Deep Reflection

MAT594G Final Project Proposal

Mert Toka

CONCEPT

We, as humans, are more connected than ever before. Communication defies spatial boundaries and is more accessible by the general public more frequently. Contacting a person half-way across the world takes a matter of seconds, and social media enables connectivity at an extended rate. However, unlike any other time in human history, the world feels eerily divided.

In today's politically divided world, listening and relating to arguments that we do not necessarily associate ourselves with is getting increasingly challenging. People become anxious about different points of view and not ready to step out of their comfort zones. As opposed to text-based communication, face-to-face interaction seems to be an effective way of alleviating this disconnect between people. By mirroring and subtly mimicking other people's facial expressions, we understand what they are experiencing better.

I wonder, instead of observing another person talking about a counter-argument, what happens if we see/hear the other side's story on our faces? Would it make us believe that their concerns are as real as ours and cause a better understanding? I hypothesize that employing Deep Fakes to transfer one person's facial expressions and lip movements onto the other person's portrait would help provide a greater understanding regarding their counter-arguments.

In COVID-19 times, the domestic violence towards women in Turkey spiked drastically, and government officials and law enforcers are still trying to downplay the importance of this issue. I imagine a possible use case for this project as a public website. The front-end would let people upload a photo of themselves. The system would animate the portrait based on the facial expressions and narrative transferred from the first-hand stories of sexual assault survivors without revealing their identities.

I imagine voice would be an essential indicator for the transmission of the message. In this regard, rather than a still image, the system may prompt to record a short video of the user (~5-seconds) reading predefined text displayed on the screen. The recording would allow sampling the user's voice, and a voice synthesis engine can "stylize" the transferred narrative with the frequency components of the user's voice.

On the other hand, when most people hear themselves on a recording, they usually won't believe how they sound from others' perspectives. This is mostly because we not only hear our voice from the sound waves that we emit to the outside world, but we also feel the vibrations from inside our bodies. Considering such an idea, we may not need to "synthesize" a correct version of the speech. It is possible that a relatively more straightforward pitch-shifting would suffice.

ARGUMENT

The idea emerged from a personal narrative. In Turkey, I am currently in a situation where I ponder how to better convey one's message to the other without causing too much distress. When we started

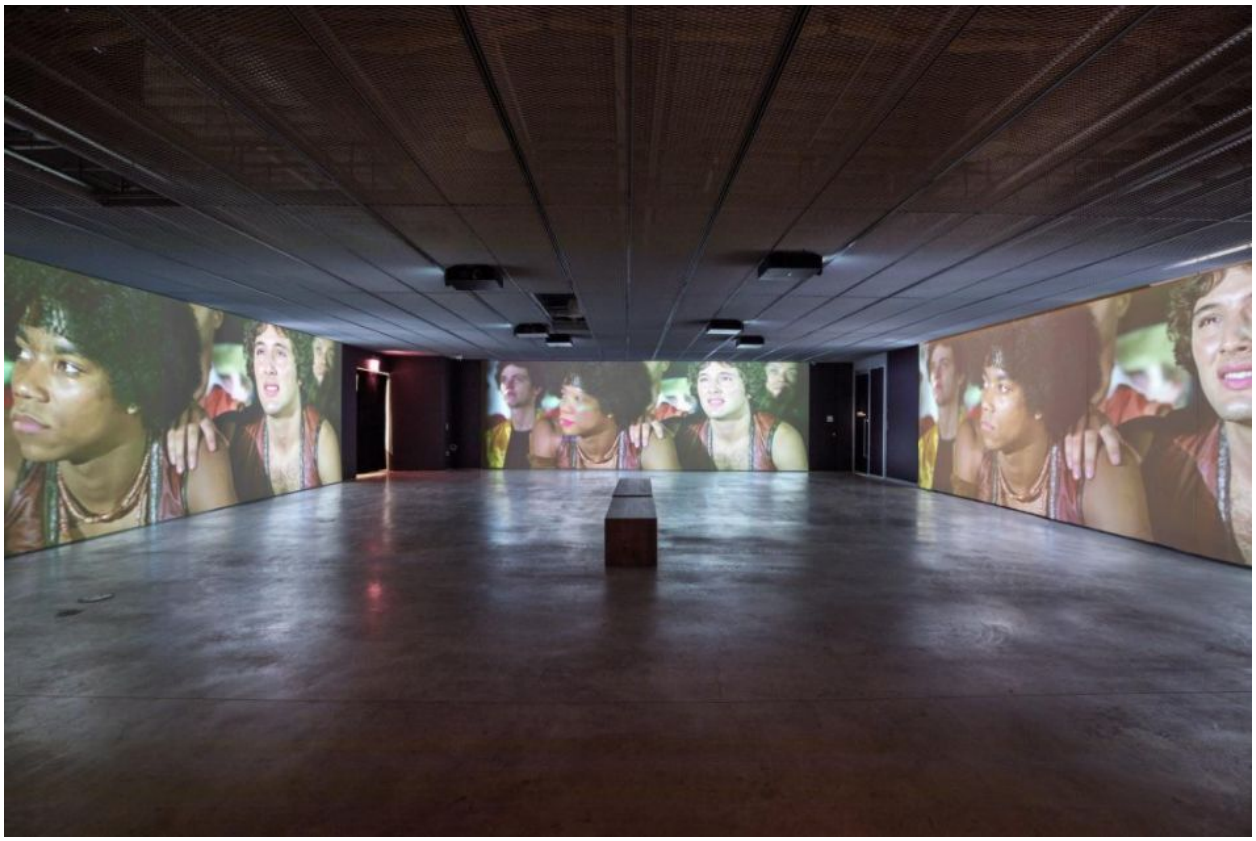
covering Deep Fakes in the class, I started speculating about employing them for my situation. If such a method is applied, I imagine the level of surprise mixed with being able to relate to one's face can invoke better empathy.

METHODOLOGY

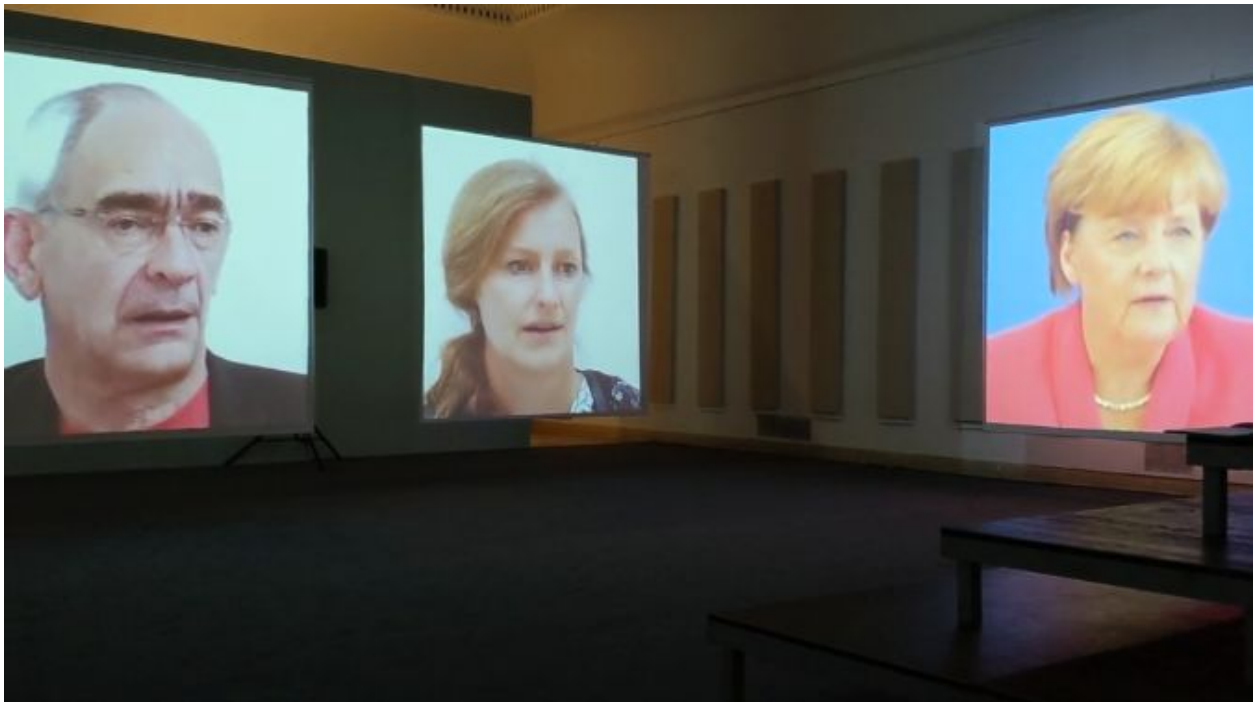
The following pipeline could be of use for the implementation:

- Designing the website for uploading image/video of the user, selecting an anonymized narrative
- Setting up an Amazon Sagemaker cloud environment for hosting trained ML models
- Running vid2vid on image/video of the user with the story, facial expressions
- [if the video] Running speech synthesis to synthesize voice
- Displaying the result on the website

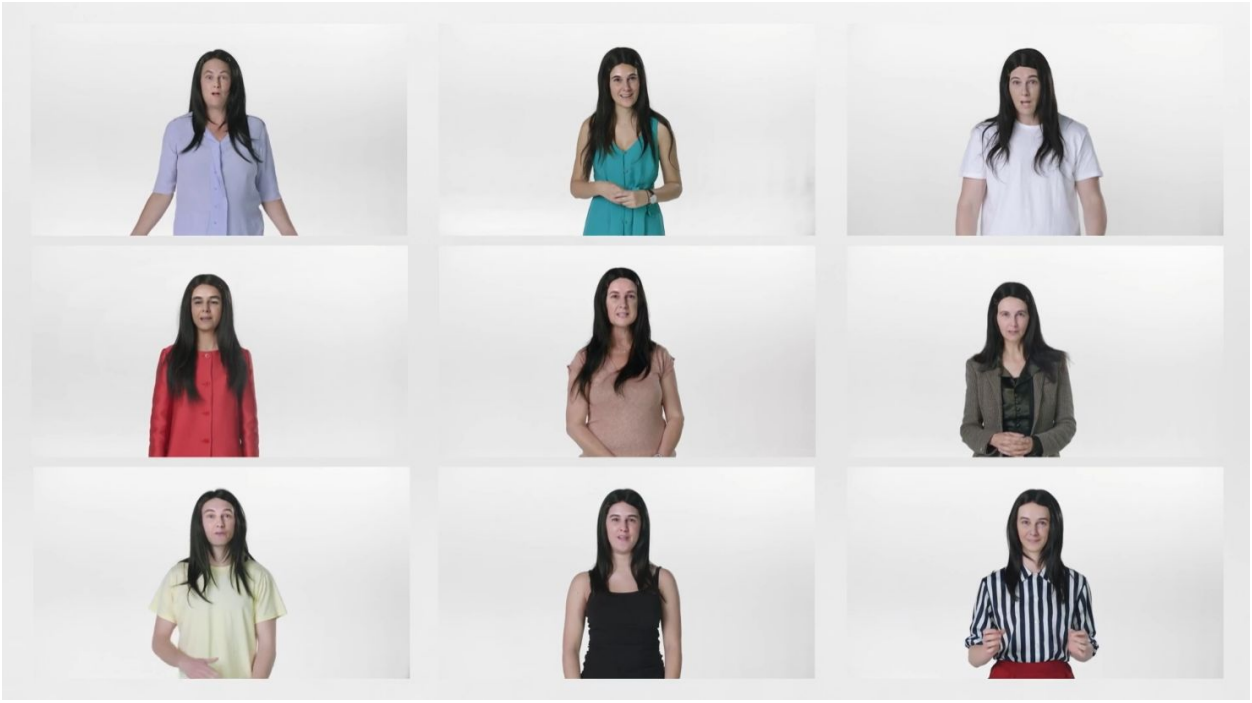
REFERENCES



James Coupe, Warriors



Tamiko Thiel, Lend me your Face!



Gillian Wearing, Wearing Gillian