being taken to detect and mitigate deepfakes effectively by discussing the importance of fostering a positive narrative around AI technologies, emphasizing their potential for societal benefit alongside the challenges they present. Further, focusing solely on the negative applications of AI, such as malicious deepfakes, risks slurring a field that has significant positive potential. This research strives to present a balanced view, acknowledging the beneficial uses of AI while addressing the ethical dilemmas posed by their misuse. By doing so, the aim is to foster a more informed and comprehensive understanding of AI technologies among the public and policymakers.

7. Conclusion

This final chapter summarizes the key findings of the study, highlights the implications for cybersecurity practices, acknowledges the limitations of the current research, and suggests directions for future work. This discussion culminates in offering strategic recommendations to combat deepfake-related security threats.

7.1. Implications

The implications of this study are important for both the field of cybersecurity and for public awareness. The research highlights the need for detection technologies and legal frameworks to mitigate the risks posed by deepfakes. As the technology evolves, the strategies to detect and counter deepfakes must also advance. Organizations and policymakers must consider adopting multi-layered security strategies that include both technological solutions and human oversight. These strategies must be adaptive and forward-looking, anticipating the developments in deepfake technology. Furthermore, this study underscores the importance of international cooperation in developing standards and regulations to manage the use and abuse of deepfakes effectively. Global guidelines can help harmonize efforts to prevent the misuse of AI technologies across borders.

7.2. Limitations

The study faced several limitations that could impact the generalizability and applicability of its findings. The primary limitation was the relatively small and potentially non-representative sample of interviewees, which was further compounded by the challenges of 'ghosting' encountered during participant recruitment. Additionally, the reliance on semi-structured and text-based interviews, while rich in detailed insights, limits the generalizability of the results. The text-based format may have also introduced a bias toward more deliberate responses, potentially overlooking the spontaneous insights that might emerge in a face-to-face dialogue.

7.3. Future Work

Future research should aim to address the limitations noted and expand the scope of inquiry. Studies involving larger and more diverse groups of participants could provide a broader perspective on the challenges and solutions related to deepfake technology. Quantitative research could complement this qualitative study, with statistical data, to see the occurrence and impact of deepfake concerns and solutions across different demographics and sectors. Furthermore, future work could explore the development of more advanced AI-driven tools for deepfake detection and consider the ethical implications of deploying such technologies. Lastly, the need for educating the general public needs attention, research into the effectiveness of public awareness campaigns and educational programs in improving the public's ability to recognize and react to deepfakes would be valuable.

7.4. Final Thoughts

As we navigate this new era of digital manipulation, the importance of maintaining the integrity of information cannot be overstated. The fight against deepfakes is not only a technical challenge but a societal imperative to preserve trust in digital communication. The collaborative efforts of technologists, legislators, educators, and the public are key in developing strategies to address this evolving threat. This study contributes to the ongoing discourse on deepfakes by

highlighting the complex interplay between technology, policies, and ethics and by calling for a proactive approach to safeguard our digital future.

The conclusion of this study is a call to action, it emphasizes the need for continued innovation, collaboration, and vigilance in the face of one of the most deceptive technologies of our time.

References

Brandi Sørensen, E., & Askegaard, S. (2007). Laddering: How (not) to do things with words. *Qualitative Market Research: An International Journal*, *10*(1), 63–77. https://doi.org/10.1108/13522750710720404

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Chen, H.-S., Rouhsedaghat, M., Ghani, H., Hu, S., You, S., & Jay Kuo, C.-C. (2021). DefakeHop: A Light-Weight High-Performance Deepfake Detector. *2021 IEEE International Conference on Multimedia and Expo (ICME)*, 1–6. https://doi.org/10.1109/ICME51207.2021.9428361

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., & Bharath, A. A. (2018). Generative Adversarial Networks: An Overview. *IEEE Signal Processing Magazine*, *35*(1), 53–65. https://doi.org/10.1109/MSP.2017.2765202

Di Dario, D., Pontillo, V., Lambiase, S., Ferrucci, F., & Palomba, F. (2023). Security Testing in The Wild: An Interview Study. *2023 49th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, 191–198. https://doi.org/10.1109/SEAA60479.2023.00037

Frolov, Dmitriy. B., Makhaev, Dmitriy. D., & Shishkarev, V. V. (2022). Deepfakes and Information Security Issues. *2022 International Conference on Quality Management, Transport and Information Security, Information Technologies (IT&QM&IS)*, 147–150. https://doi.org/10.1109/ITQMIS56172.2022.9976507

Goh, D. H.-L., Lee, C. S., Chen, Z., Kuah, X. W., & Pang, Y. L. (2022). Understanding Users' Deepfake Video Verification Strategies. In C.

Stephanidis, M. Antona, S. Ntoa, & G. Salvendy (Eds.), *HCI International 2022 – Late Breaking Posters* (pp. 25–32). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-19682-9_4

Guion, L., Diehl, D., & McDonald, D. (2011). Conducting an In-depth Interview. *EDIS*, *2011*. https://doi.org/10.32473/edis-fy393-2011

Kallio, H., Pietilä, A.-M., Johnson, M., & Kangasniemi, M. (2016). Systematic methodological review: Developing a framework for a qualitative semi-structured interview guide. *Journal of Advanced Nursing*, *72*(12), 2954–2965. https://doi.org/10.1111/jan.13031

Kaushal, A., Singh, S., Negi, S., & Chhaukar, S. (2022). A Comparative Study on Deepfake Detection Algorithms. *2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, 854–860. https://doi.org/10.1109/ICAC3N56670.2022.10074593

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

Mirsky, Y., & Lee, W. (2021). The Creation and Detection of Deepfakes: A Survey. *ACM Computing Surveys*, *54*(1), 7:1-7:41. https://doi.org/10.1145/3425780

Pawlicka, A., Pawlicki, M., Kozik, R., & Choraś, M. (2023). What Will the Future of Cybersecurity Bring Us, and Will It Be Ethical? The Hunt for the Black Swans of Cybersecurity Ethics. *IEEE Access*, *11*, 58796–58807. https://doi.org/10.1109/ACCESS.2023.3283791

Rana, M. S., Nobi, M. N., Murali, B., & Sung, A. H. (2022). Deepfake Detection: A Systematic Literature Review. *IEEE Access*, *10*, 25494–25513. https://doi.org/10.1109/ACCESS.2022.3154404

Rana, M. S., Solaiman, M., Gudla, C., & Sohan, M. F. (2024). Deepfakes – Reality Under Threat? *2024 IEEE 14th Annual Computing and Communication Workshop and Conference (CCWC)*, 0721–0727. https://doi.org/10.1109/CCWC60891.2024.10427659

Remya Revi, K., Vidya, K. R., & Wilscy, M. (2021). Detection of Deepfake Images Created Using Generative Adversarial Networks: A Review. In M. Palesi, L. Trajkovic, J. Jayakumari, & J. Jose (Eds.), *Second International Conference on Networks and Advances in Computational Technologies* (pp. 25–35). Springer International Publishing. https://doi.org/10.1007/978-3-030-49500-8_3

Robila, V. (2023). Fostering Computer Science Education through Expert Interviews. *2023 IEEE Integrated STEM Education Conference (ISEC)*, 54–57. https://doi.org/10.1109/ISEC57711.2023.10402224

Rowley, J. (2012). Conducting research interviews. *Management Research Review*, *35*(3/4), 260–271. https://doi.org/10.1108/01409171211210154

Shoaib, M. R., Wang, Z., Ahvanooey, M. T., & Zhao, J. (2023). Deepfakes, Misinformation, and Disinformation in the Era of Frontier AI, Generative AI, and Large AI Models. *2023 International Conference on Computer and Applications (ICCA)*, 1–7. https://doi.org/10.1109/ICCA59364.2023.10401723

Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). Deepfakes and beyond: A Survey of face manipulation and fake

detection. *Information Fusion*, *64*, 131–148. https://doi.org/10.1016/j.inffus.2020.06.014

Trocchia, P. J., Swanson, D. L., & Orlitzky, M. (2007). Digging Deeper: The Laddering Interview, a Tool for Surfacing Values. *Journal of Management Education*, *31*(5), 713–729. https://doi.org/10.1177/1052562906293611

Vaccari, C., & Chadwick, A. (2020). Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. *Social Media + Society*, *6*(1), 2056305120903408. https://doi.org/10.1177/2056305120903408

Walle, A. H. (2015). *Qualitative Research in Business: A Practical Overview*. https://eds-p-ebscohost-com.library-proxy.his.se/eds/ebookviewer/ebook?sid=c23c12cf-b15a-4712-98f1-44f3ccbe6b75%40redis&ppid=pp_vii&vid=0&format=EB

Westerlund, M. (2019). The Emergence of Deepfake Technology: A Review. *Technology Innovation Management Review*, *9*(11), 40–53. https://doi.org/10.22215/timreview/1282

Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., & Wesslén, A. (2012). *Experimentation in Software Engineering*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-29044-2

Zhu, K., Wu, B., & Wang, B. (2020). Deepfake Detection with Clustering-based Embedding Regularization. *2020 IEEE Fifth International Conference on Data Science in Cyberspace (DSC)*, 257–264. https://doi.org/10.1109/DSC50466.2020.00046

Appendix A – Interview Guide

1. Introduction

• Brief the participant on the study's purpose, ensuring understanding and obtaining informed consent to record for transcribing.

• Assure confidentiality and anonymity to encourage openness.

2. Warm-up questions

• "Could you briefly describe your role and experience in cyber security?"

• "How did you first become involved in issues related to deepfake technology?"

3. Understanding deepfake technology

• "In your opinion, how has deepfake technology evolved over the past few years?"

• "What are the key technical advancements that have made deepfakes more convincing and difficult to detect?"

4. Cybersecurity threats posed by deepfakes

• "From your experience, what are the most concerning cyber security threats posed by deepfakes?"

• "Can you share any instances or examples where deepfakes have significantly impacted cybersecurity?"

5. Current countermeasures and their effectiveness

• "What countermeasures are currently employed to mitigate the risks associated with deepfakes?"

• "In your opinion, how effective are these strategies, and what are their most significant limitations?"

6. Insights on future developments

• "How do you foresee the evolution of deepfake technology impacting cybersecurity in the next five years?"

• "What emerging technologies or strategies hold the most promise for detecting and combating deepfakes?"

• "In your opinion, what is the most concerning aspect of deepfake technology?"

• "What steps should organizations and individuals take now to prepare for future challenges posed by deepfakes?"

7. Ethical and societal considerations

• "What ethical considerations should guide the development of deepfake detection and mitigation strategies?"

• "How can we balance the need for security with protecting individual privacy and freedom of expression?"

8. Concluding questions

• "Is there anything of importance that you would like to add? Anything you feel like we missed?"

• "Based on our discussion, what do you believe are the key areas for future research on deepfakes and cybersecurity?"

• "Are there other experts or resources you recommend I explore to gain further insights on this topic?"

• “Can you recommend any other experts or professionals in the field of cybersecurity or deepfake technology who might provide additional insights or perspectives? Would you be willing to facilitate an introduction?”

9. Thank you and my next steps

• Thank the participant for their time and contribution.

• Discuss the next steps, including how the information will be used and any follow-up actions.