Thursday, July 7, 2022

In this fraught and fractured time, wouldn't we benefit from some humanizing assurance?

Rather than the tinny machine-generated sounds we absorb from our technology, or the universal impersonal voices chosen by Google, Apple and Amazon, it may be better to interact with own comforting people.

The warm baritone of James Earl Jones could be serving up music and weather reports this weekend.

Our joke-a-day app might be enhanced with absurd cackle of Gilbert Gottfried.

Goodnight Moon could be read to your kids by voice of their grandfather – even if he passed away in the 1990s – keeping him alive in all of your memories.
We have the technology to make this happen, and tech companies are testing ways to bring your favorite voices to respond to your questions and read to your children. While this technology sounds interesting at first, deeper thought raises serious concerns. Just because we can do this thing, doesn’t mean we should do it. This version of deep fakery may bring comfort to some, but loss, confusion, and emotional ruin to others.

Listening to the voice of a deceased loved one can hold warm memories, but allowing an AI to recreate that voice in response to questions in real time is creepy, confusing and possibly would run counter to the wishes of the person whose voice is being used. If your long-dead aunt is answering questions posed to Alexa (“Aunt Ellie, what year did Supertramp have its first number one hit?”) or Siri (“Aunt Ellie, what direction do I turn onto Chestnut Street?”), does that keep a happy memory alive or does it trivialize your relationship and her mortality? How psychologically healthy can it be to interact with people who were once important in our lives, but now are gone forever? How will children be affected by the responsive voices of people who passed?

This technology has existed for years, as anyone knows who watched the famous deep fake internet memes of Barack Obama bogus speeches. AI can start with relatively small samples of a person’s vocal pattern, build out models of how that voice would sound pronouncing many words, and use the model speech-pattern library to generate an unlimited variety of words in the speaker’s language. This library can then be the source of vocal presentation to answer questions or read text. Now some companies are considering how to personalize their responses.

The voice-capture and parroting technology was just unveiled at Amazon’s Re:Mars conference in Las Vegas, is in development, and could allow the Amazon virtual assistant to sound like any person based on less than a minute of spoken recording of that person’s voice.

But what if that person doesn’t want their voice to serve Amazon as a digital voice slave, saying anything that the company decides it wants you to hear? People could object. Of course, dead people cannot object. There may be commercial reasons for objection, where Barry White may not want his dulcet tones to fill a million bedrooms with interactive discussion for free. Celebrities should be able to sell to their own voices for your home digital operations. But many people would simply find the concept of shilling for Amazon or Apple distasteful and creepy, demanding that their voices to remain within their own control.

According to NBC News' website, “Amazon’s push comes as competitor Microsoft earlier this week said it was scaling back its synthetic voice offerings and setting stricter guidelines to “ensure the active participation of the speaker” whose voice is recreated. Microsoft said Tuesday it is limiting which customers get to use the service — while also continuing to highlight acceptable uses such as an interactive Bugs Bunny character at AT&T stores.” Presumably Mel Blanc or another voice actor has approved this message.

Amazon hopes that interacting with Amazon technology through a familiar voice will build greater trust with Alexa by capturing more "human attributes of empathy and affect." However, this seems like unearned and unwarranted trust. By simply
mimicking a human voice that is important to you, Amazon expects the trust you have built in decades with your mother can transfer to Amazon, in what may be the most cynical business move I have ever considered. Yes, the up-selling message is coming from Amazon (“You haven’t signed up yet for your Amazon music account”), but it sounds like Grandma – you trust Grandma, don’t you?

The Washington Post notes that bad actors can use this technology to put words in your in mouth that you never said. It quoted cybersecurity expert Rachel Tobac as saying, “I don’t feel our world is ready for user-friendly voice-cloning technology. If a cybercriminal can easily and credibly replicate another person’s voice with a small voice sample, they can use that voice sample to impersonate other individuals. That bad actor can then trick others into believing they are the person they are impersonating, which can lead to fraud, data loss, account takeover and more.” I imagine that voice deepfakes could fool biometric voice-based security (“My voice is my passport” for all you Sneakers fans), but I don’t know enough about either technology yet to know how they would interact.

This may be a new arena for the growing law and practice of digital estates, where people determine what happens to their electronic and online assets following death. Digital-focused wills can address what happens with online businesses, ecommerce sites, accounts and licenses with third parties, and even personal storage. Add to the list the disposition of your voice digitized in recordings. Do you want anyone to be able to use your voice as an interactive tool? If so, who can do it and who can't? Or do you want to leave that choice to the executor of your estate or two a selected heir? The effect could be financial for some famous voices – some people will want to continue to hear Walter Cronkite read the news to them every night – and simply a personal decision for others. The decision could be made on religious grounds.

And whatever happens with your voice could also be applied to your image. Using video of a person speaking can allow an audiovisual deepfake to speak words chosen by others, or even by a machine.

It is amazing how the realization that we have the tools to further humanize our technology can lead to so many legal and ethical questions. We can perform miracles with our tech. Should we?

Copyright © 2022 Womble Bond Dickinson (US) LLP All Rights Reserved.

National Law Review, Volume XII, Number 188

Source URL: https://www.natlawreview.com/article/voice-reason