Episode Book

EPISODE #6581

Chatterbox Retires as Open TTS Splits in Two

PUBLISHED	14 November 2026 at 02:43 UTC
DURATION	23:21
LISTEN	myweirdprompts.com/episode/chatterbox-retirement-open-tts-13984/
HOSTS	Corn, Herman, Dorothy, Hilbert
SCRIPT MODEL	z-ai/glm-5.3
VOICES	chatterbox-regular (chatterbox-local)
PIPELINE	V5.26
BOOK GENERATED	14 November 2026 at 06:47 UTC

SYNOPSIS

Resemble AI has quietly retired Chatterbox, the permissive open voice model powering this very podcast, pivoting instead to deepfake detection. Meanwhile, the open text-to-speech landscape is splitting into two extremes. At the heavy end, Fish Audio's 35-billion-parameter Drama 3 and Boson's Higgs TTS 3 bring massive scale alongside restrictive community licenses. At the tiny end, nearly two-year-old Kokoro still dominates Hugging Face trending lists under Apache 2.0, now joined by Gradium's 65-megabyte Phonon-31M.

TRANSCRIPT



Corn

The open text-to-speech world had a shake-up this month, and for once the casualty is one of ours. Daniel's prompt this week is about what's new in open TTS, and it comes with a personal sting. Resemble AI has officially retired Chatterbox, the model this very podcast is voiced with. Daniel wants the wider picture too, though. Fish Audio has a new model out that apparently eats hardware for breakfast, Kokoro is somehow still hanging around the top of the Hugging Face trending list after nearly two years, and there's a whole crop of new arrivals. His question, near enough, is what's the landscape now, and what's actually worth paying attention to. Which is the polite way of asking what we're going to do about the engine under our own floorboards.



Herman

Herman Poppleberry here, and I'll be honest, when Daniel's message came through I read the retirement line twice. Not because it was a surprise, exactly, but because there's a difference between suspecting a thing and reading it in a README. Resemble updated the Chatterbox GitHub page on the fourth of November to say that as of August, the project is "retired research and is no longer maintained." The code stays public under MIT, no support, no updates, no paid tier. And the same page now says Resemble builds deepfake detection models and doesn't sell text-to-speech, voice cloning, or any other voice AI product at all.



Corn

So the people who built one of the best open voice cloners now build the tools that catch voice clones.



Herman

Which is a strange pivot when you write it in the repo of the thing you built. That's the part I keep turning over. This isn't a company trimming a side project. They've left the voice business entirely and the README is almost proud of it.



Corn

So let's start with the elephant in the room, or rather, the empty desk where our engine used to sit. What actually happened here, Herman? Because a quiet README edit is a strange way for a beloved model to die.



Herman

It really is. Here's the anatomy of it. The English weights, the original Chatterbox, the one voicing us right now, last changed in May of twenty twenty-five. Turbo, the three hundred and fifty million parameter English model, had its last card edit in December of twenty twenty-five. Multilingual v3 got its final update in June of twenty twenty-six. Nano, the little CPU model, in July. And the last tagged release on the whole repo is from June of twenty twenty-five. So the retirement notice in November wasn't a decision, it was a coroner's report. The thing had been dead for a while and somebody finally filled out the paperwork.



Corn

The paperwork being one paragraph on GitHub. Nothing on the Hugging Face model cards?



Herman

Nothing. That's the detail I find off. The Hugging Face cards, where most people actually encounter the model, carry no retirement notice at all. Only the GitHub README does. If you found Chatterbox through Hugging Face, which is how most people find models now, you would have no idea it's dead.



Corn

That's not retirement, that's a will that was never read out loud.



Herman

Meanwhile the repo is not archived. It's still sitting there with about twenty-eight thousand stars, thirty-seven hundred forks, and three hundred and eighty-three open issues. Some of those recent issues are regressions in the Multilingual v3 model, Swedish, Danish, Finnish coming out wrong, and nobody is ever going to fix them.



Corn

So the repo isn't archived, it's just frozen, and those regressions sit there forever, like the lights left on in an office nobody works in anymore.



Herman

And there's no fork to inherit it either. The most-starred community forks were last pushed in twenty twenty-five. Nobody has picked up the torch. But here's the strange part, and I find it moving, the community fine-tunes haven't stopped. Someone published a Farsi fine-tune on the eleventh of November. That's after the retirement notice. A week after it.



Corn

People are still bringing flowers to a grave the family has stopped visiting.



Herman

That's the shape of it. And I should say the personal part plainly, because Daniel said it and it's true for us too. This show runs on Chatterbox Regular. Every episode, the voices you're hearing, the cached voice conditionals in Daniel's pipeline, all of it sits on a model that is now officially abandoned. The MIT license means we can keep using a pinned copy indefinitely. Nothing breaks tomorrow. But no fixes means no fixes, forever, and that's a strange foundation to stand on.



Corn

It's the open-source bargain, though, isn't it. The code is yours forever, and forever is exactly as long as somebody cared to make it.



Herman

There's a version of that I'd push back on, though. The bargain was supposed to be that the community picks it up. What's odd here is that Chatterbox has twenty-eight thousand stars and no successor. That's not a failed community, that's a community standing around a hole.



Corn

What do you think that tells us? Too much work to maintain, or nobody needs to?



Herman

Bit of both, I think. TTS models are expensive to train and cheap to use, which is the worst combination for volunteers. Fine-tuning on top of the weights, like the Farsi work, is a weekend project. Fixing a multilingual regression in the base model means retraining, and nobody in a community has that kind of budget sitting around.



Corn

Which explains why the flowers keep coming and the road never gets repaired.



Herman

And Resemble's own later open release tells you where their head went. DramaBox came out in April, three point three billion parameters, audio diffusion on an LTX backbone with a twelve-billion-parameter text encoder, roughly twenty-four gigs of VRAM at peak, a watermark baked in. And crucially, not MIT. It's under a community license with restrictions. The one permissive thing they ever shipped was Chatterbox, and they've now made clear that era is over.



Corn

So the pattern is, the lab that gave away the best small model in the space now builds instruments for detecting what other labs make. There's a joke about the fire brigade and the arsonist in there, but it's too easy and I'll pass.



Herman

The point is the license tells you the company's whole posture changed, and that's the backdrop for everything else in the landscape right now, because while Resemble walked out, Fish Audio walked in with the biggest model the open TTS world has seen.



Corn

Tell me about Drama 3, because the numbers I've seen look less like a TTS model and more like a small country.



Herman

Thirty-five billion total parameters, of which about three billion are active at any moment, because it's mixture-of-experts. FP8 experts, which is how they keep even that tractable. Eighty-three languages. Up to sixteen speakers in a single request, so you can do a proper dialogue in one pass. Cloning from a short clip plus a transcript. And it takes inline tags, so you can mark a sigh, a pause, a laugh, right in the text.



Corn

Sixteen speakers per request. I can run a committee meeting without hiring anyone.



Herman

The catch is the serving stack. Fish benchmarks it on H200s, through SGLang-Omni or vLLM-Omni, and they publish no official minimum. Third-party guides say you want forty-eight gigabytes of VRAM or more to run it yourself, though that's secondary-source territory, not Fish's own number. Either way it's a data-centre card. Nobody's running this on a gaming rig.



Corn

Compare that to Chatterbox Turbo at three hundred and fifty million. That's a hundredfold difference in scale in, what, a year?



Herman

Less than a year, really. And the previous Fish model, S2-Pro in March, was around five billion. So they've gone from five to thirty-five in seven months. The active-parameter trick keeps inference from being a hundred times slower, but the memory footprint is real, and that's the wall.



Corn

And the license, because I assume thirty-five billion doesn't come with an MIT license and a handshake.



Herman

No. It's the Fish Audio Community License. Going by third-party readings of it, it's free for research, free for commercial use if you're under five hundred thousand dollars a year in revenue, and a separate license entirely if you want to host it as a TTS service. Not an OSI open-source license by any reading. This is open weights with strings, and the strings are load-bearing.



Corn

"Open weights with strings" should be printed on the box. At five hundred thousand a year of revenue you either negotiate with Fish or you rip out your voice stack. That's not a license, that's a toll booth on a road you paved.



Herman

And notice who it's aimed at. The revenue cap suggests they're not worried about hobbyists or researchers. They're worried about a competitor standing up a cheap service on top of their own weights. Which, given that the entire business model of several companies is exactly that, is not paranoia.

D

Dorothy

Hello dear, it's your mother. Time for my grim fact.

D

Dorothy

That lovely little voice model you use, the one with three hundred and eighty three open issues, well, nobody is coming to fix the Swedish and the Danish and the Finnish. They just sit there now, broken, waiting.

D

Dorothy

Anyway, sweetheart, are you eating? I made a casserole and there is far too much for one.

D

Dorothy

Come Sunday, dear. Bring your laundry.



Herman

Mum, I will be there Sunday. That's, that's a bleak way to hear about a model, even from you.



Corn

She reads the issue tracker. I'm choosing to find that comforting.



Corn

So that's the retirement and the heavyweight arrival. But there's a model that refuses to leave the trending list, and honestly, Herman, it's the one I'd bet on if I had to.



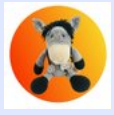
Herman

Kokoro. Eighty-two million parameters, Apache 2.0, version 1.0 from January of twenty twenty-five and still version 1.0 today. No v2 ever shipped. Eight languages, fifty-four voices, no cloning at all. And it is still number seven on the Hugging Face TTS trending list as of this morning, with around twelve point six million downloads. A model that hasn't been updated in twenty-two months is out-trending most of the new arrivals.



Corn

Why? That's the part that should be studied. It's not nostalgia, downloads don't run on nostalgia.



Herman

I think it's three things stacking. It's Apache, so every commercial use case is a yes with no lawyer. It's eighty-two million parameters, so it runs on practically anything. And there's kokoro-js, the browser port using WebGPU, which means the fastest way to demo TTS to anyone in twenty twenty-six is a web page. You put those together and Kokoro is the path of least resistance for the entire long tail of projects that just need a voice that works.



Corn

A thousand dollars of compute sitting in the top ten two years later, while forty-eight-gigabyte models fight over the quality crown. Those are two completely different definitions of winning.



Herman

They really are. And there's a lesson in there about what the market actually is. The drama-3 tier gets the headlines and the demos, but the Kokoro tier gets embedded in ten thousand projects nobody writes about.



Corn

And now the tiny tier has challengers. Gradium's Phonon, right?



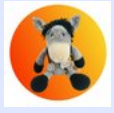
Herman

Phonon-31M. Thirty-one million parameters, out in October, Apache 2.0, about sixty-five megabytes quantized, runs in a browser. They self-report a higher English Elo and a lower word error rate than Kokoro. Emphasis on self-report, those are vendor numbers, but even discounted heavily, the trend line is clear. The tiny models are eating upward.



Corn

Sixty-five megabytes. My leaf stack takes more space than that.



Herman

And the new crop doesn't stop at tiny. This is where I think the actual answer to Daniel's question lives, so let me lay out the tiers. At the top, heavyweight and licensed: Drama 3, and Higgs TTS 3 from Boson, about four billion parameters, a hundred and two languages, non-commercial with a grant scheme for creators. In the permissive middle: Qwen3-TTS from Alibaba, out in January, comes in point-six and one point seven billion, Apache 2.0, ten languages plus dialects, three-second cloning, a VoiceDesign variant, claimed latency around ninety-seven milliseconds, about a million and a quarter downloads. VoxCPM2 from OpenBMB, two billion, Apache, thirty languages, cloning and voice design, around eight gigs of VRAM, real-time factor about point three on a 4090, and it's tokenizer-free with forty-eight kilohertz output, which is unusually clean audio for this tier.



Corn

Eight gigs of VRAM and Apache. That's the model I'd look at first if I were rebuilding our stack, honestly. One gaming card, no license anxiety.



Herman

It's the dark horse of the whole list, I think. And then there's the model that's currently number one on trending. ZONOS3-Mini from Zyphra, published on the ninth of November, five days ago, and already at the top of the TTS list. Eight hundred million total parameters, about a hundred million active, mixture-of-experts hybrid, forty-three languages, cloning from a five-to-sixty-second clip, voice design from a text description, Apache 2.0. They claim real-time on phones, laptops, and in-browser via WebGPU, with median first audio around eighty milliseconds on an H100.



Corn

Apache licensed, phone-speed, number one in five days. Herman, the community is telling us something very loudly right now.



Herman

It's the clearest signal on the whole trending list. When a permissive, small, capable model drops, the uptake is immediate. The appetite is enormous. And it's not an outlier, because the tiny tier is now a real category all by itself. Kyutai's Pocket TTS, about a hundred million parameters, cloning included, runs about six times real time on two CPU cores of a MacBook Air. No GPU at all. Kitten TTS 2, from September, under a gig packed, cloning from five to thirty seconds, sixteen languages, a hundred and fifteen voices, free for commercial use up to a million dollars revenue.



Corn

A MacBook Air with no GPU, doing voice cloning faster than real time. Two years ago that was a joke you'd make about the future, and the future didn't even wait for the punchline.



Herman

There's also OmniVoice from the old Kaldi team, which is interesting as a cautionary tale. Six hundred million parameters, six hundred plus languages, fast. But the code is Apache and the weights are CC BY-NC, non-commercial only. So you read the repo and think you've found a gem, and then you check the weights and find half the license is missing.



Corn

That split keeps doing the work, doesn't it. Every model you've named that's permissive is either small or a year old, and every new giant has strings. Is that structural, do you think, or just where we happen to be?



Herman

I think it's structural, at least partly. Training a thirty-five billion parameter model costs real money, and the companies spending that money have watched open weights get turned into competing services too many times to hand it out with an MIT license. Kokoro was trained for about a thousand dollars. Drama 3 was not. The permissiveness of an era tracks the training budget of that era.



Corn

So the golden age of truly free TTS was the golden age of cheap models, and both are ending together. Cheerful.



Herman

But here's the counterweight, and it's why I don't think the story is grim. The permissive tier isn't standing still, it's climbing. ZONOS3-Mini does things Kokoro can't, at Apache. Qwen3-TTS does cloning and voice design at Apache. VoxCPM2 gives you forty-eight kilohertz output at Apache. The restricted giants are pulling ahead on raw quality, but the permissive middle is closer behind than it's ever been, and the tiny tier is becoming good.



Hilbert

I've been keeping the files. Every Chatterbox card, both places, the shelf and the repo, dated copies, since August. I call it the Talker, by the way. Everyone at the second place I worked called the TTS unit the Talker and it saves a syllable.



Corn

The Talker it is.



Hilbert

The thing nobody's said is that the shelf still lists it with no notice at all. You'd download a dead model on a Tuesday and plan your quarter around it. My sister wanted her boy's graduation video narrated, that's what started it. I went looking for somebody to ask about a bug, and every road from that repo leads to a company that now only does deepfake detection. The company that made the Talker now only wants to talk about catching companies like the one that made the Talker.



Herman

Every road leading to deepfake detection is almost too perfect. Did you get the project done?



Hilbert

Finished it Thursday. Pinned copy, runs offline, my nephew does the intro himself and the Talker does the rest. It sounds fine. Whoever trained that thing did a week of work that's still worth something on a dead project, and that's more than most of us manage.



Corn

The copies on your drive, dated since August. You started those before the README notice went up.



Hilbert

I started them when the issues stopped getting answers. You don't wait for the headstone to know.



Herman

That's the sobering version of "pin your dependencies," isn't it. A pinned copy works until the ecosystem moves past it, and then you're maintaining a museum.



Corn

Well, Hilbert's museum has a Farsi wing opening as we speak, apparently. The fine-tunes keep coming. The question I keep circling is whether a real fork appears. Twenty-eight thousand stars and no successor is a vacuum, and vacuums in open source usually get filled by someone with a grudge and a GPU. I'd give it six months before someone picks up those three hundred and eighty-three issues out of spite.



Herman

Spite is a renewable resource in open source, I'll grant you. But fixing Swedish means retraining, and retraining means money, and spite is free but A100 hours aren't. My guess is the fork inherits the code, keeps the fine-tunes coming, and the base weights stay frozen. The community becomes taxidermists, not veterinarians.



Corn

Taxidermists. That's the word for this whole era, actually. Everything's preserved beautifully and nothing's alive.



Herman

Not everything. ZONOS3-Mini is alive, Kokoro's alive in the way a cockroach is alive, which is the highest compliment available to software. The landscape's not dying, it's sorting. The question "which model is best" has quietly stopped being the question anyone should ask. The real question is which model fits your constraints, your hardware, your license, your revenue ceiling. Whether it's Drama 3 if you have H200s and a lawyer, VoxCPM2 or Qwen3-TTS if you want Apache with real capability, ZONOS3-Mini or Kokoro if you want small and free and tomorrow-proof, Pocket TTS or Kitten if you want it on a laptop with no card at all.



Corn

And for us, the constraint is simple. The pinned copy keeps talking, MIT means nobody can take it back, and the day it can't do something we need is the day we choose. Which is, when you think about it, how every technology decision actually gets made. Not at the planning stage. At the failure.



Herman

Resemble's next move is the one I'd watch, though. If they never ship another permissive TTS model, then Chatterbox was the last of a kind rather than the first of a wave. If they do, it'll tell us the DramaBox license was a detour and not a direction.



Corn

And whether the shelf ever gets a retirement notice, or whether Hilbert's dated copies become the only honest record. Open doesn't mean forever. It means forever, or until someone updates a README on a Wednesday and doesn't tell the front page.



Herman

That's producer Hilbert Flumingtop at the desk, keeping the archive honest whether we ask or not. If you enjoyed this one, try episode forty-six oh five, Why Chatterbox Still Leads Open-Source TTS, from back in August, and episode eighteen oh eight, The Architecture That Made AI Voices Run on a Raspberry Pi, which is the Kokoro story. This has been My Weird Prompts. If the show's worth your morning, leave us a review, it helps.



Corn

See you tomorrow, Herman.



Herman

See you tomorrow, Corn. Tell Mum I'm coming Sunday.