Thesis Project Portfolio

Evaluating the Importance of Demographic and Technical Factors in Creating Authentic-Sounding AI-Generated Human Voice Clones

(Technical Report)

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

(STS Research Paper)

An Undergraduate Thesis

Presented to the Faculty of the School of Engineering and Applied Science University of Virginia • Charlottesville, Virginia

In Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Rhea Agarwal

Spring, 2025

Department of Systems and Information Engineering

Table of Contents

Sociotechnical Synthesis

In my capstone project, my team and I used human survey respondents (human actors) and a machine model called NISQA (non-human actor) to evaluate how different factors impact the perceived realness of AI-generated voice clones. To further understand how non-human actors like deepfake tools influence sociotechnical systems beyond perception, my STS research used Actor-Network Theory (ANT) to analyze how deepfake technologies interact with corporate protocols and decision-making processes in the real world. Both projects complement each other, offering insights into the risks posed by cloning tools - one by identifying the factors leading to the most realistic, and therefore dangerous, clones, and the other by analyzing how such technologies can cause instability in security networks.

In my technical project, we investigated demographic (age, gender, Hispanic identity) and technical (background noise, AI tool, training time) factors that impact how realistic a cloned voice sounds to both human listeners and a machine. We built a library of 336 cloned and authentic voices encompassing all our factors, then used optimization to select a balanced subset of 81 (67 cloned and 14 authentic) voices. This subset was used to create a survey in which 449 respondents each rated 6 of the 81 voices on a scale of 1–5 based on how real they sounded. These 81 voices were also scored by the NISQA model. Results revealed that clones built from under-30, male, non-Hispanic sources, trained for either 15 or 60 seconds, using tools other than Lovo, and with added background noise, were indistinguishable from authentic voices. Additionally, we observed no correlation between human and machine ratings, suggesting that tools like NISQA evaluate voice naturalness using different criteria from humans.

For my STS research paper, I used ANT to analyze the 2019 UK CEO fraud case, in which deepfake audio was used to impersonate a company executive and authorize a large transaction. ANT helped examine how executives and employees (human actors) and deepfake tools and voice verification systems (non-human actors) interact within a company's security network. I argued that the network's vulnerability came from the unchecked agency of non-human actors and the failure of the voice verification system. I also used parallel cases from CNN and The Guardian to show how scammers combined deepfake tools with publicly available data to create deepfakes that bypassed both human and machine judgment. This shows how deepfake tools are active participants in the network and can shape organizational behavior.

Working on both projects simultaneously helped build my understanding of the risks posed by voice cloning. The technical project gave me tools to quantify and predict which voices create the most convincing clones, while the STS research helped me contextualize the findings into real-world cases. ANT specifically helped me see how voice cloning technologies don't just generate content, but actively participate in shaping decisions. These learnings will impact the considerations I have when designing technologies. I acknowledge the importance of designing technologies that not only anticipate their intended use, but also how they may be exploited in sociotechnical environments.

Word Count: 498