




Article

Deepfake-Driven Social Engineering: Threats, Detection Techniques, and Defensive Strategies in Corporate Environments

Kristoffer Torngaard Pedersen, Lauritz Pepke, Tobias Stærmose, Maria Papaioannou , Gaurav Choudhary  and Nicola Dragoni * 

DTU Compute, Technical University of Denmark (DTU), 2800 Kongens Lyngby, Denmark;
s205354@dtu.dk (K.T.P.); s191179@dtu.dk (L.P.); s205356@dtu.dk (T.S.); marpapa@dtu.dk (M.P.);
gauch@dtu.dk (G.C.)

* Correspondence: ndra@dtu.dk

Abstract: The evolution of deepfake technology has the potential to reshape the threat landscape in corporate environments by enabling highly convincing digital impersonations. In this paper, we explore how artificial media produced by AI can be misused to assume authoritative personas, leaving traditional cybersecurity programs with significant vulnerabilities. Drawing from interviews with cybersecurity professionals across various industries, we find that the majority of organizations remain vulnerable due to their adoption of broad, vendor-centric security solutions that are not specifically designed to protect against deepfake attacks. In response to the evolving threat landscape, we introduce the PREDICT framework—a cyclical, iterative theoretical model. This model combines definitive policy direction, organizational preparedness, targeted employee training, and advanced AI detection tools. Additionally, it incorporates effective incident response plans with continuous improvement and simulations. Our findings underscore the need to revise the current security protocols and offer practical suggestions for strengthening corporate defenses against the increasingly dynamic threat landscape posed by deepfakes.

Keywords: deepfake; social engineering; detection techniques; corporate security; defensive strategies



Academic Editor: Danda B. Rawat

Received: 24 February 2025

Revised: 5 April 2025

Accepted: 24 April 2025

Published: 27 April 2025

Citation: Pedersen, K.T.; Pepke, L.; Stærmose, T.; Papaioannou, M.; Choudhary, G.; Dragoni, N. Deepfake-Driven Social Engineering: Threats, Detection Techniques, and Defensive Strategies in Corporate Environments. *J. Cybersecur. Priv.* **2025**, *5*, 18. <https://doi.org/10.3390/jcp5020018>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Deepfakes represent synthetic media generated through artificial intelligence and machine learning techniques [1]. In recent years, the technologies within AI and machine learning have grown into potent tools that can be used for malicious activities [2]. This could be impersonation or disinformation. Early discussions focused mainly on their potential misuse in politics or celebrity culture. Nowadays, however, there is growing awareness that corporate environments are also at high risk. Deepfake-driven social engineering poses a significant challenge by creating hyper-realistic forgeries such as audio or video impersonations of executives or stakeholders that can bypass traditional security measures [3].

A *critical gap* remains in both the academic literature and industry practice: there is insufficient attention devoted to deepfake-driven social engineering at scale within corporate environments [4]. Although several studies have addressed deepfake generation and detection in general [5,6], only a few have investigated how these synthetic media threats exploit structures and communication channels that are of great significance to modern enterprises. As illustrated in Table 1, this comparative overview reveals key gaps in the current cybersecurity practices that highlight the need for additional research to

improve organizational preparedness for deepfake-driven social engineering threats. The existing corporate defense protocols, largely oriented toward classic phishing or generic malware, often fail to anticipate the sophistication and persuasive realism of deepfake attacks [7].

To address these concerns, the aim of this paper is to bridge the research gap between theoretical deepfake capabilities and the limited, often reactive countermeasures employed by modern organizations. This work examines the effectiveness of the current detection methods and aligns them with practical defensive strategies. In this way, it extends the conversation beyond the abstract dangers of deepfakes with the goal of providing actionable insights for corporate stakeholders.

1.1. The Problem Statement

The rapid advancement of deepfake technology has introduced a new and potentially critical threat into the threat landscape within the corporate world. While traditional measures in cybersecurity are designed to address conventional threats, they still risk falling short in addressing the sophisticated and evolving nature of deepfake-driven social engineering attacks. Despite increasing media attention and academic interest, many organizations remain unprepared due to a lack of specialized detection tools and an over-reliance on generalized security practices.

This research attempts to bridge this critical gap by examining the current state of corporate defenses against deepfake attacks, identifying specific vulnerabilities within the existing infrastructures, and proposing a comprehensive framework (PREDICT) aiming to integrate advanced detection, robust incident responses, and continuous improvement practices. In this way, we seek to define the problem with greater clarity while also providing actionable insights to mitigate the associated risks.

1.2. Purpose and Objectives

This hybrid technical and survey paper aims to explore the emerging threat that deepfake social engineering poses to corporate enterprises in detail. It also provides a framework that aims to enable organizations to respond effectively to these emerging risks. This synthesis is aimed at cybersecurity professionals across different industries to address the gap between rapid developments in deepfake technology and the corresponding defensive strategies against it required to protect corporate assets. The specific objectives of this paper are as follows:

- Identify deepfake threats;
- Evaluate detection techniques;
- Develop defensive strategies;
- Create a comprehensive model;
- Provide actionable guidelines;
- Contribute to corporate security posture.

To achieve these objectives, in this paper, we intend to provide organizations with the knowledge and tools necessary to avoid and mitigate deepfake threats. This will hopefully help improve their ability to protect sensitive information, maintain trust, and ensure operational continuity in an increasingly digital and interconnected world.

Table 1. Comparison of related work on deepfake detection and mitigation (discussed: ✓; never mentioned: –; partially mentioned: *).

Authors	Main Contribution	Threat Landscape	Detection Techniques	Organizational Impact	AI and ML Tools	Research Challenges
[5]	Provides a broad survey on the creation and detection of deepfakes, focusing on generative models and AI-driven strategies.	✓	✓	–	✓	Organizational impact not covered
[8]	Introduces deepfakes as a threat to face recognition systems; highlights the importance of detection but less on enterprise-level mitigation.	–	✓	–	–	Threat landscape, organizational impact, and AI and ML tools not covered.
[9]	Discusses the current and future trends in deepfakes, with partial mention of organizational security measures.	✓	✓	*	✓	Organizational impact mentioned briefly.
[6]	Analyzes face manipulation techniques and ML-based detection methods; focuses on image/video authentication tests.	✓	✓	–	✓	Organizational impact not covered.
[10]	Presents a review of deepfake synthesis and detection approaches; focuses on state-of-the-art generative adversarial networks and AI-based defense.	✓	✓	–	✓	Organizational impact not covered.
[1]	Analyzes deepfake's risks to national security by highlighting potential disruptions to communications and military operations while proposing coordinated detection and mitigation strategies.	✓	✓	✓	*	AI and ML tools partially mentioned.

Table 1. Cont.

Authors	Main Contribution	Threat Landscape	Detection Techniques	Organizational Impact	AI and ML Tools	Research Challenges
[7]	Provides an analysis of deepfake-enabled cyber attacks, focusing on the implementation of targeted security controls to mitigate the risks associated with digital deception.	✓	✓	✓	–	Does not discuss AI/ML tools.
[11]	Reviews deepfake detection methods, highlighting advanced deep learning techniques, evaluation benchmarks, and emerging forensic challenges.	–	✓	–	*	Threat landscape and organizational impact not included and AI/ML tools partially discussed.
[12]	Surveys deepfake synthesis and detection techniques, integrating advanced generative models with robust machine learning to tackle emerging digital media forensic challenges.	✓	✓	–	*	Organizational impact not covered; AI/ML tools discussed partially.
[13]	Introduces a deepfake detection framework that uses spatial–temporal ensemble learning with residual networks to capture subtle manipulation artifacts.	✓	✓	*	✓	Organizational impact partially discussed.
[14]	Examines societal, legal, and security challenges while proposing a multidisciplinary framework for public trust and democratic processes.	✓	✓	✓	–	AI/ML tools not covered.
This work	This paper addresses corporate reliance on vendor security, deepfake-specific threats, real-time detection, and the Zero Trust approach (the PREDICT model).	✓	✓	✓	✓	The deepfake-driven threat landscape, detection techniques, and organizational defense strategies are covered in this study. AI-based deepfake detection, real-time media analysis, secure verification protocols, and employee training are addressed.

1.3. Research Questions and Contributions

- What threats do deepfake-driven social engineering attacks pose to corporate environments?
 - Amplification of social engineering risks: We explore how deepfakes increase the risks associated with social engineering (Sections 5 and 5.1);
 - Real-world examples: We discuss real-world cases demonstrating threats to trust and stability (Section 3.5);
 - Exploitation of communication channels: We analyze how deepfakes can be used to exploit corporate communication channels (Section 5.2).
- What are the current detection techniques?
 - Review of AI-driven tools: We evaluate AI-driven detection tools and their limitations, such as scalability and false positives (Sections 6.1 and 6.2);
 - Gaps in security solutions: We highlight the existing gaps due to reliance on generalized security solutions (Section 6.3).
- What defensive strategies are used by or recommended for organizations?
 - Employee training and multi-factor authentication: We emphasize the importance of employee training and multi-factor authentication (Sections 7.1 and 7.2);
 - Integration of policies and legal frameworks: We discuss the integration of policies, legal frameworks, and incident response plans (Section 7.3);
 - The Zero Trust deepfake framework: We propose the “Zero Trust Deepfake Framework” for robust verification processes (Section 7.4.3).
- How can organizations assess their readiness and improve their policy frameworks?
 - The PREDICT model: We have developed the PREDICT model, addressing Policies, Readiness, Education, Detection, Incident Response, Continuous Improvement, and Testing (Sections 8 and 9.2);
 - Proactive measures: We suggest proactive measures for continuous adaptation and corporate preparedness (Sections 9.2 and 11).

1.4. Outline of This Paper

Section 1 introduces the deepfake technology, as well as the problem statement, objectives, and research questions of this work. Section 4 details the semi-structured interview approach with cybersecurity professionals, including the participant recruitment, data collection, and thematic analysis. Section 5 reviews corporate vulnerabilities to deepfakes, social engineering tactics, case studies, and the impact on security, reputation, and finances. Section 6 evaluates the current deepfake detection methods, including AI tools, image/audio analysis, and general cybersecurity measures, while addressing their limitations. Section 7 proposes measures including employee training, policy frameworks, and advanced technologies such as real-time monitoring and AI-driven tools. Section 8 proposes a framework for dealing with AI-driven threats specially tailored to corporations. Section 9 reflects on the findings, highlighting existing gaps and the urgency of defensive measures. Section 10 provides insights on potential future research directions. Section 11 concludes this paper.

2. Related Work

The rapid advancement in deep learning (DL) and artificial intelligence (AI) has enabled the creation of highly realistic AI-generated media known as deepfakes. These advancements have significantly altered the cybersecurity threat landscape, transitioning from conventional disinformation tactics, such as email phishing, to more sophisticated social engineering attacks. Initially, researchers explored the impact of deepfakes on the

general public, focusing particularly on public discourse, trust, and privacy [14]. However, recent studies focus more on the dramatic improvements in the authenticity of deepfakes and the need for robust detection and defense mechanisms [8,9]. Table 1 summarizes these studies by comparing their main contributions along key dimensions such as the threat landscape, detection techniques, and the use of AI/ML tools.

2.1. Deepfake Technologies and Threats

Researchers have extensively documented how advancements in generative adversarial networks (GANs) and deep learning techniques make it easier to produce realistic synthetic media [5,6,10]. These technologies not only allow for the alteration of images and videos but also enable voice mimicry and the generation of fabricated texts, making it simpler for attackers to pose as executives, public figures, or trusted associates with very little effort and not very sophisticated techniques [1,7]. The current research warns that as these technologies continue to develop, deepfakes can be weaponized to enable a range of malicious activities, including the spread of fake news, different hoaxes, financial fraud, and public defamation, thus undermining the trustworthiness of online information [11].

2.2. Detection Techniques and Challenges

The ongoing race between the generation and detection of deepfakes drives much of the current research. Initially, the detection attempts focused on identifying visual artifacts (e.g., discoloration, out-of-place shapes, or similar unwanted visual anomalies), inconsistencies in facial expressions, or unnatural image boundaries [8]. Evolving methods have incorporated deep learning models, convolutional neural networks (CNNs), and multimodal approaches—analyzing both visual and audio cues—to detect deepfakes more accurately [11,12]. That being said, even state-of-the-art techniques (e.g., data based on FaceForensics++, DeepFaceLab detection, etc.) face large challenges. Attackers continually improve their generation techniques to bypass known detection algorithms. This leads to an ongoing cycle of attack and defense [13]. On top of this, many detection methods struggle with scalability, real-time analyses, and the typical dynamic environments of corporate networks and communication channels [5,10].

2.3. Defensive Strategies and Frameworks

In response to these challenges, recent researchers have focused their attention on building holistic defense strategies. These strategies integrate technical, procedural, and policy-based defenses. The technological solutions include improved real-time detection systems [11], the model-based detection of synthetic media artifacts [15], and distributed ledger or blockchain solutions for detecting media's origins [9]. On the organizational side, preventive measures such as the comprehensive training of employees, routine authentication checks, and the establishment of trusted communication protocols have gained traction [1]. On top of this, researchers recommend that organizations adopt frameworks that provide guidance on brand protection, legal recourse, and robust incident response plans to mitigate deepfake threats [7].

2.4. Future Directions and Research Gaps

While the body of work on deepfake generation and detection is growing, significant gaps remain. These include the ability to rapidly adapt the detection techniques to novel attack vectors and integrate seamless verification into corporate workflows. Additionally, standardized benchmarks for evaluating detection tools are still lacking [5,12]. There is also a pressing need for collaborative efforts among technologists, policymakers, and industry stakeholders to ensure that legal frameworks, standards, and educational campaigns keep pace with the rapid evolution of deepfake technologies [6]. Future research efforts will

likely center on more resilient hybrid models that combine multimodal detection techniques with strong cryptographic verification and improved AI-driven anomaly analysis to help organizations proactively identify and counter emerging deepfake-driven social engineering threats.

2.5. A Comparative Table of Related Work

Building on the previous section on related work, the subsequent table provides a concise comparison of the reviewed studies. Table 1 establishes the key dimensions, such as the threat landscape, detection techniques, the organizational impact, and the use of AI and ML tools, that are highlighted in this literature. The purpose of this comparative overview is to serve as a tool for identifying areas where further research is needed to improve organizational preparedness for deepfake-driven social engineering attacks.

This paper makes a significant contribution by presenting a comprehensive framework in the PREDICT model that directly addresses the deepfake-specific threats in corporate environments. It uniquely focuses on the limitations of existing corporate security solutions and proposes mitigation strategies to counter the evolving threat landscape. This study integrates advanced AI and machine learning tools for media analysis, secure verification, and continuous improvements in security. Overall, this work provides a more holistic strategy that bridges the gap between technical detection methods and effective corporate defense measures.

3. Deepfake Technologies and the Emergence of Deepfake-Driven Social Engineering

3.1. The Definition of Deepfakes

Deepfakes refer to artificially generated or modified digital media—videos, images, audio, or text—created through advanced machine learning (ML) techniques. In particular, *generative adversarial networks* (GANs) are particularly notable for their ability to synthesize or manipulate content in a highly realistic manner [5,6]. Initially, GANs were used for tasks such as image enhancement and style transfer [16]. However, malicious applications of GANs have significantly increased in recent years [7]. By leveraging large amounts of training data, such as publicly accessible videos or voice recordings, attackers can create deepfakes that convincingly replicate the appearance, voice, or even mannerisms of real individuals. In addition to GANs, other models like transformers and diffusion-based models have gained popularity for generating deepfakes. These models are readily available and easy to use in popular large language models (LLMs) like GPT-4 from OpenAI, Dall-E 2 or Gemini Flash 2.0 from Google [17].

From a technical perspective, different categories of deepfake manipulations exist [10]:

- **Face swaps and face reenactment:** A common application of deepfakes is to map a person's face onto another person's body or reenact someone's facial expressions in real time;
- **Lip-sync and voice cloning:** Attackers can synchronize a speaker's lip movements with *audio* that is artificially generated or adapted so that it appears as though the speaker is delivering a different message.

Together, these technologies lower the barrier of entry for highly convincing content creation, thereby opening the door to more sophisticated social engineering attacks [10].

3.2. The Definition of Deepfake-Driven Social Engineering

Deepfake-driven social engineering represents an emerging threat paradigm where malicious actors use AI-generated or manipulated media to deceive targets, often bypassing traditional security controls [1]. In classic social engineering, attackers exploit human

psychology, using impersonation, authority, or urgency to prompt victims into revealing confidential information or performing unauthorized actions. Deepfake-driven social engineering intensifies these tactics by adding hyper-realistic audio, video, or images of trusted individuals.

For instance, a deepfake video might depict a Chief Executive Officer (CEO) urgently requesting a wire transfer. Alternatively, a synthetic voice message could appear to come from a trusted stakeholder demanding sensitive documents. These AI-driven impersonations can be remarkably believable, increasing the risk that employees or even advanced security systems will fail to detect the deception [8]. The psychological cues that make social engineering successful—such as perceived authority or familiarity—become more convincing when backed by near-flawless digital impersonations [8].

3.3. Relevant Threat Vectors

Although deepfake technologies originate from advancements in generative modeling, their rapid adoption for illegal activities spans multiple threat vectors [11]:

- **Financial fraud:** Attackers can impersonate high-level executives or managers, instructing employees to authorize payments or disclose sensitive financial data;
- **Espionage and data exfiltration:** By impersonating key personnel, cybercriminals may gain privileged access to systems or coax employees into divulging intellectual property or other strategic information;
- **Disinformation campaigns:** Deepfakes can create fake public statements or events, damaging a corporation's reputation and eroding trust among stakeholders;
- **Phishing and vishing (voice phishing):** These consist of deepfake audio calls or video calls that mimic IT support staff or external vendors, tricking employees into resetting passwords or installing malware.

As outlined in the following sections, these attack vectors exploit the trust placed in digital communication, ultimately underscoring the urgency of robust detection tools and well-informed defensive strategies [11].

3.4. Relevance to Corporate Environments

The convergence of deepfakes and social engineering presents unique risks to corporate organizations [7]. Executives and employees often use and rely on virtual meetings, emails, and instant messaging for daily operations. A single convincing deepfake message can override the standard security protocols and lead to the following:

- **Unauthorized transactions:** Impersonated instructions for wire transfers or critical financial approval;
- **Reputational damage:** Falsified statements or controversies involving senior leadership, quickly spreading across social media;
- **Operational disruption:** Fake directives that disrupt supply chains, vendor relations, or production cycles.

Throughout this paper, we focus on these corporate vulnerabilities while proposing methods to safeguard assets against deepfake-driven social engineering threats. By *defining* these core concepts, we aim to establish a clear framework for understanding the subsequent sections on the threats, detection techniques, and defensive strategies in the corporate landscape.

3.5. Case Studies and the Implications of Deepfake-Driven Social Engineering

One of the more prominent cases of the application of deepfake-driven social engineering is the case of Russian President Vladimir Putin. A deepfake AI-generated avatar of Putin was engaging with the public on sensitive topics, including the perils of AI and

neural networks. This deepfake created by a student from St. Petersburg was so realistic that for a moment, it blurred the line between reality and illusion [18]. Another example of deepfake-driven social engineering was a case of Putin declaring mobilization on state TV [19]. This case further demonstrates the potential societal implications of deepfake technology when employed by malicious actors.

Both cases became talking points during our set of interviews on deepfake-driven social engineering with various corporations. This could underline how such sophisticated deepfakes may take advantage of trust in leadership and credibility. The interviews highlighted several implications for corporations. These include reputational damage, erosion of trust in official communications, and the rapid dissemination of misinformation via social media. These findings demonstrated how quickly a deepfake incident can go viral, influencing public opinion, corporate stability, or financial markets. Especially in an era where even a few believers can easily spread disinformation as genuine, this impact is magnified. A close analysis of these events underscores the critical importance of robust mechanisms, increased public awareness, and strategic corporate defenses in addressing these emerging technological threats [18].

In addition to political incidents, such as the deepfakes involving Russian President Vladimir Putin, there have been notable real-world cases of corporate deepfake-driven social engineering attacks, which emphasize the immediate risks faced by companies.

In addition to these high-profile political examples, there have been other instances targeting corporations. For example, a case reported by Damiani [20] in *Forbes* describes how a voice deepfake was used to trick a CEO out of about USD 243,000. During this incident, the scammer used AI voice-mimicking technology to impersonate the CEO of the German parent company of a UK-based energy firm. The attack chain happened over several phone calls:

1. **Initial contact:** The scammer first called the CEO, imitating the superior's voice and instructing an immediate transfer of EUR 220,000 (approximately USD 243,000) to a Hungarian supplier.
2. **Follow-up and reimbursement claim:** A second call was made to falsely claim that the payment had been reimbursed.
3. **Suspicion triggered:** On a third call—made from an Austrian phone number—the CEO grew suspicious when the promised reimbursement did not materialize. At this point, the fraudulent chain was interrupted, although the initial transfer had already been executed.

Another case reported by CNN [21] highlights an example of deepfake-driven fraud in the corporate world. In this incident, a finance worker at a global company was tricked into approving a financial transfer of about USD 25 million during a video conference call. The worker believed he was participating in a video call with several colleagues, but in reality, all of the participants were deepfakes. According to Hong Kong police, the scammers used deepfake technology to impersonate the company's CFO, among others. Initially, the worker received a message that was supposedly from the UK-based CFO, telling him to carry out an urgent money transfer. The finance worker was convinced by the realistic video feeds and voices on the call and authorized the transfer of HKD 200 million (amounting to about USD 25.6 million). The scam was only discovered when the employee later had the transaction verified with the head office of the company. This case shows how attackers can exploit trusted communication channels using deepfakes to bypass the standard verification processes, exposing critical gaps in current corporate security systems.

These two cases reveal not only significant financial losses but also expose broader vulnerabilities and issues in corporate security frameworks. Both the USD 243,000 incident and the USD 25 million scam show how deepfake technology enables scammers to

manipulate trusted communication channels, bypass established verification protocols, and execute sophisticated scams. The attack chains in these incidents show that without specialized deepfake detection tools and robust countermeasures, organizations remain at high risk. Moving forward, it is important that companies integrate dedicated deepfake detection systems, enhance employee training, and revise their security frameworks to mitigate the consequences of AI-driven fraud.

4. Methodology

4.1. Interview Development

To investigate the current measures modern companies use to detect and prevent deepfake attacks, we conducted interviews with cybersecurity professionals across various industries. These interviews provided qualitative insights into how organizations are addressing the evolving threats posed by deepfakes. Given the complexity and emergent nature of deepfake technologies, interviews (i.e., discussions) were chosen over fixed-response surveys to capture the evolving methods that companies may employ better.

The aim of the questions that were selected for the interviews was to highlight important segments related to deepfake security, including the detection technologies, challenges in operations, response strategies, and prevention measures. Our interview questions (see Appendix A) covered five main topics:

- Detection technologies and algorithms;
- Prevention measures;
- Training and awareness;
- Operational challenges;
- Future plans and adaptations.

These topics provided insights into a set of different measures that could be taken against upcoming deepfake threats. We specifically designed the questions to be open-ended, ensuring that the participants could elaborate on areas most relevant to their organizations' experiences and concerns.

4.2. The Participants and Data Collection

To capture a broad range of perspectives, we interviewed cybersecurity professionals coming from five companies in the sectors of IT, finance, and education and research. The participants, all involved in their organizations' cybersecurity operations, shared insights on countering deepfakes. The interviews were conducted remotely via video conferencing, email, or online questionnaires. Each session lasted 1–2 h. Table 2 summarizes the participating companies:

Table 2. Overview of the participants and communication methods.

Company	Participants	Method of Communication
Company A	3	Email and Video Interview (2 h)
Company B	1	Video Interview (1 h)
Company C	1	Questionnaire
Company D	1	Email
DTU	1	Email (Weak Conversation)

The data collected from the interviews helped establish a foundation for further research and highlighted key areas where corporate preparedness can be enhanced to address evolving deepfake-driven social engineering risks. Acknowledging logistical and scheduling constraints, a mix of interview formats was used for the companies surveyed. Despite the variation in the interview formats, efforts were made to ensure the comparability of the insights collected in the interviews. Considering the qualitative nature of the data collected,

we maintain that the methodology achieves a balance between effective data collection and practical feasibility.

The surveyed companies are described below according to their size, industry, and market presence. The descriptions have been kept deliberately vague to maintain the anonymity of the participant organizations:

- **Company A** is a large multidisciplinary engineering and consultancy firm headquartered in Northern Europe with a global presence;
- **Company B** is a leading IT consultancy specializing in digital solutions for public and private sector clients, with a pronounced presence across Europe;
- **Company C** is a medium-sized company dealing with debt collection and credit management, serving clients across Northern Europe;
- **Company D** is a smaller cybersecurity firm focused on threat intelligence, detection, and incident response services located in Northern Europe.

When selecting companies to interview, we prioritized those more likely to encounter deepfake-related threats in their daily operations. The interviewees held roles ranging from technical to managerial positions. This diversity ensured a comprehensive view of organizational strategies and challenges. Furthermore, the companies were selected to reflect a diverse range of organizational sizes, sectors, and levels of digital expertise. By including both highly digitized companies with established cybersecurity infrastructures and companies with developing security infrastructures, this paper aims to capture a broader range of preparedness for deepfake attacks. The goal of this sampling was to allow for comparative insights into how different industries assess the risk of deepfake-driven social engineering threats. While we recognize that this sample is not representative of all sectors, we believe that the selected organizations offer relevant and diverse perspectives that support the arguments presented in this paper, providing valuable insights into the preparedness for deepfake-driven social engineering threats and making valuable contributions to the field.

4.3. Data Analysis

We performed a thematic analysis to uncover how organizations perceive and address deepfake threats. The analysis was conducted on the interview transcripts with iterative coding, followed by theme refinement, to ensure a transparent, reproducible process of identifying the core issues in deepfake mitigation. Inspiration for this section was taken from Braun and Clarke's guidelines for a thematic analysis [22]. Our five steps were as follows:

1. Transcribing and reviewing the interviews;
2. Assigning initial descriptive codes (e.g., *awareness level*, *vendor reliance*);
3. Clustering the codes into themes (e.g., *Organizational Readiness*, *Mitigation Strategies*);
4. Refining and removing redundant categories;
5. Summarizing the findings in a thematic table (Appendix B).

Insights from the Thematic Analysis

Following the thematic analysis, we found the following core issues based on the interviews conducted.

- **Organizational readiness:** Most view deepfakes as a looming threat but remain reactive due to scarce real-life incidents;
- **Mitigation strategies:** Organizations rely on general security controls (e.g., MFA, phishing filters) rather than dedicated deepfake solutions;
- **Detection gaps:** None deploy specialized deepfake detectors, instead trusting existing security vendors;

- **Training limitations:** Deepfake-focused training is uncommon, leaving employees underprepared for AI-driven social engineering.

Overall, organizations depend on standard cybersecurity measures and vendor trust, indicating a need for targeted deepfake detection and response strategies.

4.4. The Literature Search and Study Selection

A semi-structured literature search was conducted in the literature review process. The procedure broadly involved the following stages:

- **Identification:** We searched multiple databases using a combination of keywords related to deepfake-driven social engineering, cybersecurity, and detection methods. This initial search resulted in approximately 60 studies.
- **Screening:** Duplicates were removed, and the remaining studies were initially screened based on their titles and abstracts. Studies that did not meet the preliminary inclusion criteria of relevance to the core subjects were excluded. This process resulted in approximately 45 studies being selected for full-text review.
- **Eligibility:** The full texts of the remaining studies were assessed against inclusion criteria such as their focus on deepfake detection methods, empirical evaluation of techniques, or integration into cybersecurity frameworks. Studies failing to meet these criteria were left out, which left approximately 35 studies.
- **Inclusion:** The final review comprised approximately 30 studies that were subsequently cited in this paper. This manuscript cites 31 sources in total (articles and research studies). Nevertheless, it is important to highlight that a broader set of studies was thoroughly reviewed during the review process to create a comprehensive understanding of the literature.

5. Deepfake-Driven Social Engineering Threats in Corporate Environments

Social engineering is the general term for the art of manipulating individuals to disclose confidential or personal information and to conduct certain actions that might compromise an organization's security. Rather than exploiting vulnerabilities in software or hardware, social engineering primarily targets human psychology and uses trust, authority, curiosity, or even fear to compel people to share sensitive data or take unauthorized actions [23]. Organizations and their employees may be vulnerable to fake trading practices. These include fake online accounts used in attempts at social engineering, fraudulent text and voice messages used to avoid technical defenses, and fake videos used to spread disinformation [1].

5.1. Deepfake Technology in Social Engineering

Deepfake technology has become one of the major threats to social engineering, as it gives attackers the ability to create realistic but false digital content such as videos and audio clips. Through all of the interviews conducted, it was demonstrated that although there is an increased awareness of deepfakes in corporate environments, minimal efforts have been implemented to counteract them.

During the interviews, we found that companies are aware of the potential harm of such impersonations; however, their existing measures are mostly related to general phishing and scam prevention. For instance, deepfake-based phishing often involves synthetic audio or video mimicking a manager's or an executive's voice instructing an employee to complete a financial transaction without verifying the standard protocols [12]. Company A and Company B clearly stated that they are providing training on phishing

recognition and scam avoidance. However, none have developed or implemented specific strategies or detection tools specifically designed to address deepfake-based impersonation.

Deepfakes also pose threats to corporate stability and reputation in the transmission of fabricated statements or actions allegedly carried out by prominent individuals in an organization. For instance, a deepfake video could show the CEO engaging in unethical behavior, posing a risk of severe damage to the company's image [9]. We discussed this exact example with the companies.

The potential for misuse of deepfakes is heightened by their scalability and accessibility. Increasingly, only minimal technical expertise is required to use intuitive deepfake tools. As a result, individuals with harmful intentions can execute highly persuasive social engineering tactics [12]. Deepfake technology applied in social engineering is one of the cases showing an increased need to raise awareness and implement sophisticated countermeasures to minimize its growing impact on corporate security and trust [1].

5.2. The Impact on Corporate Security

From the interview data, we learned that although these companies recognize the threat posed by deepfakes, none have deployed dedicated countermeasures against them. This underscores a heightened vulnerability, as attackers can use deepfake-driven impersonations to exploit trust in familiar voices or faces, potentially bypassing traditional verification processes. These attacks can facilitate fraudulent transactions and lead to the leakage of confidential information, ultimately damaging the reputation of a company or the key individuals within it [14].

Some companies, such as Facebook and DARPA, have already begun to invest in deepfake detection technology [24]. This investment highlights the seriousness of the threat and its potential damage to corporate environments [24]. Company A and Company B indicated that they rely on general security tools provided by well-known vendors like Microsoft. However, these tools are not specifically designed to identify deepfake manipulations. This has rendered most the current frameworks inadequate for picking up anomalies associated with AI-driven forgeries. Also, while cybersecurity training does touch on phishing and scam awareness, very little involves comprehensive deepfake education, pointing to a potential need for formal policies, specialized systems of detection, and updates to employee training specifically for the threat of deepfake-driven social engineering. Figure 1 illustrates the key insights we have gathered from the interviews and discussions with companies about deepfake-driven social engineering. It displays how deepfake attacks can result in consequences such as reputational damage, the leakage of sensitive information, and financial losses through various attacks like email scams, impersonation, and phishing calls.

Survey data indicate that 27–50% of individuals cannot reliably distinguish real video content from a deepfake [25]. In the corporate setting illustrated by Figure 1, this statistic highlights how easily attackers could take advantage of fake video impersonations to compromise sensitive information or trigger fraudulent transactions. Intriguingly, adults and working professionals, those most likely to have decision-making roles in a company, have exhibited greater vulnerability to deepfakes than younger student populations [25]. This finding reinforces the need for specialized detection strategies and employee training, as even experienced staff can be susceptible to deepfake-driven social engineering attacks.

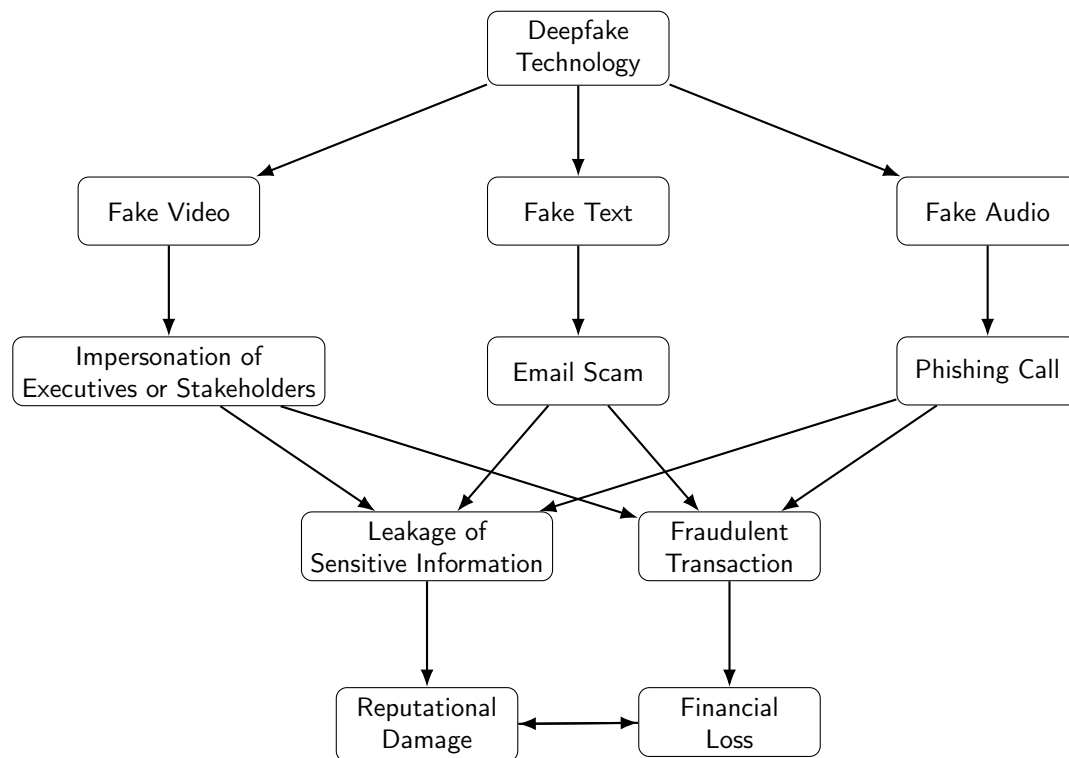


Figure 1. The threats of deepfake technology.

6. Detection Techniques for Deepfake Attacks

6.1. General Cybersecurity Measures

The detection of deepfake attacks within corporate environments remains a significant challenge, as evidenced by the interviews with industry professionals. Many of the organizations we spoke with currently lack tools specifically designed to identify deepfake content. Instead, there is a tendency to rely on broader cybersecurity solutions provided by major technology vendors, such as Microsoft. Throughout our discussions with cybersecurity professionals, we frequently encountered variations of the sentiment “We use and rely on whatever Microsoft provides” when it came to security. This reliance underscores a clear trend: companies prioritize established security frameworks rather than adopting niche detection technologies specifically designed to combat deepfakes.

In particular, the companies often adopted these vendor-provided solutions due to their reliability and ease of integration and the comprehensive support offered by established providers. As an example, multi-factor authentication (2FA), often through an SSO solution, is a standard security measure implemented to improve user verification processes, reducing the risk of unauthorized access that could be exploited by deepfake-driven social engineering attacks. Some companies recognized the need for deepfake-focused security. However, none were willing to implement dedicated detection measures until deepfakes were perceived as a widespread threat. Relying solely on comprehensive security solutions presents significant drawbacks in addressing deepfake-driven social engineering attacks. No universal solution can be fully tailored to the specific needs of companies given the diverse threat landscape that they might face [26]. As evidenced by our data derived from the interviews conducted and our review of the current studies, there is currently no pre-tailored security solution implemented to specifically target and detect deepfake attacks.

In the absence of dedicated deepfake detection measures, the organizations generally focused on strengthening their overall cybersecurity posture. This included conducting

regular phishing simulations, securely managing their cloud environments, and addressing vulnerabilities identified through penetration testing. By fortifying these foundational security aspects, the companies aimed to create a resilient infrastructure capable of mitigating various types of cyber threats, including those posed by deepfakes. For instance, Company A had security measures in place to detect abnormal metadata in Microsoft Teams. However, these measures were not specifically implemented to detect deepfakes but rather to address a broader spectrum of social engineering attacks that could potentially include deepfake attacks.

6.2. Image-, Audio-, and Video-Based Detection

6.2.1. The Current Approaches and Technologies

During the interview, organizations such as Company B recognized the importance of image- and video-based deepfake detection mechanisms. While Company B currently does not employ specialized deepfake detection tools, its internal security framework emphasizes the integration of innovative technologies including convolutional neural networks (CNNs) and Deep Neural Networks (DNNs) to enhance its overall cybersecurity measures. These neural network architectures are used to identify subtle artifacts and errors within visual and auditory media that indicate the presence of deepfake manipulations. For a possible high-level setup, see Appendix C.

6.2.2. Forensic Analysis Integration

Company B and Company A both underscore the role of using a version of forensic analysis in their security protocols. Forensic techniques are utilized to examine the metadata and frame-by-frame integrity of images, audio, and videos. By using forensic analyses, Company B and Company A would like to be able to detect signs of tampering, such as irregular lighting, inconsistent shadows, and unnatural facial movements, which are often telltale signs of deepfake content. While this is currently not part of their setups, this again seemed to be a case of being an external dependency. Company A specifically mentioned that it would wait for Microsoft itself to implement this into its services like Microsoft Teams instead of attempting to set up its own forensic analysis pipelines for deepfakes specifically.

6.3. Limitations of the Detection Techniques

6.3.1. Scalability Challenges

One of the primary limitations mentioned by Company B and Company A during the interviews is the problem of scalability in the current deepfake detection techniques. As deepfake technologies improve, the volume and complexity of synthetic media increase. Detection systems need to scale efficiently without compromising their performance. Company B highlighted the need for scalable solutions capable of the real-time analysis of vast amounts of image, video, and audio data across its extensive operations. During the interviews, we primarily discussed this issue in the context of video streams through Microsoft Teams, which was used by all of the companies.

6.3.2. Speed and Real-Time Detection

The speed at which deepfake detection tools can process and analyze media content is another significant challenge. During the interviews, we discussed the importance of real-time detection capabilities. Such capabilities are crucial for identifying and mitigating deepfake threats without hindering productivity. Current techniques such as CNNs and DNNs, while effective, often require substantial computational resources and time, limiting their practicality for an immediate threat response. This is especially critical in video call scenarios, where even slight latency is highly noticeable.

6.3.3. Accuracy and False Positives

Achieving high accuracy in deepfake detection is also a concern. We noted that the existing methods can struggle with precision, leading to false positive rates that may cause distrust in the detection systems. Balancing the sensitivity to detect genuine deepfakes while minimizing false alarms is essential for maintaining operational efficiency and employee confidence in the security measures.

6.3.4. Integration with Existing Protocols

Another limitation is the seamless integration of deepfake detection technologies with existing cybersecurity frameworks. We found that there are operational challenges in integrating advanced detection tools into existing company protocols and policies. Such integrations would require adjustments and regular updates to remain effective.

6.4. An Evaluation of the Detection Approaches

Generally, detection approaches are currently built on the assumption that a deepfake threat is not advanced enough to be a real threat yet. As mentioned, we see large reliance on externally provided security measures such as software and systems, and none of these seem to provide any software-based detection specifically tailored to deepfakes. Some areas of these companies have higher deepfake-related risks. These are typically outward-facing roles like sales or reputational risks (e.g., deepfake videos of a CEO meant for the public to harm the company). The detailed overview of detection techniques is given in Table 3.

Table 3. Overview of detection techniques.

Category	Findings	Challenges
General Cybersecurity Measures	Companies rely on broad cybersecurity measures from major vendors (e.g., Microsoft). Multi-factor authentication and phishing simulations are common measures. No dedicated deepfake detection tools are in use.	There is a lack of specific deepfake detection solutions. Companies only react when threats become extensive.
Image-, Audio-, and Video-Based Detection	Some companies are exploring the use of CNNs and DNNs to detect deepfake artifacts. A forensic analysis can help detect inconsistencies in media (e.g., shadows, facial movements).	High computational cost and dependency on external tools (e.g., Microsoft).
Scalability Challenges	There is a need for detection systems to handle increasing occurrences of deepfakes and volumes of media. Real-time monitoring is a priority (e.g., Microsoft Teams).	Processing large-scale media efficiently without performance issues.
Speed and Real-Time Detection	Companies require real-time deepfake detection, especially for live interactions (e.g., video calls).	CNNs and DNNs require significant computing power, which can cause potential delays.
Accuracy and False Positives	Balancing the sensitivity to detect deepfakes while avoiding false positives is crucial.	High false positive rates can lower trust in detection systems.
Integration with Existing Protocols	Companies prefer security solutions that integrate seamlessly with their current infrastructure.	Difficulty embedding advanced deepfake detection within existing cybersecurity frameworks.
Overall Evaluation	The interviewed companies primarily identify deepfakes as a future risk but do not yet implement targeted detection measures. Higher-risk areas include external-facing roles (e.g., sales, executive communication).	Reliance on third-party security vendors rather than in-house solutions.

6.5. The Performance Metrics for Deepfake Detection Algorithms

Table 4 summarizes notable performance metrics reported in recent studies on deepfake detection [27,28]. Instead of using accuracy and false positive rates, the Area Under the Curve (AUC) is utilized, providing a more comprehensive evaluation of the algorithms'

performance across different threshold settings. While accuracy and the false positive rate (FPR) are common indicators for evaluating classification models, the referenced papers primarily report the AUC, as it reflects the detection performance in imbalanced datasets better. Future evaluations could incorporate these additional metrics for a more comprehensive assessment. Two datasets—Celeb-DF and FaceForensics++ (FF++)—are used for benchmarking.

Table 4. Performance metrics of prominent deepfake detection algorithms.

Model	Methodology	AUC (%)	Source Datasets
Capsule	CNN-based analysis	96.6	FF++
		93.2	Celeb-DF
DSW-FPA	Face warping artifact analysis	93.0	FF++
		94.8	Celeb-DF
DCViT	Convolutional vision transformers	98.3	FF++
		97.2	Celeb-DF
Xception	A depthwise-separable CNN	99.7	FF++
		97.6	Celeb-DF
CNN + RNN	A CNN and an RNN	99.7	FF++
		61.5	Celeb-DF
LRNet	A lightweight, two-stream RNN architecture	99.9	FF++
		56.9	Celeb-DF

As shown in Table 4, several deepfake detection models achieve high AUC scores across both Celeb-DF and FF++. Models such as Xception and DCViT demonstrate a particularly strong performance, with their AUC values exceeding 97% on both datasets. Since the results for some models appeared in more than one study, the highest reported AUC values were selected to reflect the best performance achievable. This ensures that the table highlights the strongest capabilities of each model under the optimal conditions.

While these results indicate strong potential for deepfake detection, deploying these models in cybersecurity requires additional work. The performance may vary depending on several factors. These include the video resolution, compression artifacts, and differences between the training data and the actual deepfake generation techniques. To achieve reliable detection, the models would most likely need additional fine-tuning on domain-relevant datasets, calibration to reduce false positives, and integration with other techniques, such as temporal consistency analyses or multimodal detection incorporating audio cues. It is also worth noting that as deepfake technology continues to evolve, cybersecurity strategies must include ongoing model updates and adversarial robustness testing to maintain their effectiveness against increasingly sophisticated manipulations.

7. Defensive Strategies Against Deepfake-Driven Social Engineering

As deepfake technology advances, organizations must implement comprehensive defensive strategies to mitigate the risks associated with deepfake-driven social engineering attacks. In this section, we outline the current defensive measures, explore technological defenses, discuss corporate policies and legal frameworks, and highlight directions for enhancing corporate security against deepfake threats. These insights are supported by industry professionals and visualized in Appendix D. The detailed overview of defensive strategies is presented in the Table 5.

Table 5. Overview of defensive strategies.

Category	Findings	Challenges
Employee Training and Awareness Programs	Organizations focus on training their employees so that they can identify phishing attempts and verify communications. Reporting mechanisms for suspicious content are established.	The integration of deepfake recognition into training modules is still in development.
General Security Frameworks	Reliance on security frameworks from major vendors like Microsoft. These provide basic protection but are not tailored specifically to deepfake detection.	Potential gaps in defenses emerge as deepfake technology continuously advances.
AI- and Machine-Learning-Based Detection Tools	AI and machine learning (e.g., GANs, CNNs) are used to detect visual and audio anomalies in videos/audios that might have been manipulated. Multimodal approaches combining visual and audio analysis will be considered in the future.	Companies have yet to implement specialized deepfake detection tools.
Automated Alerts and Real-Time Monitoring	Real-time monitoring systems can detect and respond to deepfake incidents quickly.	Specialized deepfake detections systems are not yet widely implemented. There are high computational costs and potential delays.
Integration with Existing Cybersecurity Infrastructure	Seamless integration with current cybersecurity measures (e.g., firewalls, intrusion detection systems) is crucial.	Difficulty embedding advanced deepfake detection within current frameworks.
Privacy Policies	Comprehensive privacy policies help protect sensitive information from deepfake attacks.	The existing policies may not fully address future advancements in deepfake technology.
Legal Regulations and Compliance	Adherence to legal regulations (e.g., the GDPR) and industry standards is essential. Continuous monitoring and adaptation are required.	Keeping up with new laws related to digital impersonation and cybercrime.
Incident Response Plans	Robust incident response plans are in place, but specific plans for deepfake-related incidents are underdeveloped.	The need for more detailed and specific response strategies for deepfake attacks.
Real-Time Monitoring Systems	There is set to be further focus on real-time systems capable of detecting deepfake content at all times in the future.	Technologies are still in the early stages and not yet widely adopted.
Collaboration with Cybersecurity Experts and Policymakers	Partnerships with cybersecurity firms can provide access to advanced detection technologies and expert knowledge.	Requires investment and coordination with external entities.
Zero Trust Deepfake Framework	The Zero Trust framework can be adopted, including principles such as continuous verification, micro-segmentation, least privilege access, and advanced authentication methods.	This framework is still under development.

7.1. The Current Defensive Measures

7.1.1. Employee Training and Awareness Programs

One of the primary defenses against deepfake-driven social engineering is extensive employee training and awareness programs [1]. Through our research, we observed that organizations recognize human vigilance as an important line of defense.

- **Phishing avoidance:** From the interviews with Company A and Company B, training focuses on educating employees to identify phishing attempts, including those potentially using deepfake technologies. For example, from the interview with Company A, we learned that it could foresee a future where deepfake recognition training would be integrated into its annual training modules.
- **Verification protocols:** Employees should be trained to verify how authentic communications are, where especially those requesting sensitive information or financial transactions are focused on. From the interview with Company A, we learned that it highlighted the importance of validating communications through known channels.
- **Reporting mechanisms:** It has become evident to us that establishing clear procedures for reporting suspected deepfake content or other suspicious activities is crucial.

7.1.2. General Security Frameworks

From the interview data, we have learned that many organizations continue to rely on established security frameworks provided by major technology vendors, such as Microsoft's security solutions. These frameworks offer foundational protections against a wide range of cyber threats, including traditional phishing and malware attacks. However, as follows from the interview data, there is a prevalent reliance on these general solutions, which may not be specifically tailored to detecting or mitigating deepfake threats specifically. For instance, Company B currently relies on Microsoft's security tools without specialized deepfake detection systems. This reliance could leave a gap in its defenses if deepfake technology advances significantly.

7.2. Technological Defenses

7.2.1. AI- and Machine-Learning-Based Detection Tools

We maintain that advanced technological defense strategies can play a large role in identifying and mitigating deepfake threats. These tools leverage AI and machine learning to analyze digital content for signs of tampering or manipulation [10]. The approaches include GANs and CNNs, which are used to find visual anomalies in images or videos that could indicate deepfake manipulation. However, as we can extrapolate from the conducted interviews, many corporate organizations have yet to implement specialized tools like these, as they are relying on broader cybersecurity solutions. A multimodal approach where organizations combine visual and audio analyses to achieve a greater detection accuracy, especially in multimedia content, could be implemented broadly in the future. This was also a point of discussion in the interview with Company B, as it saw this as a potential area for further implementation.

7.2.2. Automated Alerts and Real-Time Monitoring

Automated alert systems and real-time monitoring can help organizations quickly detect and respond to potential deepfake incidents. These systems continuously scan established communication channels for suspicious content and generate alerts when anomalies are detected. However, such specialized systems for detecting deepfake content have not yet been implemented within the organizations interviewed. Instead, they rely on the existing capabilities of their cybersecurity tools. Furthermore, both Company A and Company B do not currently perceive deepfake-driven cybersecurity attacks as a significant enough threat to justify specialized detection systems. They acknowledge that deepfake technology could become inexpensive and sophisticated enough to warrant specialized systems. However, they currently do not consider this scenario likely enough to integrate into their immediate plans.

7.2.3. Integration with Existing Cybersecurity Infrastructures

In order to make technological defenses effective, they must be seamlessly integrated with existing cybersecurity infrastructures. This also includes ensuring that deepfake detection tools work in tandem with existing defenses like firewalls, intrusion detection systems, and other measures to provide a unified defense strategy. As evidenced by our interview data, some organizations rely on Microsoft's integrated security solutions for their operations but have not yet incorporated niche deepfake detection technologies, which could indicate that an area for potential improvement exists. As we have argued in this paper, deepfake technology may yet become sufficiently reliable and economical for threat actors to employ, and the integration of a deepfake defense strategy should be, at the very least, in the plans for future developments in the cybersecurity departments of corporate organizations.

7.3. Corporate Policy and Legal Frameworks

7.3.1. Privacy Policies

Establishing comprehensive privacy policies can help safeguard sensitive information from being accessed and exploited in deepfake attacks. These policies should outline data protection methods, access controls, and general guidelines for handling and sharing confidential or sensitive information. In our interview with Company B, the role of privacy guidelines in protecting against the misuse of sensitive data through deepfakes was highlighted. Company B argued that the current threats of deepfake attacks are covered by the general guidelines and best practices employed in its privacy policies. It believes it remains to be seen whether this threat will increase with technological advancements.

7.3.2. Legal Regulations and Compliance

It is also critical that organizations adhere to and follow the relevant legal regulations and industry standards when mitigating deepfake threats. They must stay informed about new laws related to digital impersonation, data protection, and cybercrime. Collaboration with legal entities could help ensure defensive strategies comply with the law. Additionally, it could provide clear legal resources in the event of a deepfake-related incident. For example, Company B's compliance with the GDPR and other regulations involves continuous monitoring and adaptation to ensure its defensive measures are legal.

In addition to the GDPR, which protects personal data by requiring informed consent, the proposed Artificial Intelligence Act (AI Act) aims to harmonize EU regulations by mandating that creators declare artificially generated or manipulated media. While neither framework outright bans deepfakes, both enhance traceability and accountability, though their enforceability remains challenging [29]. Organizations should work closely with legal experts to align their strategies with both current and forthcoming regulations.

7.3.3. Incident Response Plans

From the data derived from the interviews, we can highlight the importance of developing and maintaining robust incident response plans to handle deepfake attacks more effectively. Such plans should encompass key elements, including immediate containment to isolate affected systems and prevent further spread; investigation procedures to assess the breach and identify the source; recovery processes to restore operations and mitigate damage; and communication strategies to inform stakeholders about the state after an attack. Immediate containment involves quickly isolating the affected systems to prevent the spread of a deepfake attack, minimizing its impact and protecting the network. Detailed investigation procedures are essential to assess the breach, identify the source, and understand the damage. Recovery processes restore normal operations, repair systems, recover

data, and reduce the downtime. Effective communication strategies inform stakeholders, maintain trust, and manage an organization's reputation during and after an incident.

Our interviews with Company B and Company A revealed that while it is an industry standard to have extensive and detailed incident response plans in place, plans specific to deepfake-related cybersecurity incidents are underdeveloped or missing entirely. This gap highlights the need for organizations to adapt their existing response plans to address the unique challenges posed by deepfake technology. In this way, they can prepare for and respond to the evolving threat landscape better.

7.4. Future Directions

7.4.1. Real-Time Monitoring Systems

Future defensive strategies will likely be focused on the development of real-time systems which are capable of detecting deepfake content as it is released or received by members of a corporate entity. Both Company A and Company B expressed that they expect this to be the case as the technology behind deepfake attacks continues to become more sophisticated and less costly to employ. Systems like these can likely use advanced AI algorithms to provide instantaneous analyses and alerts, which could significantly reduce the size of the attack vector or window of opportunity for attackers. Currently, the organizations interviewed in the development of this paper determined that such technologies are still in the early stages but are expected to become more advanced and prevalent in the coming years.

7.4.2. Collaboration with Cybersecurity Experts and Policymakers

Corporate organizations could foster partnerships with specialized cybersecurity firms which could improve their ability to defend themselves against deepfake threats. Collaborations like these can provide access to cutting-edge detection technologies, expert knowledge, and security assessments specifically tailored to the needs of the organization. From the interview with Company A, it was suggested that collaborating with external experts could offer valuable insights which could help bolster the defense against more sophisticated deepfake technology.

7.4.3. The Zero Trust Deepfake Framework

Another strategy for addressing deepfake threats is the adoption of a Zero Trust framework, a strategy which is under development [30]. The Zero Trust model operates on the principle of “never trust, always verify”, which can ensure that all access requests are thoroughly authenticated and authorized no matter what their origin is. In the context of deepfakes, such a framework could include *continuous verification*, where regular authentication of users' identities and the integrity of communication channels would be conducted, even within internal networks. *Micro-segmentation* is another aspect which divides the network into smaller, isolated segments to prevent the lateral movement of threats. *Least privilege access* restricts user permissions to the minimum necessary, reducing the potential impact of a successful deepfake attack. Finally, *advanced authentication methods* can incorporate biometric verification and behavioral analytics to improve the accuracy of user authentication.

8. PREDICT: A Cyclical Framework for Defending Against Deepfake-Driven Social Engineering Threats

In this section, we introduce *PREDICT*, a theoretical framework, shown in Figure 2, designed to guide organizations in proactively defending against the multifaceted challenges posed by deepfake-driven social engineering attacks. *PREDICT* is not just a high-level

concept; it outlines a structured, cyclical process. This process integrates organizational policies, technical tools, and iterative improvements to address the evolving threat landscape. Below, we elaborate on each phase of PREDICT in greater technical and procedural depth, clarifying the rationale and mechanisms of the model.

Policies. Rather than only providing a set of guidelines, this phase also establishes the legal, ethical, and accountability structures necessary for AI-driven media within an organization. At the technical level, *Policies* may include setting baseline requirements for content verification tools, cryptographic signing of official digital communications, or the mandated use of secure channels for the transfer of sensitive media. These requirements create a defensive perimeter where employees, contractors, and stakeholders are formally guided in how to handle AI-generated content and are held accountable if they fail to adhere to the best practices or engage in its misuse.

Readiness. This phase encompasses the governance models and risk assessment methodologies that ensure an organization can rapidly adopt new detection and forensic tools. From a technical standpoint, *Readiness* includes establishing or licensing specialized AI infrastructure capable of analyzing media in real time. Additionally, it involves creating cross-functional committees (e.g., IT security, legal, and compliance) to regularly review emerging deepfake capabilities and vulnerabilities. In this way, the organization formalizes a chain of responsibility for quick resource allocation when potential threats arise.

Education. In addition to general security awareness, *Education* entails detailed workshops and general education on how different deepfake models work. This includes identifying common artifacts or anomalies produced by these models, such as subtle facial distortions or mismatched lighting. This technical grounding helps employees, especially those in high-risk roles, recognize suspicious media more easily. Furthermore, practical exercises—such as exposing staff to a curated set of deepfake examples—build an intuitive, experience-based capability to detect manipulation. Integrating these educational programs into broader security training ensures that both technical and non-technical personnel remain vigilant against ever-evolving tactics.

Detection. While detection capabilities remain an active research area, *Detection* in PREDICT emphasizes integrating cutting-edge techniques directly into existing workflows. This includes machine-learning-based classifiers that analyze audio, video, or image streams for signs of tampering, as well as rule-based systems that cross-reference metadata inconsistencies (e.g., timestamps, camera signatures, or voiceprint mismatches). Organizations might employ vendor solutions for advanced detection or even develop in-house detection pipelines using open-source libraries. Additionally, digital watermarking and content fingerprinting methods can be incorporated to confirm authenticity. By weaving these tools into communication platforms, suspicious content can be flagged in near-real time, allowing security teams to respond quickly.

Incident Response. Though incident response is often described at a high level, *Incident Response* in PREDICT uses forensic validation and decision-making protocols tailored to deepfake incidents. For example, once a piece of suspicious media is detected, an automated workflow might (1) quarantine the content, (2) run it through multiple AI-based deepfake detection models, and (3) generate a confidence score for authenticity. If it is flagged as high-risk, the system could automatically alert a response team comprising technical experts and communications professionals. This dual approach ensures that any resulting containment actions (e.g., locking down accounts, halting financial transactions, or issuing public statements) are informed by timely, evidence-based intelligence. Adopting well-defined escalation pathways allows organizations to respond in a coordinated fashion, preserving trust among stakeholders.

Continuous Improvement. To sustain effectiveness, PREDICT encourages a structured review of each phase after an incident or detection test. This includes examining false positives/negatives from detection algorithms, re-evaluating which employee positions might be at risk, and reviewing updates in the deepfake generation and detection research. Technically, this phase could involve re-training machine learning models on fresh datasets or refining the detection thresholds as generative models become more sophisticated. Partnerships with academic institutions or specialized cybersecurity firms may also be leveraged to stay informed on new types of manipulations and detection techniques.

Testing. Finally, the *Testing* phase validates the entire PREDICT framework. It is able to through realistic simulations of deepfake-driven attacks. This could be spear-phishing attempts using voice mimicry or manipulated videos. Conducting table-top exercises can reveal vulnerabilities in policy guidelines, employee training, or detection algorithms. These tests can employ known deepfake generation tools to produce media specifically designed to bypass existing defenses, thereby providing a critical stress test. The insights gained from testing feed back into the cycle, prompting refinements to Policies, Readiness, Education, and so forth. Furthermore, the detailed overview of the Predict framework is presented in Table 6.

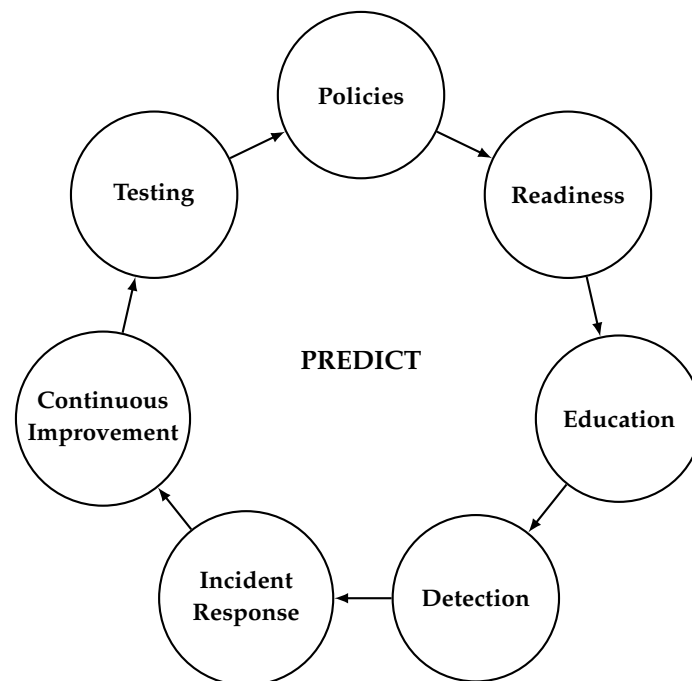


Figure 2. PREDICT lifecycle diagram.

Future Work: Effective Implementation and Validation with Organizations

Although PREDICT outlines a strategic and cyclical approach to mitigating deepfake-based social engineering, further validation in real-world organizational settings is crucial. Future efforts should focus on collaborating with companies across different sectors to implement each phase of the model under authentic constraints. This includes tailoring the detection technologies, refining the educational content, and testing incident response capabilities in simulated deepfake scenarios. Feedback derived from these experiments can then be used to assess the model's impact on reducing vulnerability and to refine the framework for broader industry adoption.

Table 6. Overview of the PREDICT model.

Phase	Purpose	Key Actions	Desired Outcome
Policies	Create guidelines for the acceptable use of generative AI and ensure accountability.	<ul style="list-style-type: none"> Define acceptable AI use and what counts as abuse; Ensure alignment with legal/compliance requirements; Establish secure communication channels for sensitive requests and reports. 	Clear organizational rules and responsibilities that allow for an environment of ethical AI usage and compliance.
Readiness	Build governance, infrastructure, and “preparedness” to face deepfake risks.	<ul style="list-style-type: none"> Conduct risk assessments; Allocate resources for detection tools and training; Form governance committees for oversight. 	An organization equipped with the necessary tools, structures, and expertise to address deepfake threats effectively.
Education	Promote a security-first culture and enable staff to identify and report deepfakes.	<ul style="list-style-type: none"> Offer awareness campaigns with real-world examples; Provide specialized training for high-risk roles; Regularly update training to address new tactics. 	A knowledgeable workforce capable of quickly recognizing and responding to potential deepfakes.
Detection	Deploy and integrate specialized tools to identify deepfakes in real time or near-real time.	<ul style="list-style-type: none"> Use AI-driven algorithms for anomaly detection; Partner with vendors for advanced detection; Apply digital forensics to validate authenticity. 	The timely and accurate identification of suspicious or manipulated media before it causes harm.
Incident Response	Contain and manage deepfake-driven incidents to minimize damage.	<ul style="list-style-type: none"> Create clear escalation workflows; Verify authenticity and freeze actions if suspicious; Coordinate internal and external communications. 	A fast and coordinated response that limits the impact of deepfake attacks and maintains stakeholder confidence.
Continuous Improvement	Refine defenses through ongoing learning, partnerships, and updates.	<ul style="list-style-type: none"> Review incidents and feedback loops; Leverage threat intelligence and expertise; Update detection technologies regularly. 	Evolving defenses that remain effective against emerging threats and tactics.
Testing	Validate the effectiveness of policies, tools, and training through simulations.	<ul style="list-style-type: none"> Conduct simulated deepfake attacks; Run table-top exercises to assess readiness; Identify gaps in policy, training, and technology. 	A proactive improvement cycle that ensures all components of the PREDICT model stay current and robust.

9. Lessons Learned and Recommendations

9.1. Anticipating the Deepfake Threat

As we were able to confirm through our interviews, many organizations now see deepfakes as a niche or low-probability threat, despite the fact that they have garnered a lot of attention in the media and cybersecurity circles. Our study’s interviewees reported that although they are aware of the possible dangers, the majority have not personally experienced fraud or impersonation connected to deepfakes. Companies have adopted a “wait and see” mentality, concentrating on more urgent cybersecurity issues like ransomware, insider threats, or conventional phishing attacks through well-established protocols and controls, as a result of this sense of relative safety and the comparatively low sophistication of many *currently circulating* deepfakes.

However, the overall direction of AI's development shows that people might not stay satisfied for long. As described by Zhou Shao et al. in their paper [31], artificial intelligence is in an extraordinary period of accelerated growth. They found that the number of AI researchers and academic papers on AI is increasing very fast, which means knowledge about AI is being created and shared faster than before.

Only a few years ago, the generation of believable synthetic media required expert-level expertise, high-end, cutting-edge computers, and a deep understanding of machine learning and AI [32]. Today, a rapidly expanding ecosystem of user-friendly tools and cloud-based AI services has enabled individuals, even those with moderate to no technical skills, to create very convincing audio, images, and videos of people. We can imagine that a small leap in generative technology relative to what we have seen just in the past few years could dramatically lower the barrier for producing near-flawless deepfakes to trick people. When combined with language generation tools, such an advance could enable attackers to create *lifelike impersonations* of practically anyone. Celebrities, political figures, or high-level corporate executives could be targeted more than they are now, at a scale and quality that could outdo the current detection methods. In this scenario, malicious actors or social engineers could convincingly instruct employees to authorize financial transactions or disclose sensitive data while defenders and security systems struggle to distinguish authentic communications from fraud. From a corporate standpoint, even a single successful deepfake-based attack can result in reputational damage, legal liabilities, and a substantial loss of trust in digital communications.

9.2. Concrete Recommendations

Our interviews indicate that many organizations still regard deepfake attacks as a low-probability threat, and therefore they lean towards “waiting and seeing” rather than proactive security planning. Nonetheless, an awareness of the accelerated pace of generative AI technologies is required in order to pre-empt potential risks before an alarming event triggers reactive action. One beneficial action is creating an internal task group or specifying stakeholders that are always responsible for monitoring the development of deepfakes so that senior management can appreciate such danger as a strategic problem and not as a hypothetical chance.

Even if most firms already have rigid cybersecurity procedures in place, they are better adapted to treating identified challenges such as phishing. To defeat the nascent threat of deepfakes, organizations must repurpose their current work processes to intercept the characteristic warning signals of AI-driven impersonation. Repurposing can include the integration of domain-specific AI applications into existing security infrastructures to verify digital content in real time. It can also involve collaborating with cybersecurity experts and industry consortiums to share new detection techniques and knowledge on deepfake threats.

Meanwhile, our findings also indicate that limited companies utilize distinct detection technology for deepfakes. Although numerous tools like multi-factor authentication or phishing filters offer some degree of protection, they are not particularly designed to identify AI-driven impersonations. Consequently, firms must invest in technologies that will scan images, videos, or voice communications for tampering and work with existing security providers to maintain periodic updates in algorithms for new deepfake tactics. Closing these detection gaps offers a critical additional layer of protection, cutting down on the likelihood that sophisticated synthetic media attacks will go unnoticed.

A further area of vulnerability centers on the limited training most employees receive on deepfakes. Enhancing employees' knowledge involves explaining the existence and potential dangers of deepfakes. It also requires practical training on identifying imitation giveaways, such as graphical errors, awkward vocal cadence, or unlikely environmental details. In order to establish these skills, organizations should conduct deepfake-related training in their training sessions. Establishing a culture of awareness in an organization should reduce the chances of deepfake attacks that are successful.

Even with proactive measures, companies need a clear and efficient incident response plan. This allows quick reactions when a deepfake threat is identified, minimizing potential damage and preserving stakeholder trust.

Our interviews ultimately uncover a significant discrepancy between the way that organizations view deepfake threats and the pace at which AI-powered social engineering is changing. Although conventional security controls and dependency on known vendors are still prevalent, more proactive approaches are crucial in order to keep up with increasingly sophisticated impersonation attacks. By boosting overall awareness, enacting targeted detection tools, partnering with vendors committed to deepfake countermeasures, providing deepfake-specific training, and formalizing incident response plans, businesses can reinforce their defenses against the rapidly evolving deepfake threat. Such recommendations acknowledge the gravity of dealing with deepfakes not as a potential future or niche threat but as an existent and possibly disruptive force in cybersecurity.

10. Future Research Directions

This paper gives valuable insights into corporate preparedness for deepfake-driven social engineering, but it has limitations. The small sample size, of cybersecurity professionals from five organizations, limits the *generalizability* of our findings. Including more diverse industries, like finance or healthcare, could provide other perspectives. Also, our sample may show some bias, as most of the interviewees were from tech or consultancy organizations, where the awareness of digital threats is more commonly higher. Non-tech sectors may have different readiness levels and policies, which were not captured in these data. Despite these limitations, we think the analysis provides meaningful insights into corporate perspectives on deepfake threats. Future research should definitely broaden the scope of the sectors included.

More extensive empirical work on the relationship between organizational policy and technological adoption is warranted. While this paper highlights the role of governance and incident response, future research could go deeper into how policies, both internal and broader regulatory policies, impact the selection, handling, and effectiveness of deepfake detection tools. As legal and ethical considerations around AI-generated media continue to evolve, organizations must navigate the increasing complexities in policy implementation, stakeholder communication, and data protection.

Further investigations into human factors and company culture are also recommended. Understanding how employees interpret or trust suspicious content could help improve training programs focused on deepfake awareness. Performing case studies might deepen the understanding of how individual bias or social dynamics shape the response to potential AI-generated content. Integrating technical, organizational, and human perspectives would allow future research to more comprehensively address how best to defend companies from increasingly sophisticated deepfake threats.

11. Conclusions

In this paper, we have explored the rising threat of deepfake-driven social engineering in corporate environments, which exploits the gaps in traditional cybersecurity frameworks. Despite advancements in the detection methods, such as AI-driven anomaly detection, many organizations remain underprepared to address these threats due to their reliance on generalized cybersecurity measures and lack of tailored solutions. The implications of deepfake attacks are significant, ranging from financial losses to reputational damage and operational disruptions. To mitigate these risks, corporate organizations must adopt specialized detection tools, integrate deepfake prevention into employee training, and develop robust incident response plans. We propose the PREDICT framework as a structured and proactive approach, combining governance, education, detection, and continuous improvement to strengthen corporate defenses. As deepfake technologies become increasingly accessible and sophisticated, corporate organizations must innovate and collaborate to safeguard their assets, trust, and reputation. By *proactively* addressing this emerging challenge, organizations can better prepare for the evolving threat landscape, ensuring their resilience in an increasingly digital and interconnected world.

Author Contributions: Conceptualization: K.T.P., L.P., T.S., M.P., G.C. and N.D. Methodology: K.T.P., L.P., T.S., M.P., G.C. and N.D. Validation: K.T.P., L.P. and T.S. Investigation: K.T.P., L.P. and T.S. Resources: K.T.P., L.P. and T.S. Data curation: K.T.P., L.P. and T.S. Writing—original draft preparation: K.T.P., L.P., T.S., M.P., G.C. and N.D. Writing—review and editing: M.P., G.C. and N.D. Visualization: K.T.P., L.P., T.S., M.P., G.C. and N.D. Supervision: M.P., G.C. and N.D. Project administration: M.P., G.C. and N.D. Funding acquisition: N.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Dataset available on request from the authors.

Conflicts of Interest: The authors declare that there are no conflicts of interest regarding the publication of this work.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
ML	Machine Learning
GAN	Generative Adversarial Network
CNN	Convolutional Neural Network
DNN	Deep Neural Network
MFA	Multi-Factor Authentication
SSO	Single Sign-On
PREDICT	Policies, Readiness, Education, Detection, Incident Response, Continuous Improvement, Testing
CEO	Chief Executive Officer

Appendix A. Interview Questions

Appendix A.1. Introduction and Background

1. Can you please introduce yourself and describe your role within the organization?
2. How familiar are you with deepfake technology and its applications in cybersecurity?

Appendix A.2. The Deepfake Threat Landscape

1. From your perspective, how significant is the threat of deepfake-driven social engineering in today's corporate environment?
2. Can you share any specific instances or case studies where deepfakes were used in social engineering attacks within your organization or industry?
3. Have you had any instances of attacks you prevented or didn't in terms of deepfake?

Appendix A.3. Detection Technologies and Algorithms

1. What technologies or tools does your organization currently use to detect deepfakes?
2. How effective do you find these detection methods in identifying sophisticated deepfake attacks?
3. Are there any machine learning models or AI technologies that you find particularly promising for deepfake detection?
4. What challenges do you face in implementing and maintaining deepfake detection systems?

Appendix A.4. Prevention Measures and Strategies

1. What preventative measures have your organization implemented to mitigate the risk of deepfake-driven social engineering attacks?
2. How do you incorporate multi-factor authentication or media watermarking in your defense strategy?
3. Can you discuss any proactive monitoring strategies your company employs to identify potential deepfake threats?

Appendix A.5. Training and Awareness

1. How does your organization train employees to recognize and respond to deepfake threats?
2. What kinds of awareness programs or campaigns are in place to educate staff about deepfakes and social engineering tactics?
3. How frequently are these training sessions conducted, and how is their effectiveness evaluated?

Appendix A.6. Operational Challenges

1. What operational challenges have you encountered in integrating deepfake detection systems with existing cybersecurity protocols?
2. How scalable are your current deepfake mitigation strategies, especially as deepfake technologies evolve?
3. Can you discuss the cost-effectiveness of the deepfake defense measures your organization has adopted?

Appendix A.7. Future Directions and Innovations

1. What future plans does your organization have for enhancing deepfake detection and prevention capabilities?
2. Are you exploring any emerging technologies, such as blockchain or advanced AI, to bolster your defenses against deepfakes?
3. How do you anticipate deepfake technology evolving, and how is your organization preparing for these changes?

Appendix A.8. Recommendations and Best Practices

1. Based on your experience, what best practices would you recommend to other organizations aiming to defend against deepfake-driven social engineering?
2. What gaps do you see in current deepfake defense strategies that need to be addressed through further research or development?

Appendix A.9. Conclusion and Next Steps

1. Is there anything else you'd like to add that we haven't covered regarding deepfake threats and defenses?
2. Would you be open to participating in future discussions or providing additional insights as our research progresses?

Appendix B. Deepfake Preparedness**Table A1.** Overview of deepfake preparedness in interviewed companies.

Company	Employee Training on Deepfakes	Mitigation Techniques	Detection Techniques	Notes/Other
Company A	Partial	General Security Policies	None Specific	Relies mostly on vendor solutions; planning to integrate specialized tools if deepfake threats become more prominent.
Company B	Partial	General Security Policies	None Specific	Offers broad cybersecurity training but deepfakes are only briefly mentioned. Considering adding dedicated modules.
Company C	None	General Security Policies	None Specific	Focuses on standard phishing and scam awareness; no tailored approach to deepfake-based audio/video manipulation.
Company D	None	Under Discussion	None Specific	Acknowledges the issue but has not yet allocated resources; evaluating detection tools for the next fiscal year.
DTU	N/A	Policies for Educational Use	Limited Research Tools	General guidelines for AI-related research, including deepfakes, but no formal corporate-level strategies.

Appendix C. High-Level Detection Flow

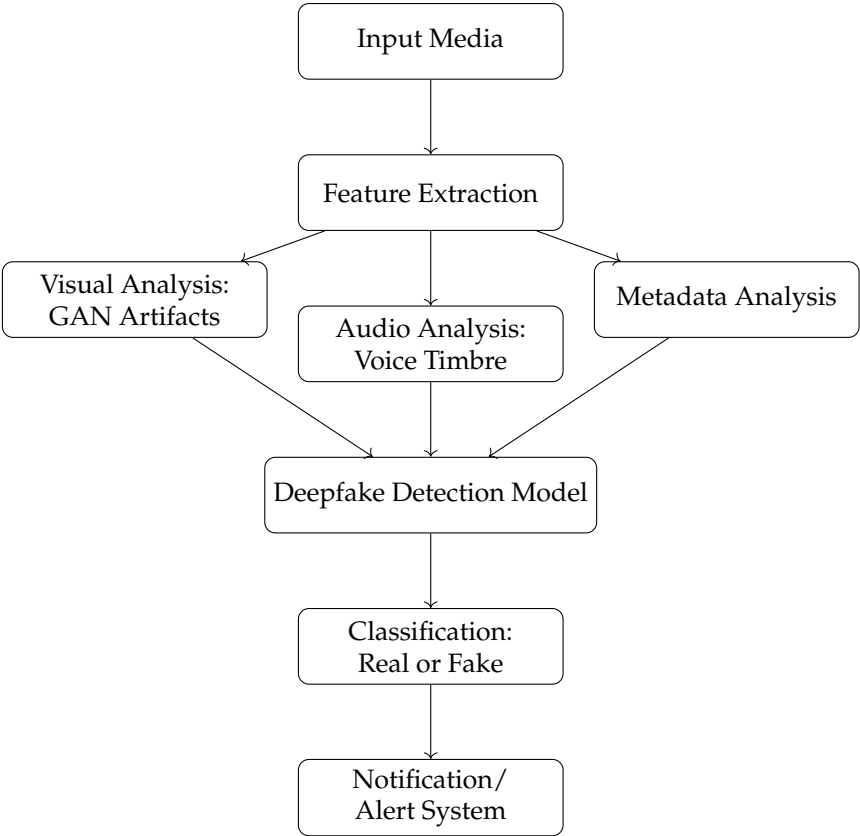


Figure A1. High-level detection flow using machine learning.

Appendix D. Defensive Strategy Model

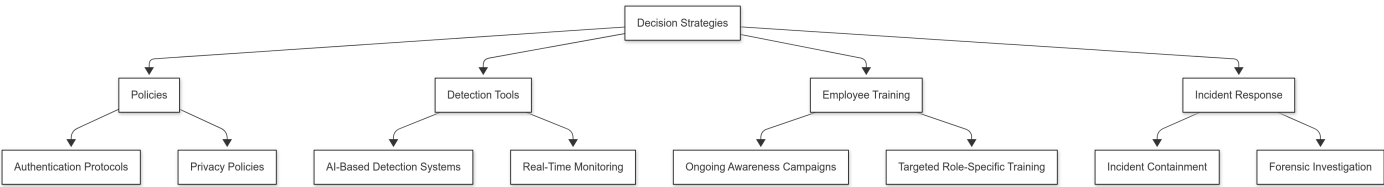


Figure A2. Defensive strategy model.

References

1. National Security Agency; Federal Bureau of Investigation; Cybersecurity and Infrastructure Security Agency; U.S. Department of Defense. Contextualizing Deepfake Threats to Organizations. Technical Report 2023. Available online: <https://media.defense.gov/2023/Sep/12/2003298925/-1/-1/0/CSI-DEEPFAKE-THREATS.PDF> (accessed on 25 November 2024).
2. Ferrara, E. GenAI against humanity: Nefarious applications of generative artificial intelligence and large language models. *J. Comput. Soc. Sci.* **2024**, *7*, 549–569. [CrossRef]
3. Control Risks. How Deepfakes Threaten Organisational Security. *Control Risks*, 30 October 2024. Available online: <https://www.controlrisks.com/our-thinking/insights/how-deepfakes-threaten-organisational-security> (accessed on 30 October 2024).
4. Schmitt, M.; Flechais, I. Digital deception: Generative artificial intelligence in social engineering and phishing. *Artif. Intell. Rev.* **2024**, *57*, 324. [CrossRef]
5. Mirsky, Y.; Lee, W. The creation and detection of deepfakes: A survey. *ACM Comput. Surv.* **2021**, *54*, 1–38. [CrossRef]
6. Tolosana, R.; Vera-Rodriguez, R.; Fierrez, J.; Morales, A.; Ortega-Garcia, J. Deepfakes and beyond: A survey of face manipulation and fake detection. *Inf. Fusion* **2020**, *64*, 131–148. [CrossRef]

7. Lynn, J. Addressing Deepfake-Enabled Attacks Using Security Controls. *SANS Whitepaper*, 14 September 2022. Available online: <https://www.sans.edu/cyber-research/addressing-deepfake-enabled-attacks-using-security-controls/> (accessed on 15 November 2024).
8. Korshunov, P.; Marcel, S. Deepfakes: A new threat to face recognition? assessment and detection. *arXiv* **2018**, arXiv:1812.08685.
9. Gambín, Á.F.; Yazidi, A.; Vasilakos, A.; Haugerud, H.; Djenouri, Y. Deepfakes: Current and future trends. *Artif. Intell. Rev.* **2024**, *57*, 64. [CrossRef]
10. Nguyen, T.; Nguyen, C.M.; Nguyen, T.; Duc, T.; Nahavandi, S. Deep learning for deepfakes creation and detection: A survey. *Comput. Vis. Image Underst.* **2022**, *223*, 103525. [CrossRef]
11. Pashine, S.; Mandiya, S.; Gupta, P.; Sheikh, R. Deep fake detection: Survey of facial manipulation detection solutions. *arXiv* **2021**, arXiv:2106.12605.
12. Naitali, A.; Ridouani, M.; Salahdine, F.; Kaabouch, N. Deepfake attacks: Generation, detection, datasets, challenges, and research directions. *Computers* **2023**, *12*, 216. [CrossRef]
13. Dhesi, S.; Fontes, L.; Machado, P.; Ihianle, I.K.; Fassihi Tash, F.; Adama, D.A. Mitigating adversarial attacks in deepfake detection: An exploration of perturbation and AI techniques. *arXiv* **2023**, arXiv:2302.11704.
14. Chesney, R.; Citron, D. Deep fakes: A looming challenge for privacy, democracy, and national security. *Calif. Law Rev.* **2019**, *109*, 1753. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3213954 (accessed on 3 December 2024). [CrossRef]
15. Khan, R.; Sohail, M.; Usman, I.; Sandhu, M.; Raza, M.; Yaqub, M.A.; Liotta, A. Comparative study of deep learning techniques for deepfake video detection. *ICT Express* **2024**, *10*, 1226–1239. [CrossRef]
16. Jeganathan, B. Exploring the Power of Generative Adversarial Networks (GANs) for Image Generation: A Case Study on the MNIST Dataset. *Int. J. Adv. Eng. Manag.* **2025**, *7*, 21. [CrossRef]
17. Lawton, G. Generative Models: VAEs, GANs, Diffusion, Transformers, NeRFs. *TechTarget*, 8 July 2024. Available online: <https://www.techtarget.com/searchenterpriseai/tip/Generative-models-VAEs-GANs-diffusion-transformers-NeRFs> (accessed on 5 December 2024).
18. Suci, P. Putin's Deepfake Doppelganger Highlights the Danger of the Technology. *Forbes*, 15 December 2023. Available online: <https://www.forbes.com/sites/petersuci/2023/12/15/putins-deepfake-doppelganger-highlights-the-danger-of-the-technology/> (accessed on 13 March 2025).
19. Vasilyeva, N. Deepfake Video of Putin Declaring Mobilisation Broadcast on State TV. *The Telegraph*, 5 June 2023. Available online: <https://www.telegraph.co.uk/world-news/2023/06/05/fake-video-putin-mobilisation-russia-broadcast-state-tv/> (accessed on 15 November 2024).
20. Damiani, J. A Voice Deepfake Was Used To Scam A CEO Out Of \$243,000. *Forbes*, 3 September 2019. Available online: <https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/> (accessed on 14 March 2025).
21. Chen, H.; Magramo, K. Finance Worker Pays Out \$25 Million After Video Call with Deepfake “Chief Financial Officer”. *CNN*, 4 February 2024. Available online: <https://edition.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html> (accessed on 14 March 2025).
22. Braun, V.; Clarke, V. Using thematic analysis in psychology. *Qual. Res. Psychol.* **2006**, *3*, 77–101. [CrossRef]
23. Carnegie Mellon University, Information Security Office. Social Engineering. *Don't Take the Bait*, n.d. Available online: <https://www.cmu.edu/iso/aware/dont-take-the-bait/social-engineering.html> (accessed on 10 December 2024).
24. Kietzmann, J.; Lee, L.W.; McCarthy, I.P.; Kietzmann, T.C. Deepfakes: Trick or treat? *Bus. Horizons* **2020**, *63*, 135–146. [CrossRef]
25. Doss, C.; Mondschein, J.; Shu, D.; Wolfson, T.; Kopecky, D.; Fitton-Kane, V.A.; Bush, L.; Tucker, C. Deepfakes and scientific knowledge dissemination. *Sci. Rep.* **2023**, *13*, 13429. [CrossRef] [PubMed]
26. TG8 Security. Breaking Free: The Drawbacks of Embracing a Single Cybersecurity Vendor. Available online: <https://web.archive.org/web/20240418000244/https://www.tg8security.com/breaking-free-the-drawbacks-of-embracing-a-single-cybersecurity-vendor/> (accessed on 20 November 2024).
27. Yasser, B.; Hani, J.; Amgad, O.; Ahmed, N.; Amr, H.; Salah, M.; El-gayar, S.; Ebied, H.M. Deepfake Detection Using EfficientNet and XceptionNet. In Proceedings of the 2023 IEEE Eleventh International Conference on Intelligent Computing and Information Systems (ICICIS), Cairo, Egypt, 21–23 November 2023; pp. 598–603. [CrossRef]
28. Sun, Z.; Han, Y.; Hua, Z.; Ruan, N.; Jia, W. Improving the Efficiency and Robustness of Deepfakes Detection through Precise Geometric Features. *arXiv* **2021**, arXiv:2104.04480.
29. Mifsud, E.; Massa, P. Deepfakes and the Law: Are We Protected? *Lexology*, 28 August 2023. Available online: <https://www.lexology.com/library/detail.aspx?g=6de74674-e2d5-4e0c-9e26-33f4d6aadf40> (accessed on 1 April 2025).
30. Yerushalmi, D. Embracing a Zero-Trust Mindset to Combat Deepfakes in Identity Verification. *Forbes*, 30 September 2024. Available online: <https://www.forbes.com/councils/forbestechcouncil/2024/09/30/embracing-a-zero-trust-mindset-to-combat-deepfakes-in-identity-verification/> (accessed on 2 April 2025).

31. Zhou, S.; Yuan, S.; Xu, J. Evolutions and trends of artificial intelligence (AI): Research, output, influence and competition. *Library Hi Tech* **2021**, *40*, 704–724. [[CrossRef](#)]
32. Millière, R. Deep Learning and Synthetic Media. *Synthese* **2022**, *200*, 231. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.