UVA and Greenway Solutions AI Fraud Detection

Sociotechnical impact of fraud within Commercial Banking

A Thesis Prospectus
In STS 4500
Presented to
The Faculty of the
School of Engineering and Applied Science
University of Virginia
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Systems Engineering

By
Vishnu Lakshmanan

November 9, 2024

Rhea Agarwal, Drake Ferri, Baani Kaur, Vishnu Lakshmanan, Padma Lim, Fahima Mysha

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.

ADVISORS

Prof. Pedro Augusto P. Francisco, Department of Engineering and Society

Prof. Gregory J. Gerling, Department of Systems and Information Engineering

Introduction

The topic I am researching is fraud detection for banks using artificial intelligence (AI) technologies. This research aims to explore how banks can identify fraudsters and protect their data amidst increasing cybersecurity risks posed by AI. As AI capabilities grow, modern businesses must understand both the potential risks and opportunities these tools present for day-to-day operations, particularly in financial institutions where security is critical.

My Capstone Project will focus on developing AI-based solutions to improve fraud detection systems in banking institutions. Specifically, I will investigate techniques that fraudsters use to impersonate humans, such as voice spoofing, and develop systems capable of distinguishing between real and AI-generated voices. The technical aspect of this project includes analyzing features like meaningless silence, background noise, and other factors that impact a voice to clone a voice that is capable of passing bank security measures. This solution aims to counter the increasing availability of spoofing tools on the internet, which fraudsters use to bypass security measures like voice authentication.

On the sociotechnical side, my STS research will investigate the societal impacts of AI-driven fraud detection systems. This will include an examination of how fraud affects consumers and businesses on a personal level and how fraudsters may exploit vulnerabilities, such as attempting to access accounts of deceased individuals.

Additionally, I will explore the ethical concerns associated with AI-based fraud detection, including privacy, consumer trust, and the broader social implications of using AI in financial security.

The connection between my Capstone Project and STS Research lies in the interplay between technical and societal concerns. While AI-based fraud detection technologies promise to enhance security, they must also address ethical issues such as bias, consumer trust, and ensuring fairness. Both projects highlight the importance of developing robust technical solutions that are also socially responsible. This study aims to create a holistic fraud detection solution that not only mitigates fraud but also ensures the ethical deployment of AI in financial systems, balancing overall security.

AI Voice Detection and Fraud Measures

The problem that I am addressing is voice spoofing with audio deepfakes in the banking industry. With several commercially available and open-source tools available, fraudsters are able to create voices that aim to bypass security measures that have been put in place by banks. Some of the commercial tools available allow users to sign up and create cloned voices in minutes (Inamdar et. al, 2024). A user has to sign up and provide a short sample ranging from 15 seconds to a minute. After the software is able to analyze the voice from the recording or file, it lets the user know that it's ready for use. From here, the user is either able to use a text-to-speech or a speech-to-speech feature that will produce output. These tools make it increasingly easy for attackers to bypass voice authentication systems, as discussed by Kassis and

Hengartner (2023), who highlight that current methods, while effective in certain cases, struggle with issues such as background noise and distortions that fraudsters exploit.

The purpose of the project is to understand where the major clients, banks, have weaknesses in their systems. This will be achieved through several small steps, which culminate in testing against bank software to understand weaknesses. Firstly, the important factors that make up a voice will be studied. Initial factors such as pitch, background noise, and tone of voice have been identified. Once all the factors have been determined, a small library of voices will be created with the use of commercially available software such as ElevenLabs and FineVoice. Both tools are open to use by the public and have more premium versions that allow for more aspects of a generated voice to be tweaked and longer samples to be uploaded to the software, potentially resulting in a better output.

This small library of voices will be shown to a sample of strangers, acquaintances, and friends to classify the voices. The two questions this small study aims to answer are whether the voice shown is real or not real and whether it sounds like the person being cloned or whether it does not sound like the subject. The former question is more heavily weighted as current bank verification systems surround liveliness checks. Huang et al. (2021) highlight that voice authentication systems integrating Automatic Speaker Verification (ASV) are highly vulnerable to spoofing attacks. They found that attackers could easily bypass systems, especially when systems are not designed with robust countermeasures like liveliness checks.

These checks pose the question of whether the voice that an agent or machine is hearing is alive or not, and will further allow an understanding of what factors are deemed to be more important than others when evaluating a voice. Once results are yielded from this small-scale experiment, a larger library of voices will be built out and run through a verification system referred to as an ASV. An ASV's primary purpose is to verify the identity of a speaker based on their voice, comparing a person's voice to a pre-recorded voiceprint to determine if the speaker is who they claim to be (Kassis & Hengartner, 2023).

Based on these results, the library of voices will be refined to adjust for the strengths and weaknesses that the system determined. Djjalali (2023) emphasizes the differences between AI-generated voices and human voices, noting that while AI-generated voices can be highly convincing, they often lack certain human nuances that make them detectable. This insight could be crucial in refining the voice library and enhancing the system's ability to differentiate between real and fake voices. Afterward, the library will be tested against client software. This should allow all parties involved to understand the weaknesses that a system may have and work to fix it. Moreover, as the study by Blue and Traynor (n.d.) demonstrates, even sophisticated AI-generated deepfake voices tend to show certain tell-tale signs that can be exploited by detection systems. By applying such insights, this project aims to enhance fraud detection in the banking industry.

Sociotechnical impact of fraud within Commercial Banking

To frame the problem from an STS perspective, I will examine how AI-driven fraud detection systems impact individuals and society at large. From an STS standpoint, it is essential to understand the complex interactions between technology, ethics, and social consequences. My research will focus not only on the technical efficiency of these systems but also on the broader ethical and social implications, such as privacy concerns, trust issues, and the potential exploitation by fraudsters. I believe that a detailed analysis of these aspects is necessary to understand the full scope of AI's role in financial security.

The central research question guiding my investigation is: How do AI-driven fraud detection systems impact consumer privacy, trust, and ethical practices in financial security? This question is crucial because AI has enhanced security and operational efficiency in financial systems (Boukherouaa, et. al., 2021) but its use raises significant ethical concerns. Issues surrounding consumer privacy are especially important when AI systems process sensitive data to detect fraud. Additionally, the growing skepticism among consumers regarding data management practices makes this topic timely and relevant (Barr-Pulliam, et. al, 2023). Exploring these issues is essential to ensure that AI technology improves security without undermining consumer rights or perpetuating societal harm.

The importance of my research question is further justified by the increasing sophistication of fraud tactics. AI-driven fraud detection systems, while beneficial, are not immune to exploitation. Fraudsters continuously look for ways to bypass security

measures, often targeting vulnerable individuals or systems (Hilal, et. al, 2021). Understanding these vulnerabilities and the broader social impact of AI in fraud detection is critical for developing more ethical, robust, and socially responsible systems.

My research will also explore the development and history of AI-driven fraud detection systems. AI, particularly machine learning algorithms, is designed to detect patterns indicative of fraudulent activity, allowing for rapid responses to potential threats. However, as this technology has evolved, it has introduced new ethical challenges. AI systems typically rely on vast datasets, including personal and financial information, which could be misused if not carefully regulated.

In my analysis, I will look at the ethical and societal dimensions of AI-driven fraud detection using a qualitative approach. I plan to use case studies and literature reviews on AI ethics and fraud prevention to explore the perceptions of consumers and businesses about these systems. I will focus on how they address the ethical concerns that arise and the steps taken to mitigate risks. Additionally, I will investigate specific cases where AI-based systems have failed or been exploited by fraudsters, offering insight into system vulnerabilities.

To gather evidence, I will collect qualitative data from case studies and a review of relevant literature. I will also look at performance data from AI systems, focusing on the rates of false positives and negatives in fraud detection. This evidence will allow me to assess the accuracy and reliability of these systems, and I will interpret the findings to understand how these technologies impact consumer trust, privacy, and security from

both a practical and ethical standpoint. Through this analysis, I aim to identify gaps in current regulatory frameworks and recommend changes to ensure AI-driven fraud detection systems safeguard consumers without compromising ethical standards or social values.

Conclusion

The overall technical deliverable of the technical research should be a recommendation or set of recommendations that banks should follow and implement to strengthen their verification and security processes. Another report will be generated to highlight the STS aspect of the technical project. Both the recommendation and report will allow readers to understand the impact of AI in the commercial banking industry while also allowing the banks to work on developing their software and practices.

References:

1. Kassis, A., & Urs Hengartner. (2023). *Breaking Security-Critical Voice Authentication*. https://doi.org/10.1109/sp46215.2023.10179374
2. Huang, K.-L., Duan, S.-F., & Lyu, X. (2021). Affective Voice Interaction and Artificial Intelligence: A Research Study on the Acoustic Features of Gender and the Emotional States of the PAD Model. Frontiers in Psychology, 12. https://doi.org/10.3389/fpsyg.2021.664925
3. djjalali. (2023, April 12). How AI Voices Differ from Natural Voices. Voices. https://www.voices.com/blog/ai-vs-natural-voice/
4. Our brains respond differently to human and AI-generated speech, but we still struggle to tell them apart. (2024, June 24). EurekAlert! https://www.eurekalert.org/news-releases/1048879
5. Hai, X., Liu, X., Tan, Y., & Zhou, Q. (2023). SiFDetectCracker: An Adversarial Attack Against Fake Voice Detection Based on Speaker-Irrelative Features. https://doi.org/10.1145/3581783.3613841
6. Lalchand, S., Srinivas, V., Maggiore, B., & Henderson, J. (2024, May 29). Generative AI Is Expected to Magnify the Risk of Deepfakes and Other Fraud in Banking. Deloitte Insights; Deloitte. https://www2.deloitte.com/us/en/insights/industry/financial-services/financial-services-industry-predictions/2024/deepfake-banking-fraud-risk-on-the-rise.html
7. Techniques that Help Prevent Deepfakes in Banking. (2024, May 31). Infopulse. https://www.infopulse.com/blog/deepfake-detection-techniques-banking
8. Noyes, J. (2002). *Designing for Humans*. Psychology Press.
9. Blue, L., & Traynor, P. (n.d.). Deepfake audio has a tell – researchers use fluid dynamics to spot artificial imposter voices. The Conversation. https://theconversation.com/deepfake-audio-has-a-tell-researchers-use-fluid-dynamics-to-spot-artificial-imposter-voices-189104
10. Almutairi, Z., & Elgibreen, H. (2022). A Review of Modern Audio Deepfake Detection Methods: Challenges and Future Directions. Algorithms, 15(5), 155. https://doi.org/10.3390/a15050155
11. Inamdar, F. M., Sateesh Ambesange, Mane, R., Hussain, H., Wagh, S., & Prachi Lakhe. (2023). Voice Cloning Using Artificial Intelligence and Machine Learning: A Review. *Journal of Advanced Zoology*, *44*(S7), 419–427. https://doi.org/10.17762/jaz.v44is7.2721

12. Boukherouaa, E. B., AlAjmi, K., Deodoro, J., Farias, A., & Ravikumar, R. (2021). Powering the Digital Economy: Opportunities and Risks of Artificial Intelligence in Finance. *Departmental Papers*, *2021*(024). https://www.elibrary.imf.org/view/journals/087/2021/024/article-A001-en.xml
13. Barr-Pulliam, D. D., Brazel, J. F., McCallen, J., & Walker, K. (2020). Data Analytics and Skeptical Actions: The Countervailing Effects of False Positives and Consistent Rewards for Skepticism. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3537180
14. Hilal, W., Gadsden, S. A., & Yawney, J. (2021). A Review of Anomaly Detection Techniques and Applications in Financial Fraud. *Expert Systems with Applications*, *193*(1), 116429. https://doi.org/10.1016/j.eswa.2021.116429