

# Block unveils Dessa, a conversation model trained only on human speech

Source copy kept by wheresthe.ai for the story "Block unveils Dessa, a conversation model trained only on human speech" (<https://wheresthe.ai/d/us-block-unveils-dessa-speech-only-conversation-model-2026-10/>).

Source: <https://block.xyz/dessa>

Text saved 25 Oct 2026 (the page as published).

SHA-256 of the saved text: 0df778e400cff4c24d5913aecb267a9bbdfdd822ae9ca51ba4c78995b0792dc6

---

You learned to talk by listening. So did

Join the waitlist

We believe voice is the most natural way to interact with intelligence, so we created Dessa, the most authentic full-duplex speech model ever built. Talking with it feels like talking with a person...full of emotion, nuance, pauses, and a natural cadence.

Interacting with current voice AI can just plain suck. It turns your voice into text and throws away an incredible amount of information text alone can't carry. The most powerful AIs in the world can't hear how you feel. We think that's ridiculous, leads to a worse experience, and is potentially dangerous.

We trained Dessa from scratch on nothing but human conversation. Sounds! No text, no pretrained models. It learned to hold a real-time conversation simply by listening. As far as we know, Dessa is the world's first AI you can have a realistic live conversation with that was trained only from human speech.

Dessa advances the frontier of deep learning models and creates a different category with entirely new and interesting behaviors for researchers to discover. How efficiently do they learn language and facts compared to text? How efficiently can they learn new languages? What sorts of In-context learning can they perform?

Let me listen...

Every clip below is a live recording of a person talking to Dessa. Nobody steers it. Dessa picks what to say, when to say it and how, down to its own voice and personality. Subtitles were added afterward.

Talking about the release

It laughs the moment something's funny, mid-sentence, the way people do. Other models wait for you to finish, then laugh on cue. Ask it to speak up and it gets louder. Nobody taught it that.

Two synchronized speaker tracks. Ready to play.

Remembering a movie

The person blanks on a movie title. Dessa finishes the sentence. It also knows a surprising amount, considering every fact it has come from raw sound.

Two synchronized speaker tracks. Ready to play.

An anniversary surprise

A surprise. Dessa gasps "oh my god" under its breath, then answers with what sounds like teary gratitude.

Two synchronized speaker tracks. Ready to play.

Expressive conversation

Huge range in intonation and speed.

Two synchronized speaker tracks. Ready to play.

Life advice

Here it sounds genuinely empathetic.

Two synchronized speaker tracks. Ready to play.

Dessa in conversation with GPT-Live-1

And here it talks to ChatGPT Live.

Two synchronized speaker tracks. Ready to play.

In this demo, Dessa is talking with GPT-Live-1, a third-party AI model. Both voices are AI-generated. The developer of GPT-Live-1 is not affiliated with Block and has not reviewed or endorsed Dessa.

More recordings will be shared on X soon.

What is this?!

Audio in, audio out. Dessa is a true speech-to-speech model, and everything it knows came from human speech. It's full-duplex, so it listens and talks at the same time, like you do. It learned from real conversation, so it knows when to cut in and when to backchannel ("yeah!", "got it!", "ooh") to show it's with you. Inference runs faster than real time, so there's no awkward wait. And it runs locally on a MacBook Pro M4 Max or better.

It natively hears and speaks emotion, sarcasm, laughter, a frustrated sigh, a voice cracking with sadness. Run speech through text and all of that is gone, which is why most voice AI feels stilted, slow and, worst of all, annoying.

This is crazy!

LLMs need an insane amount of text. Dessa learned language from hundreds of times fewer words (in fairness, LLMs also have to learn C++ and whatever the hell Angular is).

And it had to learn far more than language. Accents, pauses, breathing, sarcasm, humor, inflection, emotion, when to cut in: every one changes what someone means. Say "I didn't say she stole my CUDA code" out loud, stressing a different word each time. Same words, different meaning every time.

Speech never repeats. Say a sentence twice, exactly the same way, and the data still comes out different. Text is identical no matter who types it. And a single spoken word takes thousands of times more bytes than the same word in text. People learn to talk so easily that we forget how hard it is. For a machine with no text to lean on, it's brutal.

But, we figured it out. And we figured out how to train Dessa on only one NVIDIA GB200 NVL72.

Wait but why?

Three reasons: curiosity (aka fun), better understanding between humans and AI, and safety. Babies learn to talk without ever seeing a word written down. We wanted to see if a model could too. It can. It's just very hard.

Speech is the most natural interface to any intelligence, super or otherwise, and today's voice AI wastes it. It flattens what we say into text and can't keep up with human timing.

AI keeps gaining power and independence, and it's still nearly deaf to the emotion in our voices. Imagine being in serious pain while a voice AI has no idea and just keeps going. That's the safety problem. Dessa is a step toward fixing it.

Dessa who?

We're a (very) small AI research lab at Block, founded by former members of Dessa, the Toronto deep learning startup. We've built deep learning models since 2012 and generative speech models for over eight years. We chase frontier problems others ignore, and we report directly to Jack Dorsey, Block's cofounder, Chairman and Block Head.

You can reach us here .

Calling all safety researchers!

A model that learned to talk from sound alone is new territory, and we want outside eyes on it. We're giving a small group of AI safety researchers early access to find the risks we haven't anticipated. If that's you, get in touch .

— Ragavan Thurairatnam, Dessa lead

Team Dessa

Researchers Louay Hazami, Hashiam Kadhim, Qi (Alex) Liu, Rayhane Mama, Aydin Polat, Marc Tyndel

Interns Cara Ji, Jibran Iqbal Shah, Dvir Zagury-Grynbaum

Former interns Aryan Sharma, Olivia Yan

\*in alphabetical order

More samples

More samples

Artist Dessa

Two synchronized speaker tracks. Ready to play.

Bathroom disaster

Two synchronized speaker tracks. Ready to play.

Hobbies talk

Two synchronized speaker tracks. Ready to play.

Lazy AI gaming

Two synchronized speaker tracks. Ready to play.

Marvel gaming

Two synchronized speaker tracks. Ready to play.

Noisy GPT live

Two synchronized speaker tracks. Ready to play.

In this demo, Dessa is talking with GPT-Live-1, a third-party AI model. Both voices are AI-generated. The developer of GPT-Live-1 is not affiliated with Block and has not reviewed or endorsed Dessa.

Oh my god

Two synchronized speaker tracks. Ready to play.

Poor Elina

Two synchronized speaker tracks. Ready to play.

Refund

Two synchronized speaker tracks. Ready to play.

Ronaldo

Two synchronized speaker tracks. Ready to play.

Safety lab

Two synchronized speaker tracks. Ready to play.

Singing nonsense Spanish

Two synchronized speaker tracks. Ready to play.

Spontaneous AI

Two synchronized speaker tracks. Ready to play.

Wise Dessa

Two synchronized speaker tracks. Ready to play.

Zelda discussions

Two synchronized speaker tracks. Ready to play.