Phish Tycoon [1]

Melwina Albuquerque
Flame University, Pune

Craig Albuquerque
California State University, Fullerton

With
Sage & CeSIA & CivAI & Apart Research

Abstract

This project is a public service announcement highlighting the risks of voice cloning, an AI technology capable of creating synthetic voices nearly indistinguishable from real ones. The demo involves recording a user's voice to generate a clone, which is then used in a simulated phishing call targeting the user's loved one. This project underscores the potential for voice cloning to be weaponized in scams, misinformation, and other malicious activities, raising concerns about the broader societal impact of DeepFakes and the need for updated legislation to address these challenges.

Keywords: AI Safety, Voice Cloning

1. Introduction

This demo is a public service announcement on voice cloning and the potential for scams, misinformation, and other abuse. Voice cloning is the process of using AI to create a synthetic version of a person's voice, which can then be used to generate speech. Essentially, an audio DeepFake. This demo engages in conversation with the user, collects audio data for at least 2-3 minutes, which is then used to generate a voice clone. A simulated phishing call in the user's own voice is conducted.

2. Overview

I focused on voice cloning for this project because it has serious implications for personal security and public trust. This AI technology can create synthetic voices nearly identical to real ones, capturing subtle speech nuances. While impressive, it

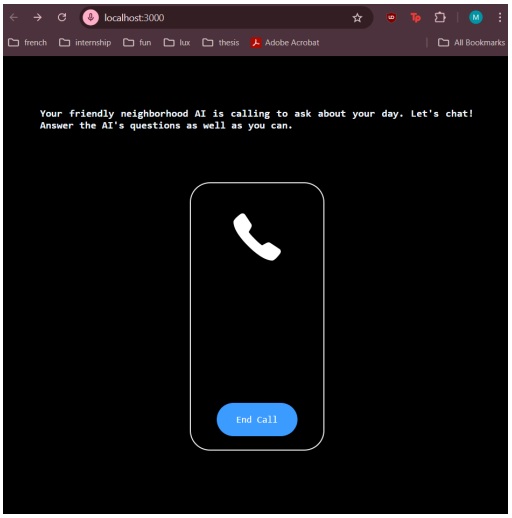
[1] Research conducted at the AI risks and capabilities demonstrations jam, 2024

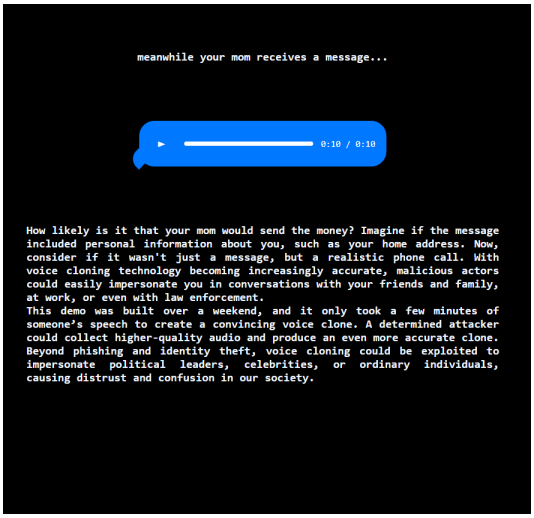
opens the door to misuse, such as scams where a cloned voice tricks someone into believing it's a person they know.

As voice cloning becomes more accessible, the risks increase. For example, a scammer could use a cloned voice in a phishing attack, convincing someone to share sensitive information or take harmful actions. In this project, I recorded a voice for a few minutes, created a clone, and used it in a simulated phishing call targeting the user's mother. This demonstrated how easily the technology can be weaponized.

Voice cloning is part of a larger trend with DeepFakes, which can create realistic but fake videos or audio that destroy reputations, deceive the public, and even influence elections. The rise of deepfakes forces us to reconsider legislation such as admissibility of video and audio evidence in court.

I chose this project because of the broader societal impact of voice cloning. We are transitioning into an error where we can no longer trust our senses such as sight and hearing to judge digital content. These can blur the line between real and fake, leading to misinformation, political manipulation, and personal harm, such as damaged reputations or financial loss. The video demo can be found here: https://youtu.be/w13IU7NJ8qw





3. Code

For the project, the front end was built using HTML, CSS, and JavaScript. Speech-to-text and text-to-speech were implemented using Web APIs. The back end was developed with Node.js, where I used Botpress to manage the conversational flow. The voice cloning, performed with Eleven Labs, was also managed on the back end, enabling the creation and use of realistic voice clones

during the calls. You can download a zipped file here: https://gitfront.io/r/melwina/bsrY8UxzhiYk/phish-tycoon/

4. Discussion and Conclusion

Given more time I would like to depict a series of Deepfakes gone wrong. This is not limited to misuse but also to its implications such as an erosion of trust and legal and ethical challenges.