Technical Topic: Deep Fake Detection in Bank Fraud

STS Topic: Impact of Deep Fake Technology on Public Trust

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
Rhea Agarwal

November 8, 2024

Technical Team Members:
Baani Kaur, Drake Ferri, 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
Rider Foley, Department of Engineering and Society
Gregory Gerling, Department of Systems Engineering

Problem Definition: The Rise of Deep Fakes

Over the past few years, AI has helped grow multiple industries, including healthcare, education, and finance (Salokannel, 2023). However, the growth of AI has and is further predicted to lead to a substantial increase in deep fake bank fraud. “Deloitte’s Center for Financial Services predicts that generative AI could enable fraud losses to reach $40 billion in the United States by 2027, from $12.3 billion in 2023, a compound annual growth rate of 32%” (Deloitte, 2024, p.1). A deep fake is a video or audio of a person in which their face or voice have been altered so that they look or sound like someone else. (Ongtingco, 2021). The past few decades have also seen a significant increase in the use of social media (Chaffey, 2024). As a result, it has become easier for criminals to find enough data about people to create convincing deep fake audios, hence increasing deep fake based bank fraud.

Along with increased social media use, there has also been an increase in the availability of voice cloning tools. Today, there are hundreds of freely available tools on the internet, such as ElevenLabs, Lovo, ReSpeeecher, and Voice.ai, that people are able to use to not only create new, human-like voices, but also clone voices of specific people. Additionally, there has been a great improvement in the quality of voices produced by easily available tools. Today, we are able to make cloned voices that can easily convince humans that they are human voices (EurekAlert!, 2024). The increase in access to all these resources means that it is increasingly important for banks to be aware of, and improve their security systems to ensure that they are not vulnerable to deep fake attacks (Khan, Malik, Ryan, & Saravanan, 2023). With the speed at which voice cloning technology is improving, it is difficult, but very important, for banks to keep pace, to ensure that their systems are keeping up with this growth.

Speaking with our capstone client, the group has realized that the most important question banks are trying to answer with their security systems is “Is this voice live and human” as opposed to if the voice is the specific person. Through this project, the team hopes to understand the extent to which banks’ security systems are currently able to identify a cloned voice, and the specific factors/qualities of a voice that the systems are particularly vulnerable to. This will be done by creating a system on python that compiles voice cloning tools and a bank-like security system, to output the probabilities of different voices passing through. By doing this, we will be able to help banks better their AI voice detection security systems, and hopefully help reduce deep fake based bank fraud occurrences. This project aims to help the client understand their vulnerabilities and improve their security system in order to reduce successful deep fake fraud attacks.

Building Resilience in Bank Authentication Systems

Banks are having a hard time keeping up with the growth of AI voice cloning tools, and need to improve their security systems in order to ensure they have minimal vulnerabilities (Goswami & Ross, 2024). The countermeasures and security systems in place today need to be developed to ensure minimal success of these deep fake fraud attempts (Khan, Malik, Ryan, & Saravanan, 2023). In order to help the bank identify their vulnerabilities in their security system, the team will be creating a system on python that gives us a sensitivity analysis which identifies the vulnerabilities of a security system (see Figure 1).

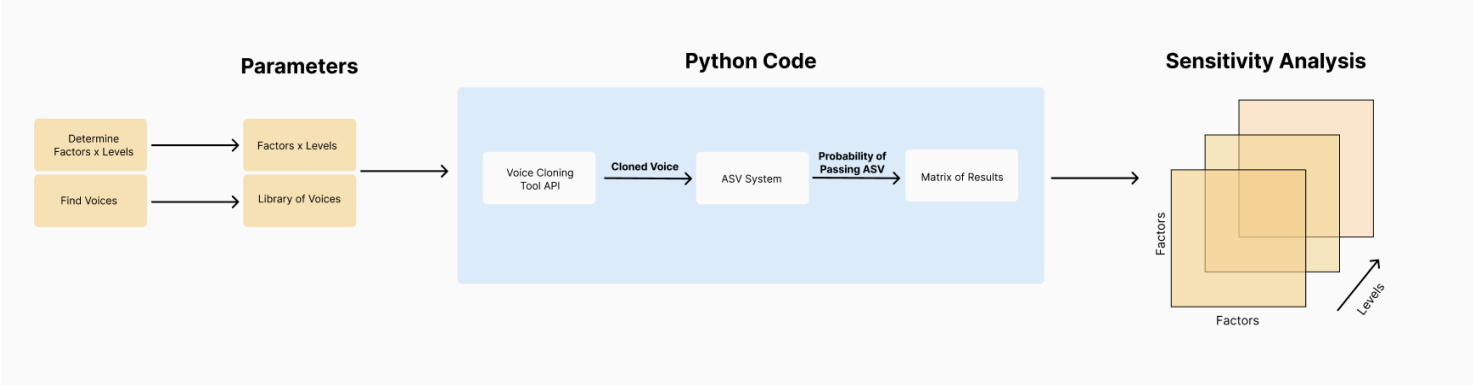


Figure 1: Python System Visualization (Source: Capstone Team, 2024)

The first step in building the system is compiling a list of factors and creating a voice library. In order to create the library of voices (see Appendix 1), our team has identified 3 factors, and associated levels, to work with: tool training times (varying by the ability of the tool), sex (male and female), and volume (soft, medium, and loud). The team created a 5 minute script, to ensure consistency in all tool training audios, which will be recorded by all team members in each of the 3 volumes. These audios will be cut to obtain audios for all other lengths. Each member of the team will then use 3 out of the 5 tools we are working with to clone their voice, in a way that ensures that each tool is being used by at least one male and one female. This will result in approximately 27 clones and 9 training audios per person.

The library of voices will then be used as a parameter in our python system, but will first be used to test manually against our client's bank security system. To do this, 162 calls will be placed to their help agent, and play one of our clones each time. After making all the calls, the bank will send us results in the form of a score for each call, indicating if the system identified it as live and human, not human, or indeterminate. Carrying out this experiment with the bank's system will help narrow down the library to voices most likely to pass the security system.

After identifying factors to narrow down the library, the library will be built further using those factors. Part of our final library will be input into python, along with APIs (Application Programming Interfaces) for the cloning tools (ElevenLabs, Lovo, Voice.ai, FineVoice, and an Open Source tool) and an ASV (Automatic Speaker Verification) system. An ASV is used to identify a real voice from a spoofed voice (Kassis & Hengartner, 2023), which is the banks' primary focus. The goal of this system is to automate the testing process carried out manually with the bank. This will allow for an increased number of trials (100 times per audio) as well as increased number of audios, resulting in more accurate results. The voices will pass through each tool to create clones, and these clones will pass through the ASV 100 times each. We will then obtain probabilities of each voice successfully passing through the ASV, and the results will be in the form of a sensitivity analysis which will tell us, quantitatively, which combinations of factors of a voice result in the most successful clone. This sensitivity analysis will be our final deliverable, pointing out to banks their biggest vulnerabilities for them to work on in their security system.

The problem explored by the technical aspect of this project can be broadened to include the impacts of deep fake technology on more than just bank security. The next section will look at the aspect of public trust in relation to deep fake technology.

Societal Trust in the Age of Deep Fakes

AI voice cloning tools are having a big impact on banks and other financial institutions. However, this problem has more social and human aspects as well. With increased awareness of these tools and the quality of clones the tools are able to make, it makes sense that there would be reduced trust in digital media and financial institutions (Deloitte, 2024). This problem can be understood using Wyatt's Technological Determinism (Wyatt, 2008) framework. It is built on the basis that technology is the primary driver of social change, which is its justificatory foundation, arguing that societal changes are largely shaped by the advancement and use of technology. Wyatt argues that technology is not simply a tool used by society but an autonomous force that influences and sometimes dictates the direction of social progress, which describes the descriptive aspect of the framework by showing how AI tools correlate with reduced societal trust. The methodological aspect lies in analyzing how the adoption of AI voice cloning tools has directly contributed to increased deep fake-based bank fraud attacks through empirical and case-based research. Finally, the normative aspect emerges in the ethical implications of this shift, as reduced public trust in financial institutions and digital media raises critical questions about the desirability and control of these technological advancements. Here, the AI tools are the technology that are driving change in societal trust. This goes to show the impact that technology has on our society and supports Wyatt's argument that it may be the primary driver of social change.

The main issue discussed in this paper (deep fake based bank fraud) is driven by the growth in AI cloning technologies, and how that impacts overall public trust. This problem can also be seen from a broader perspective than just bank fraud. Rise in the use of deep fakes has led to an overall increase in misinformation that is convincing enough to have an impact on

society (Helmus, 2022). This goes to show how this technology of AI tools has societal impacts beyond the scope of the technical topic, and must be explored in a human and social context to understand it. This is another way to look into how this technology can impact trust in a multitude of ways, and not only in relation to bank fraud and financial institutions. Today, people are able to create not only audio, but also video clones of well known, possibly influential people, saying anything they wish, making it difficult to trust almost any source of digital media out there.

Investigating Public Trust in Digital Environments

The research question I will answer is to what extent does the increase in access to deep fake technology impact public trust in digital media and financial institutions? By looking into financial institutions, this research question will relate to my technical capstone project, but going beyond that and exploring general digital media will help me explore the bigger picture impact of increased access to, awareness of, and use of deep fake technologies today. In order to answer this question, the method that I will be using is a survey. Since this question is one in which responses are subjective to a person's understanding of deep fake technology, it is important to include as many people as possible in order to get a variety of opinions. To do this, I will be surveying 300 people. I will use social media messaging platforms to send my google survey to all my friends and family, and ask that they forward it to as many people as possible. This way, I will be able to collect responses from a wide range of age groups, and hopefully a representative sample of the population. I will not, however, be restricting numbers to make my population specific, since I am making the assumption that 300 responses will give me an overall holistic view.

The survey will include 9 questions adapted from prior research conducted by the Pew Research Center (Pew Research Center, 2019). Data collected from the survey will then be analyzed using Wyatt's technological determinism framework to understand how much people really think deep fakes and misinformation is impacting their level of trust, hence seeing the impact of technology on society. To further analyze the results, they will also be compared to those from the Pew Research Center results, in order to see if and how these results have changed in the past 5 years.

Conclusion

The rise in use of deep fake technology is impacting various aspects of the world. The technical deliverable of this project will help banks, specifically our client bank, identify the biggest vulnerabilities to deep fake voices in their security system, allowing them to make specific improvements. The ultimate goal of this is to help reduce the number of successful deep fake based fraudulent attacks on the bank. Analyzing results from the survey about trust will help understand the bigger picture impacts of deep fake technology on our society. Overall, the research paper is expected to result in a deeper understanding of the scale of the impacts deep fake technology.

Resources

Amazon Web Services. (n.d.). *What is an API?* https://aws.amazon.com/what-is/api/#:~:text=API%20stands%20for%20Application%20Programming,other%20using%20requests%20and%20responses.

Carnegie Endowment for International Peace. (2020). *Deepfakes and synthetic media in the financial system: Assessing threat scenarios*. https://carnegieendowment.org/2020/07/23/deepfakes-and-synthetic-media-in-financial-system-assessing-threat-scenarios-pub-82399

Capstone Team (2024) System Outline Image

Chaffey, D. (2023, April 26). *New global social media research summary 2023*. Smart Insights. https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/

Contributor, A. B. J. G. (2024, February 16). *Challenges in voice biometrics: Vulnerabilities in the age of deepfakes*. ABA Banking Journal. https://bankingjournal.aba.com/2024/02/challenges-in-voice-biometrics-vulnerabilities-in-the-age-of-deepfakes/

Deloitte. (2024, May). *Deepfakes expected to magnify bank fraud*. The Wall Street Journal. https://deloitte.wsj.com/cio/deepfakes-expected-to-magnify-bank-fraud-c500b0a2

EurekAlert! (2024, October 17). *Human-like AI voices can drive emotional connection and human-computer interaction*. https://www.eurekalert.org/news-releases/1048879

Goswami, S., & Ross, R. (2024, April 24). *Ai voice cloning pushes 91% of banks to rethink verification*. Bank Information Security.

https://www.bankinfosecurity.com/ai-voice-cloning-pushes-91-banks-to-rethink-verification-a-24932

Helmus, T. C. (2022). *Artificial Intelligence, Deepfakes, and Disinformation: A Primer*. doi:10.7249/pea1043-1

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*. doi:10.3389/fpsyg.2021.664925

Kassis, A., & Hengartner, U. (2023). Breaking security-critical voice authentication. *2023 IEEE Symposium on Security and Privacy (SP)*, 951–968. doi:10.1109/sp46215.2023.10179374

Khan, A., Malik, K. M., Ryan, J., & Saravanan, M. (2023). Battling voice spoofing: A review, Comparative Analysis, and generalizability evaluation of state-of-the-art voice spoofing counter measures. *Artificial Intelligence Review*, *56*(S1), 513–566. doi:10.1007/s10462-023-10539-8

Ongtingco, B. (2021, November 1). *Deepfakes, our future or our downfall?* [Prospectus, University of Virginia]. University of Virginia School of Engineering and Applied Science.

Pan, J., Nie, S., Zhang, H., He, S., Zhang, K., Liang, S., … Tao, J. (2022). Speaker recognition-assisted robust audio deepfake detection. *Interspeech 2022*, 4202–4206. doi:10.21437/interspeech.2022-72

Patalauskaitė, E. (2024). Ethical aspects of content creation. *Filosofija. Sociologija*, *35*(3). doi:10.6001/fil-soc.2024.35.3.14

Pew Research Center. (2019, June 5). *Americans think made-up news and videos create more confusion than other types of misinformation*. https://www.pewresearch.org/journalism/2019/06/05/3-americans-think-made-up-news-and-videos-create-more-confusion-than-other-types-of-misinformation/

Rabhi, M., Bakiras, S., & Di Pietro, R. (2024). Audio-deepfake detection: Adversarial attacks and countermeasures. *Expert Systems with Applications*, *250*, 1–13. doi:10.1016/j.eswa.2024.123941

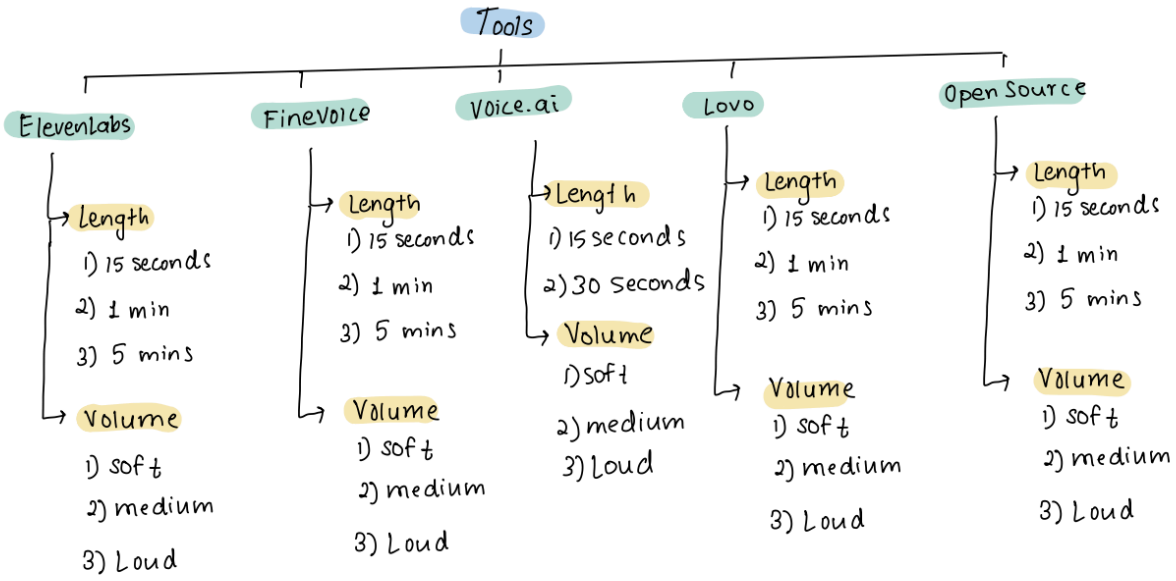
Salokannel, P. (2024, July 1). *The impact of AI: How artificial intelligence is transforming society*. 3DBear. https://www.3dbear.io/blog/the-impact-of-ai-how-artificial-intelligence-is-transforming-society

Shen, Y., Heacock, L., Elias, J., Hentel, K., Reig, B., Shih, G., & Moy, L. (2021). *Artificial intelligence in medical imaging: Opportunities, challenges, and practical applications. Current Radiology Reports*, 9(1), 35–45. https://pmc.ncbi.nlm.nih.gov/articles/PMC8129507/

Vaccari, C., & Chadwick, A. (2020b). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society*, *6*(1). https://doi.org/10.1177/2056305120903408

Wyatt, S. (2008). Technological determinism is dead; Long live technological determinism. In E. J. Hackett, O. Amsterdamska, M. E. Lynch, & J. Wajcman (Eds.), *The handbook of science and technology studies* (3rd ed., pp. 165–180). MIT Press. https://doi.org/10.1177/00027649921955326

Appendix 1: Voice cloning chart



Appendix 2: Questions that will be in the survey

From: https://www.pewresearch.org/journalism/2019/06/05/3-americans-think-made-up-news-and-videos-create-more-confusion-than-other-types-of-misinformation/ - topline document

1. How much of a problem do you think made-up news is in the US today? (very big, moderately big, small, not a problem)
2. How often do you get news from a social media site? (often, sometimes, hardly ever, never)
3. How often do you get news from a news website/app? (often, sometimes, hardly ever, never)
4. How much do you trust the accuracy of news you get from news outlets? (a great deal, some, not much, not at all)
5. How much do you trust the accuracy of news you get from social media sites? (a great deal, some, not much, not at all)
6. How much do you think each type of news and information leaves Americans confused about the basic facts of current issues and events? (a great deal, some, not much, not at all)
    a. Made-up information that is intended to mislead the public
    b. Satire about an issue or event
    c. Video or image that is altered or made up to mislead the public
7. How do you feel about the average American's ability to recognize the following types of news and information? (average american should be able to recognize, too much to ask fo average american to recognize)
    a. Made-up information that is intended to mislead the public
    b. Satire about an issue or event
    c. Video or image that is altered or made up to mislead the public

8. How do you feel about your own ability to recognize the following types of news and information? (average american should be able to recognize, too much to ask fo average american to recognize)
    a. Made-up information that is intended to mislead the public
    b. Satire about an issue or event
    c. Video or image that is altered or made up to mislead the public
9. Have you ever shared news that you later found out was made up? (yes/no)