When audio deepfakes put words in Jay-Z’s mouth, did he have a legal case?

Enlarge / This is not an article about Daft Punk remixing or mashing up Jay-Z classics. It’s a photo illustration about machine learning models being applied to famous people’s voices. (But, hey, we’re ready for that Daft Punk + Jay-Z collab over here.)

reader comments

55 with 41 posters participating

Share this story

In late April, audio clips surfaced that appeared to capture Jay-Z rapping several unexpected texts. Did you ever imagine you’d hear Jay-Z do Shakespeare’s “To Be, Or Not to Be” soliloquy from *Hamlet*? How about Billy Joel’s “We Didn’t Start the Fire,” or a decade-old 4chan meme? All of these unlikely recitations were, of course, fake: “entirely computer-generated using a text-to-speech model trained on the speech patterns of Jay-Z,” according to a YouTube description. More specifically, they were deepfakes.

“Deepfakes” are super-realistic videos, photos, or audio falsified through sophisticated artificial intelligence. The better-known deepfakes are probably videos, which can be as silly as Green Day frontman Billie Joe Armstrong’s face superimposed on Will Ferrell’s, or as disturbing as non-consensual porn and political disinformation. But audio deepfakes— AI-generated imitations of human voices—are possible, too. Two days after the Jay-Z YouTubes were posted,

they were removed due to a copyright claim. But just as quickly, they returned. The takedowns may have been a first attempt to challenge audio deepfake makers, but musicians and fans could potentially be grappling with the weird consequences of machine-generated voice manipulations long into the future.

Here's a breakdown of Jay-Z's copyright dispute, the laws around audio deepfakes, and what all this could mean in the years to come.

What happened with the Jay-Z audio deepfakes?

The Jay-Z clips are hosted on a YouTube channel called Voice Synthesis, which is full of famous voices delivering unexpected material. All posted over the past several months, these Reddit-friendly pairings include Bob Dylan covering Britney Spears, Frank Sinatra crooning "Dancing Queen," and various presidents reciting rap lyrics—even the particularly believable George W. Bush take on 50 Cent's "In Da Club."

[embedded content]
Jay-Z "covers" the Billy Joel song "We Didn't Start The Fire," as posted by YouTube channel Vocal Synthesis
On April 26, in a new video posted to the channel, the simulated voices of Barack Obama, Donald Trump, Ronald Reagan, JFK, and FDR claimed that YouTube had taken down two Jay-Z clips at the request of his company Roc Nation. Two days later, both of those videos—the Jay-Z-ified snippets from *Hamlet* and "We Didn't Start the Fire"—were back in place. A YouTube spokesperson told us that the takedown requests were found to be "incomplete" but did not specify who filed them. The spokesperson said the videos have been "temporarily reinstated" pending more information from whoever filed the claims. The ball now seems to be in Jay-Z's court. A spokesperson for the rapper has not responded to our requests for comment.

Does Jay-Z have a winning case?

Probably not. According to the anonymous creator of the Voice Synthesis channel, Roc Nation's takedown requests claimed, "This content unlawfully uses an AI to impersonate our client's voice." But legal experts think that using an AI to impersonate someone's voice does not violate existing copyright law. "Their copyright claim is ridiculous," says Bill Hochberg, a music and media lawyer whose clients have included the Bob Marley estate. "You can't copyright a vocal style." Adds Jim Griffin, managing director of digital music consultancy OneHouse and a former Geffen Records tech executive, "I do not see rights issues here."

If Roc Nation had taken legal action against the creator of Voice Synthesis, that would be a matter of public record, and it doesn't appear that's the case. YouTube's takedown process isn't so transparent. "The most important part, to complainants, seems to be that DMCA provides a quick tool to get something off the Internet," notes Meredith Rose, policy counsel at Public Knowledge, a nonprofit that advocates free expression, citing the Digital Millennium Copyright Act. "Now that the videos are back, I'm curious to see how far they push the issue."

Could audio deepfakes of rappers or singers violate laws other than copyright?

It depends. Some states have a right of publicity, which allows an individual to control the commercial use of their name and likeness. In California, the entertainment industry has been lobbying for updating publicity-rights rules to address deepfakes. A handful of states have recently enacted laws against deepfakes used for non-consensual porn or to interfere with an election.

A deepfake of Jay-Z rehearsing for *Hamlet* is unlikely to run afoul of current laws against soundalikes. Tom Waits won a $2.5 million verdict against Frito-Lay in 1990 over a Doritos ad that had an impersonator rumbling in Waits' recognizable style. But the Voice Synthesis clips are clearly labeled as fakes, and their creator has maintained that they were "intended as entertainment," not for any malicious purpose. "'Weird Al' Yankovic didn't get sued for imitating the top pop stars he targeted," Hochberg observes.

How are audio deepfakes different from sampling?

The self-described "hobbyist" behind the Voice Synthesis channel told the blog Waxy that the Jay-Z deepfakes were created with Tacotron 2, a text-to-speech program developed by Google. The software has to be "trained" with audio samples and text transcripts. The actual voice is used in the creation but from there it's all ones and zeros from the AI. Musicians sue all the time over unauthorized samples of their work in other artists' songs, so it may not seem unreasonable that they could sue for unauthorized samples in a simulator of their own voices. It seems, though, the algorithms may have the law in their favor. Recently, a judge ruled that Drake's use of jazz musician Jimmy Smith's 1982 song "Jimmy Smith Rap" on the opening of his track "Pound Cake/Paris Morton Music 2," from 2014's *Nothing Was the Same*, "adds something new" and is "transformative." Jessica Meiselman, a lawyer who has written for Pitchfork.com, says, "This one smacks of that." That Drake song's featured guest, in fact, was Jay-Z.

What might be next for music and audio deepfakes?

More experimentation and more legal challenges. AI research group OpenAI recently unveiled Jukebox, where you can hear such creepy-yet-fascinating, machine-generated lyrics as a

Frank Sinatra-style Christmas carol about hot tubs, or an Elvis Presley-style rockabilly jaunt with lyrics about cell division. It’s not hard to imagine the estates of deceased artists seeing potential in audio deepfakes as the technology matures, just as it’s easy to see lawsuits being filed over deepfake “guest appearances” from featured artists who never agreed to collaborate. How long until a SoundCloud rapper has “Jay-Z” on their track?

“Deepfake is very democratizing,” says Peter Martin, CEO of virtual reality-focused creative agency V.A.L.I.S. Studio, which has worked with Janelle Monáe and Run the Jewels. “There’s a load of strange kids from Serbia in their bedroom recreating Hollywood. They haven’t really touched music yet.”

The creator of the Voice Synthesis channel tells me that he expects to get more takedown requests. If necessary, he plans to switch to another video platform and keep sharing links on Reddit. An improved version of OpenAI’s jukebox, he said, could generate realistic audio of any singer, living or dead, covering any song of one’s choice, allowing for unique combinations of performers and styles. “So far we’ve barely even scratched the surface in terms of the possibilities in this space,” he says. “I’m really excited to see how it develops over the next few years.”

This story originally appeared on Pitchfork.com.

https://arstechnica.com/?p=1676510