How Interactions Between Deep Fake Tools and Organizational Protocols Lead to Instability in Corporate Security Networks

STS Research Paper

Presented to the Faculty of the

School of Engineering and Applied Science

University of Virginia

By

Rhea Agarwal

Feb 26 , 2025

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

ADVISOR

Benjamin J. Laugelli, Assistant Professor, Department of Engineering and Society

Introduction

In recent years, deep fake technology has been a growing threat to corporate security, enabling significant forms of fraud (Hubbard, 2025). One such instance was in 2019 when fraudsters created deep fake audios impersonating the CEO of a UK-based firm and successfully transferred $230,000 (Stupp, 2019). This case raises issues about the weakness of security networks. Scholars and cybersecurity professionals see deep fake fraud as a misuse of trust-based systems. Existing research focuses on technical detection methods but often overlooks how interactions between human and non-human actors create vulnerabilities in corporate networks. This narrow view ignores the complex dynamics between technology, decision-makers, and corporate protocols that enable real-world fraud. Without examining these interactions, companies risk implementing ineffective security measures that fail to address the core problem.

A deeper analysis of how human and non-human actors interact is essential to identify and mitigate the systems' vulnerabilities, without which security systems will always lag behind AI threats. I argue that the fragility of sociotechnical networks depends on both human and non-human actors functioning as intended. In the 2019 UK CEO deep fake fraud case, the vulnerability of the corporate security network stemmed from the unchecked agency of non-human actors and the failure of critical control points that were supposed to keep the security network secure. Using Actor-Network Theory (ANT), a framework that maps how human and non-human actors interact in network settings, this paper analyzes how interactions between deep fake tools, corporate decision-makers, and organizational protocols destabilized the network, leading to its failure. It also contributes to a deeper understanding of how socio-technical networks, specifically in the scenario of a financial institution, are constructed and manipulated, offering insight into the vulnerabilities within trust-based verification systems.

This analysis uses primary sources, including a regulatory advisory from the FCC and corporate responses like Accenture's Investment in deep fake detection technologies, to understand how control points in corporate security networks failed in the 2019 fraud case.

Literature review

A number of scholars have highlighted the impacts of deep fake fraud on financial systems, specifically focusing on the vulnerabilities in financial institutions' authentication processes. However, few have used case-specific analysis to understand how this applies to banks in the real world. For example, in a paper titled "Analysis of Deep Fake Problems for Banks and Financial Institutions", Chief Technical Officer Pavlo Sidelov (Sidelov, 2022) emphasizes the role of the exploitation of trust in fraud but does not explore the role of the interplay between human decision making and specific organizational protocols in facilitating fraudulent attacks. He also argues that modern fraud techniques use deep fake technology in order to manipulate systems based primarily on trust, particularly in financial institutions. He does a good job in explaining what makes financial institutions' systems more vulnerable to fraudulent attacks by connecting it back to the idea that these systems are built on trust. By doing this, he emphasizes the insufficiency of traditional fraud prevention measures, bringing to light the fact that banks need to work on improving their security systems in order to reduce their vulnerabilities since technological advancements like deep fake tools are able to bypass existing measures. This work provides an excellent foundation for identifying vulnerabilities. However, it does not analyze the socio-technical networks and interactions that are likely leading to these vulnerabilities as well as the technicality of detecting deep fake voices.

In terms of understanding deep fake voice detection methods, Zahra Khanjani, Babrielle Watson, and Vandana P Janeja do a great job (Khanjani et al., 2023). They focus heavily on

technological development and detection tools for deep fakes, something that was not covered by Sidelov. The study provides some attention to how these detection technologies interact with organizational protocols and decision making processes in the real world, but does so in a more general way, without discussing specific cases. It also does not address the destabilizing impact that deep fake audio tools have on corporate security networks. This source does an excellent job in offering a detailed review of audio deep fake technology, giving the readers a strong background. It goes into the applications of these technologies and the current challenges in AI-generated voice detection. It also highlights the increasing accessibility of voice cloning tools which ties in with the overall problem statement associated with my research. The paper contributes to the understanding of the technical aspects of deep fake tools, but lacks a discussion about case specific applications, which I hope to achieve in my paper. In my research, I will build on this by integrating the technical aspects with an analysis of human and organizational actors, by specifically looking into the 2019 UK fraud case (Stupp, 2019).

To summarize, current scholars fail to adequately cover the insufficiency of current security systems and the technicality of voice detection tools. They also fail to analyze a specific case and understand how the interactions between technological tools, human actors, and organizational systems create the vulnerabilities in the security networks. Although both studies offer valuable insights into the technological and organizational aspects of deep fake fraud, they lack these dimensions and case-specific analysis. My research aims to fill this case by using Actor-Network Theory to examine the 2019 UK CEO fraud case.

Conceptual Framework

My analysis of the 2019 CEO deep fake fraud case in the UK draws upon actor network theory, which allows me to explore how the interactions between deep fake audio tools (non

human actors) , corporate decision makers (human actors), and organizational protocols (non human actors) contribute to the instability of the corporate security network involved. Actor network theory shifts the focus from individual actions to networks of relations and allows the analysis of both human and non-human actors in shaping outcomes. It also helps in highlighting the fragility or stability of networks based on these interactions.

Developed by Bruno Latour, Michel Callon, and John law, Actor Network Theory (Latour, 2005) emphasizes the interconnectedness of human and non-human actors in forming networks. Networks are formed by network builders who assemble as human and non-human actors into a network to solve a problem or accomplish a goal. The actors, or actants, can be people, technologies, documents, or protocols - anything that has an influence on the network. The networks in question are dynamic. They are constantly shaped and reshaped through the interactions between the actors. An actor is any entity that influences the network's behavior. In this case, deep fake audio tools act as non-human actors that directly impact decision making. A translation is a process by which the actors involved align their interests to stabilize the network. An obligatory passage point (OPP) is a critical juncture that all actors must pass through to maintain network stability. For example the bank's voice certification system acts as an OPP, and its failure allowed the fraud to succeed. The network builders discussed earlier are usually identified as the network's OPP. By establishing themselves as the OPP, they secure their influence over the other actors so that they work together to advance their network's goals. Finally, the network fragility refers to points where weak links or misalignments can cause breakdowns in the system (e.g., security vulnerabilities). What ANT does not do, is moralize or assign blame to specific actors, but instead it maps the interactions between them.

In what follows, I will use Latour's ANT to analyze how the actors involved in the CEO bank fraud case (deep fake audio tools, corporate decision makers, and organizational protocols) interact and where the points of instability occur. I will do this by examining how the non-human actors disrupt the network's balance leading to the vulnerabilities, and will explore the translation process - how scammers used deep fake tools, in this case, to align the actors' interests toward the fraudulent goal. Finally, I will identify OPPs in the voice verification systems and understand how they were bypassed.

Drawing on ANT, I will first map the key human and non-human actors involved in the 2019 UK CEO deep fake fraud case, focusing on how their interactions formed points of instability. I will then analyze how non-human actors, like deep fake audio tools, disrupted the network's balance, leading to security vulnerabilities. Finally, I will examine how scammers bypassed the critical control points, such as voice verification systems, to successfully execute the fraud.

Analysis

The fragility of sociotechnical networks becomes visible when malicious actors successfully exploit both technological tools and human trust to impact organizational systems. In order to analyze how a similar breakdown took place in the 2019 UK CEO fraud case, this section begins by mapping the network involved.

In this case, the network builders were the fraudsters that created the deep fakes, whose goal was to bypass the company's authentication system and deceive the company executive into sending them money. These fraudsters used both human actors (including the executive who was deceived into authorizing the fraudulent transaction), and non-human actors (including the AI voice cloning tools and corporate voice verification protocols). They successfully constructed a

network in which these actors all came together toward a deceptive goal, allowing the deep fake voice to be accepted as authentic by the verification system.. This paper focuses on the successful fraudulent network created by the fraudsters, focusing on how the system overcame the company's verification system. While the UK CEO deep fake fraud case is the focus of this paper, this section uses 2 parallel cases (one by CNN and another by The Guardian) to aid this analysis and show how similar dynamics appear in multiple contexts. They also help clarify how deep fake tools function as active agents within sociotechnical systems, and how their unchecked agency contributes to system failures.

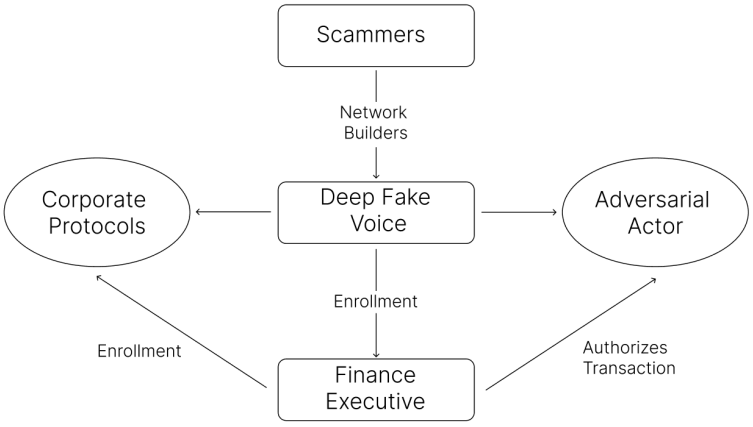

Figure 1: Network Mapping

Analysis 1: Deep fake Tools as Active Manipulators – Influencing Decisions and Evading Security

The deep fake audio tools used in the 2019 UK CEO fraud case functioned as influential actors within the network. They directly influenced the human decision makers and evaded security protocols. Their unchecked agency allowed them to manipulate human trust and bypass

the company's verification methods, which in turn created vulnerabilities within the security system.

A CNN report titled "Finance Worker Pays Out $25 Million After Video Call with Deep Fake CEO" (Chen & Magramo, 2024) details an incident, similar to the case I discuss in this paper, in which a finance employee was deceived by a deep fake video call, leading to substantial financial losses. It is important to note that in this incident, the scammers didn't just use a single fake voice. Instead, they replicated multiple familiar staff members, creating the illusion of a routine group call. This complex imitation, combined with the realistic timing and visuals of the call, provided no reason for the employee to be suspicious. The worker, unaware of the presence of deep fake video on the call, trusted that he was on a video call with his colleagues, highlighting the deep fake technology's ability to convincingly mimic human interactions. Note that the scammers did not just replicate the voice of the CEO, but instead recreated the entire staff using deep fake videos, forming a complex system of non-human actors that convincingly stimulated a regular workplace interaction (video call between colleagues). This strongly suggests that the deep fake tool was able to function as an agent within the network, and manipulate the human actor involved into transferring the money. Reliance on things like voice familiarity and visual consistency exposes a critical weakness in corporate security networks. In terms of ANT, the employee's trust in the video call acted as a weak link, making the network vulnerable to manipulation. The deep fake videos exploited this trust in order to bypass the system, revealing that human actors can themselves become unstable nodes in a network. According to ANT, non-human actors can have equal influence in a network, and the realistic portrayal of the non-human actors, in this case, emphasize the blurred boundaries between human and machine, which then lead to the breakdown of the company's security protocol.

Moreover, the employee not questioning the authenticity of the deep fake highlights further the fragile nature of the corporate security networks when trust is placed in surface level verification. Connecting this back to ANT, the deep fake videos did not just passively exist in the network, but instead actively played a role in reshaping it by causing a temporary reorganization in which the deep fake version of the CEO became an important force in driving decision making. This demonstrates how non-human actors are able to disrupt established hierarchies within a corporate network. The deep fake didn't just mimic the CEO, but effectively replaced a human actor's role in the decision-making process. This highlights how quickly power can shift within human-technology systems when security measures break down.

A report by The Guardian (Hern, 2024) also helps support this idea. In 2024, they detailed a deep fake scam that targeted Mark Read, the CEO of WPP, the world's largest advertising firm. The fraudsters created a fake WhatsApp account using AI generated voices as well as existing YouTube videos of Read, which enabled them to stage a virtual meeting. The goal was to make the participants of the video call believe that they were interacting with the CEO. Similar to the source analyzed above, this one also showcases the power of deep fake technology in infiltrating corporate communication networks. The report notes that "During the meeting, the impostors deployed a voice clone of the executive as well as YouTube footage of them" (Hern, 2024, para. 2). This illustrates how the fraudsters were able to leverage human tendencies to trust familiar visuals in order to bypass suspicion from the other attendees' end. In this situation, what's important to note is that the scammers did not rely solely on the deep fake content, but also added publicly available material as an actor into their network. This blend of authentic and cloned material made it more difficult to detect fraud. The use of these easily accessible YouTube videos further emphasizes the vulnerability created by digital artifacts. In

ANT, these videos became part of the network, strengthening the effectiveness of the deep fakes. This highlights how public data, something that is often overlooked in security measures, can be used to improve the believability of synthetic media sources, allowing scammers to get through security systems more easily. Moreover, the use of a commonly used workplace tool like WhatsApp added to the legitimacy of this call. In this case, the deep fake did not just impersonate the CEO, but instead functioned as an actor that with the help of real footage, increased its believability, making it a more influential actor overall. The fact that videos of the CEO that were so readily available on the internet were enough to create this manipulation proves the critical vulnerabilities within corporate networks.

Both the CNN and The Guardian reports show how the deep fake tools used were able to control human decision makers and bypass security checks. In the CNN case, the deep fake technology recreated an entire staff, creating a convincing video call that led the finance employee to authorize a fraudulent transaction, without knowing. The video generated using the deep fake tools acted as an active non-human participant and directly shaped the human actor's (finance employee) decisions and exploited the network's reliance on surface-level verification. Similarly, the case in The Guardian's report reveals how an already influential non-human actor is able to amplify its influence by combining readily available, authentic content with the deep fake content. Together, these cases highlight a critical vulnerability in the finance space: deep fake tools have the ability to integrate into trusted communication systems in a way that can manipulate human actors. Their unchecked agency allowed them to bypass the security measures involved, and be a driving force in the decision making process.

Analysis 2: The Breakdown of Security – How Voice Verification System Failures Enabled Deep fake Fraud

In the 2019 UK CEO deep fake fraud case, the voice verification system functioned as an adversarial actor that was expected to support secure communication but ultimately became a vulnerability within the network. The system's inability to detect the deep fake audio used in the fraud allowed the scammers to bypass the authentication system, and led to a breakdown in the company's security.

The Federal Communications Commission (FCC) issued a warning (Federal Communications Commission, n.d.) for financial institutions about the growing threat of deep fake audio and video scams. The report highlights that deep fake technology has become advanced enough to bypass traditional voice verification systems, creating vulnerabilities in corporate security networks. It notes that "Also known as voice cloning, these technologies emulate recognizable voices for robocalls to consumers and are often used in imposter scams that spread misinformation, endorse products, or steal money and personal information" (Federal Communications Commission, n.d., para. 2) This warning emphasizes the need for stronger authentication systems to prevent scammers from exploiting companies. Note also, that the FCC advisory specifically points out the ability of deep fake tools to replicate vocal cues, which voice verification systems are programmed to recognize. This shows that deep fakes not only bypass the authentication system but also reveal the system's inability to detect non-human actors. According to ANT, in the 2019 CEO fraud case, the voice verification system should have acted as a gatekeeper, but its inability to differentiate between human voice and AI generated voice allowed the unauthorized actor to pass through. This failure in identifying a deep fake voice weakened the network, and created a clear pathway for scammers.

Accenture's press release, in October 2024 (Accenture, 2024), announcing its investment in Reality Defender reflects the industry's acknowledgement that deep fake technology is a

critical security threat. The company emphasized the need for advanced detection tools to prevent deep fakes from bypassing current security checkpoints, specifically voice verification systems. The press release suggests that deep fakes have reached a level of sophistication that renders conventional security measures ineffective. It states that "Reality Defender provides detection against advanced threats posed by deepfakes and AI-generated content" (Accenture, 2024, para. 2) . The fact that leading firms are seeking more advanced tools highlights the inadequacy of traditional voice verification systems as a security checkpoint. The voice verification systems are meant to protect the company from fraud. However, Accenture's acknowledgement of their own vulnerability shows that these verification systems are no longer working as intended, allowing non-human actors to exploit gaps in the network.

In the previous paragraph, I argue that the voice verification systems are meant to protect the company, suggesting that the deep fake fraud in the 2019 UK CEO fraud case was a result of a failure in the company's voice verification system. However, it can be argued that the fraud was not a result of the failed verification system, but instead a lapse in human judgement. From this perspective, it is possible to say that the executive that transferred the money making the fraud successful should have followed internal protocols or taken additional steps to ensure authenticity before allowing such a large transaction. However, this argument puts too much responsibility on the human actor and does not recognize the role of the non human actor in this situation. A study by the University of Florida (Schlenker, 2024) explains that people's ability to detect deep fake audio, when created well, is only a little better than chance, even if they are aware of the possibility of it being a deception. This supports the idea that relying solely on human judgement is not enough when it comes to detecting such sophisticated deep fakes. Therefore, it is important to recognize that the absence of robust enough verification systems,

rather than a human error in judgement, played a major role in the success of the fraudsters' system.

Both the FCC advisory and Accenture's press release support the idea of the growing failure of voice verification systems as critical control points in corporate security networks. The FCC reflects vulnerability by intensifying how deep fake audio can convincingly mimic human voices, allowing scammers to bypass voice authentication protocols. Similarly, Accenture's investment in Reality Defender serves as a reinforcement that outdated verification systems leave networks exposed to deep fake manipulation. It highlights the need for better security systems that can adapt to the capabilities of modern deep fakes. In both the sources, the failure of voice verification enabled non-human actors to bypass security protocols and infiltrate the network.

Conclusion

The 2019 UK CEO deep fake fraud case highlights a critical question: Why is the corporate security network vulnerable to deception, and how do interactions between deep fake audio tools, corporate decision-makers, and organizational protocols contribute to its instability? In this paper, I argue that the vulnerability of the corporate security network in this case stemmed from the unchecked agency of non-human actors and the failure of critical control points that were supposed to secure the network. The AI cloning tools acted as active agents, manipulating the human decision-makers and bypassing security protocols, and the failure of the voice verification system allowed the fraudsters to succeed, breaking the network. This analysis emphasizes the complex interactions within socio-technical networks. Ignoring these dynamics leads to weak security protocols that fail to address the systems' vulnerabilities. Using Actor-Network theory, this research highlights the need for stronger security strategies, to keep pace with growing AI threats.

Beyond this single case, the findings of this paper highlight a broader concern for corporate security. More broadly, this case advances our understanding of how malicious actors can strategically use both human and non-human actors in order to achieve their deceptive goals. By using ANT, the analysis of this case further highlights how non-human actors are not passive but actively participate in networks. These insights contribute to STS research by expanding on the overall understanding of how we conceptualize agency, trust, and failure in complex networks.

Word Count: 3640

References

Accenture. (2024, October 22). Accenture invests in Reality Defender to help fight deepfake extortion, fraud and disinformation. Accenture Newsroom. https://newsroom.accenture.com/news/2024/accenture-invests-in-reality-defender-to-help-fight-deepfake-extortion-fraud-and-disinformation

Chen, H., & Magramo, K. (2024, February 4). Finance worker pays out $25 million after video call with deepfake 'chief financial officer'. CNN. https://www.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html

Federal Communications Commission. (n.d.). Deep-fake audio and video links make robocalls and scam texts harder to spot. Federal Communications Commission.

https://www.fcc.gov/consumers/guides/deep-fake-audio-and-video-links-make-robocalls-and-scam-texts-harder-spot

Hern, A. (2024, May 10). CEO of world's biggest ad firm targeted by deepfake scam. The Guardian. https://www.theguardian.com/technology/article/2024/may/10/ceo-wpp-deepfake-scam

Hubbard, L. (2025, January). The rising threat of deepfake fraud. Business Credit Magazine. https://bcm.nacm.org/the-rising-threat-of-deepfake-fraud/

Kaur, A., Hoshyar, A. N., Saikrishna, V., Firmin, S., & Xia, F. (2024). Deepfake video detection: Challenges and opportunities. Artificial Intelligence Review, 57(3), 1–27. https://doi.org/10.1007/s10462-024-10810-6

Khanjani, Z., Watson, G., & Janeja, V. P. (2023). Audio deepfakes: A survey. Frontiers in Big Data, 5, Article 1001063. https://doi.org/10.3389/fdata.2022.1001063

Latour, B. (2005). Reassembling the social: An introduction to actor-network-theory. Oxford University Press.

Schlenker, D. (2024, November 15). *Listen carefully: UF study could lead to better deepfake detection*. University of Florida News. https://news.ufl.edu/2024/11/deepfakes-audio/

Sidelov, P. (2022). Analysis of deepfakes problem for banks and financial institutions série "Vyšetřování bankovních podvodů". Věda a Perspektivy, 3(10), 97–108. https://doi.org/10.52058/2695-1592-2022-3(10)-97-108

Stupp, C. (2019, September 4). Fraudsters used AI to mimic CEO's voice in unusual cybercrime case. The Wall Street Journal. https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402