

Voice Stealing

17 August 2023

Jim Stevenson



This essay is slightly tangential to my usual fare, but it is prompted by a most amazing video that convinced me that the impact of AI this time is not hype, but rather a real threat to our society. I found the video ([1]) at 3 Quarks Daily and it was of Johnny Cash singing a song called “Barbie Girl” to the tune of his trademark Folsom Prison Blues—only it wasn’t the late Johnny Cash (1932–2003), it was AI!

One of the essential draws of mathematics as a way of life is that it has a method of establishing truth. It is so powerful, that it undermines charlatans and fakes, the powerful and the authoritarian. Everyone is on an equal footing and can be challenged to prove their assertions using established principles. This is why, for example, the Japanese claim to have solved the abc conjecture is still not accepted by the mathematical community at large ([2]). Mathematics is so powerful that attempts to port its methods to nonmathematical realms have been growing since the Enlightenment,¹ perhaps to its detriment, since non-mathematical fields require other forms of evidence and justification. But the logical model at its heart is the basis of rationalism and has been fundamental to our understanding of a democratic society.

That is why the latest developments in technology that undermine how we understand things is so dangerous, especially when they falsify our senses, like seeing and hearing. We have long known the dangers of Photoshop that can fake photos and render them useless for proving anything. But faking the voice seems new to me.

I had followed the ups and downs of AI for years, every since the days of rule-based “expert systems” of the 1970s and 80s and the rise of neural nets and other machine learning techniques at the same time. Their evolution to the latest encroachments on human expression exemplified by the likes of ChatGPT have proved interesting, but still limited ([3], [4]).

For humans the Achilles heel of current machine learning AI system, such as those behind driverless vehicles and anything using neural nets, is that they cannot explain their results, which is at the heart of a mathematical proof or a judicial determination derived from previously established positions (or at least it is supposed to be). In general no one knows what factors a neural net has used to “recognize” something by virtue of a correlation with its training data. It could have nothing to do with the sort of reason a human would give, but rather some artifact of the data that a human would ignore as not being relevant.

I confess I don’t know the details of how the Johnny Cash video was able to model someone’s voice after the patterns detected in that voice. The authentic sounds of an electronic keyboard seem easier to imagine, since samples of different musical instruments all involve lengths of fixed, discrete pitches, whereas the voice seems to be continuously modulated. As I grow older and meet friends I haven’t seen for years or view later images of aging actors, I find I don’t immediately recognize them. But once they speak, I find their voices have lost none of their distinctive pattern and are completely familiar. So faking a voice seems a higher order of imposture than faking a picture. I now see the

¹ “Newton and the Declaration of Independence” (<http://josmfs.net/2019/10/24/newton-and-the-declarationof-independence/>)

dystopian concern of actors fearing their “likeness”, including their voice, would be used without their permission, especially after death. And think of what Amazon is going to be able to do with all the voice captures it has from Alexa—and Apple with Siri.

In this age of lies the power of AI fakery is an order a magnitude worse than any text falsification. The mind boggles at how in the world we can establish authenticity in anything (other than mathematics, of course). Maybe we will have to jettison our ubiquitous devices and return to old fashioned face-to-face encounters.

References

- [1] Johnny Cash sings “Barbie Girl” (A.I.) ([@ThereIRuinedIt](https://www.youtube.com/watch?v=HyfQVZHmArA) (<https://www.youtube.com/@ThereIRuinedIt>) The Man in...Pink? (AI voice model trained by @therealmockingbird (https://www.youtube.com/channel/UCUqxTYpZ8lhXuJR42n6_UQw) / singing by me)
- [2] Woit, Peter, “Latest on abc” *Not Even Wrong* blog, 3 April 2020. (<https://www.math.columbia.edu/~woit/wordpress/?p=11709>)
- [3] Hofstadter, Douglas, “Gödel, Escher, Bach, and AI” *The Atlantic*, 8 July 2023. (<https://www.theatlantic.com/ideas/archive/2023/07/godel-escher-bach-geb-ai/674589/>)
- [4] Collier, Angela, “AI does not exist but it will ruin everything anyway”, 8 July 2023. (https://www.youtube.com/watch?v=EUrOxh_0leE)

© 2023 James Stevenson
