AI-narrated books are here: Are humans out of a job?

March 13 2024, by Roxana Popescu, The San Diego Union-Tribune

Credit: Pixabay/CC0 Public Domain

If you've listened to an audiobook or a narrated news article in the past year or so, there's a chance it was created not by a human, but by AI software that mimics the sound of a human voice.

Some day soon or further off, synthetic narrators and actors might be commonplace and accepted without so much as a blink, but for now there is room for debate: What is lost and what is gained when a machine does the work of a human performer? Who should earn revenue when jobs like audiobook narration are outsourced to AI?

The co-founder of a San Diego software company called Yembo is wading into this quagmire with an unprecedented answer to an unprecedented scenario. Voice actors in San Diego and beyond are watching this approach to paying a human for AI-enhanced labor with interest and apprehension.

The scenario: Yembo's co-founder wrote and self-published a book about AI, and an actor recorded the English audiobook last year and got paid for that recording time. Now her AI-cloned voice is being used to narrate 15 translations of that audiobook.

The narrator does not speak Swedish, Ukrainian and Turkish, but her voice does.

"US English is narrated by the flesh-and-blood Hailey, (the) rest is AI in her likeness," Zach Rattner, the book's author and publisher, and Yembo's co-founder, wrote in an email.

Hailey refers to Hailey Hansard, the actor whose voice is being cloned. Through her contract, Hansard will be paid royalties for audiobooks in her voice, even though she did not narrate the book in any of those languages.

While AI narration of audiobooks and articles is increasingly prevalent, this may be the first instance of royalty payment for AI-cloned translations in the audiobook realm—a booming industry that is expected to reach $39 billion globally by 2033, according to market

research company market.us.

"As far as I know, this audiobook project is the first one where the narrator gains royalties on a product that uses their AI likeness, but they didn't create," Rattner said. "It's the first that I know of, and it was enough that when I tried to figure it out, I couldn't find anything. We had to figure it out from scratch. There weren't templates we could find."

Sandra Conde, a San Diego actor whose likeness has been scanned into a generative AI gaming project, reviewed details from the contract and said it addresses the interests of publisher and voice actor in an uncharted, fast-shifting territory.

"It's a new frontier kind of thing, where we don't know what it's going look like, even like two years from now, or a year from now," Conde said.

Robert Sciglimpaglia, a Connecticut-based voice actor and entertainment attorney, said the contract is noteworthy because it is groundbreaking—touching upon audiobook narration, translation and AI.

"This is the wild, wild west," he said. "The (actors') union doesn't have anything for (AI) translation that I know about."

The contract matters because of what's at stake: "This is a big issue in the audiobook world right now: whether you use human voices or use cloned voices. Because there are some audiobooks being done with AI, and narrators are trying to protect live narration—trying to protect their livelihood," he said.

AI will without a doubt replace human narrators, he added.

"The question in my mind is how far is it going to go? Is it gonna take 50 percent of the business? 25 percent? 75 percent? 100 percent? That's the question we should be asking," Sciglimpaglia said.

Tim Friedlander, the president and co-founder of the National Association of Voice Actors, said this contract is significant, even if it's just one example, because it allows for human narration to be replaced or supplemented by AI generated material.

"Any kind of instance where you have normalization of synthetic content, (the contract terms are) going to matter," Friedlander said from Los Angeles.

Human actors, he added, have something over AI tools: their humanity—which lets them give nuanced readings based on lived experience, culture and context. Machines might try to mimic that, but they can't interpret the words in a story or an essay in an authentic way, he said.

Kind of like the Robert Frost saying about poetry being lost in translation.

Rattner agrees. He did, after all, hire a human to record the English audiobook instead of using a cloned voice from the start. Just the translations are cloned.

"I mean, there are inflections. In the audiobook, she snickers and chuckles a couple of times. Like, you do lose something by being AI," he said. But there are scenarios when AI makes sense, he said.

Will listeners care if a voice is human or synthetic?

That might depend on the book. Or the voice.

Help wanted: chromosomes optional

While actors have been paid for projects that record and recombine their voices for decades—Siri debuted in 2011, with a foreboding backstory around consent and compensation—the use of generative AI to clone voices is new and far more efficient, requiring just a small sample to create new material.

Ten years ago "you couldn't do this—you would have had to have a voice actor and pay him for a month" to record a deep bank of sounds and words, said Sciglimpaglia, a member of SAG-AFTRA. "Now you can take a three minute sample and you can do anything you want with it. You can do an audiobook, a film, a TV show, you can put it in three different languages. It only takes a very small amount of data."

The Atlantic magazine uses an AI narration plug-in, as does inewsource, a San Diego investigative news outlet. A cottage industry of AI text-to-speech narration services have proliferated: ElevenLabs, Podcastle, Speechify, Murf AI, Revoicer, Audiobook.ai and others.

Before this leap, acting and audiobook narration were harder to outsource. The price of labor may be cheap in Malaysia and Sri Lanka, but a California cadence is one thing they can't manufacture. AI is a workaround: instead of farming out to people, use machines.

That's why actors and other creative professionals see generative AI as an existential threat.

And that was why actors and writers clashed with studios in strikes last year, Sciglimpaglia said.

Should studios be allowed to scan actors' faces and generate new material using those scans, or should they keep hiring humans even if

synthetic actors—which don't need bathroom breaks or paychecks—could replace them? And if studios do scan an actor's features, can they pay just for that scan, or should they pay for the future or potential uses—uses that would have been fulfilled by the human actor?

The actors' strike settlement allows for AI cloning, but set limits around future uses and added rules around compensation that better protect actors, Sciglimpaglia said.

There are about 100,000 working voice actors in the U.S., a conservative estimate, and around 80 percent of voiceover work is nonunion, Friedlander said.

Human and machine

Last summer, Rattner—who worked in software innovation at Qualcomm before co-founding Yembo—self-published a book called "Grow Up Fast: Lessons from an AI Startup." It's an entrepreneurship memoir about how he helped build Yembo, a company that uses a subset of AI called computer vision to make tools for the moving and insurance industries.

The book's Spanish translation will be released this month, followed by Ukrainian and more than 10 other languages. All could come out within months—with time built in for tweaking and revising, Rattner said.

AI narration "definitely brings the barrier of entry down for people who wouldn't have been able to get their message out," he said.

He broke down the time and money costs of human and machine. The English audiobook took about four weeks to record. (Hansard could only record on weekends and her vocal cords needed breaks.) "Factoring in

mastering, editing, QA listening, and retakes, I'd estimate the US English audiobook took about 65 man-hours of work across all parties to create," he wrote.

Next, they used three hours of her book recording to train an AI tool called a speech synthesis model and used that model to create the other books in translation.

Not counting translation by humans (Rattner hired people to write translations, because "AI translation makes funky mistakes in unpredictable ways"), each AI audiobook narration takes five hours, with the bulk of that spent on quality assurance—weeding out mistakes like reading 2nd not as "second" but as "two-en-dee."

The dollar difference is more staggering: A human narrator might charge a few hundred dollars per recording session or perhaps $2,500 for an audiobook, he estimated. The voice synthesis software costs $22 a month.

A fair contract

The narrator didn't have to do extra work to create 15 translated books, but the publisher didn't have to go out and hire 15 other narrators. When part of an audiobook's production is outsourced to AI, what payment is fair to creator and publisher?

This contract, which Rattner shared with the Union-Tribune, attempts to minimize losses to one human worker while maximizing the benefits of AI, which for audiobook translations include expanded access to information. Every time a translation of "Grow Up Fast" sells, the narrator will earn money—even though she never recorded in those other languages. So will the publisher, who used AI to narrate translations at a fraction of the cost of using a human actor.

- Hansard was paid $500 per four-hour day of studio recording and gets 10 percent royalties on translated works that use her cloned voice. Payments are quarterly over a 10-year term.
- Her cloned voice can only be used for this book's translations. Other uses require a new license.
- The narrator gets 30 days to review the product, including translations, and ask for edits before it goes live.
- The publisher can sell the book at any price and do giveaways.

One section covers labeling. "The use of AI must be disclosed in product markings," Rattner said. This way, readers or listeners will know if the audiobook was "Narrated by Hailey Hansard" or "In the voice of Hailey Hansard."

Other actors who reviewed the contract's key points called it "encouraging" and said it appears generally fair to both parties, though some shared reservations.

All agreed the narrator should get royalties. The publisher is making a greater profit by using AI instead of human actors, and future narrators are losing out on potential income because of AI, Sciglimpaglia said.

"They just have one person to read in one language and they can use a machine to convert it for nothing," he said.

Friedlander likes that the contract addresses consent, control, compensation and transparency. But he said even an equitable contract raises questions about precedent being set.

"This one voice actor gets to do all of these different languages," he said. He mentioned the "damage it's done to all of the other narrators who would have done this, in those different languages."

Some day there might be "a handful of four or five narrators who become the voice of everything," he said. Audiobooks in particular are "one of the places that a lot of people get their start" in voice acting. If the norm becomes synthetic voices, how will those new people get started, he asked.

Conde wondered why royalties stop after 10 years. "Does the contract drop off and her voice can be used anywhere?" she asked. "I would be worried about what happens after the 10-year clause."

Wendy Hovland, a San Diego voice and on-camera actor, said the time limit can help the narrator renegotiate. She also said the publisher "appears to be working openly with her, to tell her how it's going to be used and come up with a way to compensate that works for both parties." Voice actors don't always get that, she said.

"That is a big issue: voices being—I don't know if 'stolen' is the right word, but used in a way that was not originally intended. Voice actors thought they were voicing one thing and found out that their voices are used for something else," she said.

Hansard feels "very protected" by the contract because it forbids other uses for her cloned voice without her OK.

"Like other actors and creators, I do worry about being exploited by AI. But this particular agreement was win-win. Zach was very receptive to taking care of all of the concerns I had," Hansard said.

AI audiobook as proof of concept

To understand why Rattner prioritized creating a fair contract with the narrator—what's it in for Yembo—it helps to understand what Yembo sells. Yembo's software scans the insides of homes and creates

inventories and 3D models for moving, storage and insurance reconstruction estimates.

The biggest challenge to signing new customers has not been competitors but resistance to change, Rattner said. In an industry where using a typewriter is still feasible—as one moving company he encountered does—how will they trust new technology, whether or not it's AI? If things have worked fine for decades, why risk it?

His solution: prove that AI can be used for more than profit.

"I find the business arrangement just as interesting as the book itself. I ... think it's an interesting story about how AI can be used for good, especially with all the anxiety around AI actors," Rattner wrote .

"AI allows for economic value to be tied to the output created, not the effort exerted (e.g., time for dollars)," he added.

Rattner said he wouldn't have pursued the foreign language narrations without AI, given that his job is running a tech startup and not a publishing house. He found the narrator from within Yembo's ranks: Hansard is a product manager employed by Yembo and a former professional actor. She is SAG-eligible, but not a union member.

"The alternative (to AI) was nothing," he said. By nothing, he explained, he meant no translations and no hiring narrators in various languages.

In an interview from Los Angeles, Hansard talked about the uncanniness of hearing her vocal clone. This is her first audiobook, both in English and in translation.

"It's almost jarring to hear my voice speaking languages that I've never spoken before, but also amazing that this possibility exists," she said.

She was comfortable with the project because she was assured it was not taking work from others.

"I think the best outcome would be that AI doesn't replace human actors or human voices," Hansard said. "It only supplements if it wouldn't have been possible without it."

She continued, "I think that's where everyone is going to have to reach into their humanity to make sure that AI doesn't replace humanity. That it only enhances—if something wasn't going to be possible, then it fills the gap."

2024 The San Diego Union-Tribune. Distributed by Tribune Content Agency, LLC.