The Best AI Conversation I Have Heard Is Not Powered by the Best Model

Sesame’s demo may be the strongest public AI conversation experience I have used as of March 24, 2026, but the public evidence suggests the effect comes less from a visibly unique text-model breakthrough than from the voice layer and the way the system is assembled.

2026.03.24 01:52 DENNIS HEDEGREEN INVESTIGATE OPEN-NOTE

HTTPS://HEDEGREENRESEARCH.COM/ARTICLES/THE-BEST-AI-CONVERSATION-I-HAVE-HEARD-IS-NOT-POWERED-BY-THE-BEST-MODEL/

This is the second Sesame investigation.

The first asked a narrower question: what has Sesame actually shipped?

As of March 24, 2026, the answer was still limited. A public demo. A partially open voice model. A locked beta app. A company with major funding and very little broadly released product surface.

This investigation asks a different question.

What is actually doing the work inside the demo that impressed so many people?

The answer may be more interesting than Sesame’s public story around it.

What Everyone Heard

The strongest thing Sesame has built is not subtle.

People heard it immediately.

A voice that sounded less like a machine and more like a person.

Natural pauses.

Breathing.

Hesitation.

Laughter that landed in roughly the right place.

That part is real.

Firsthand observation: I have used Sesame’s public voice demo repeatedly since its release and continue to use it daily as of March 24, 2026. It is the strongest public AI conversation experience I have personally used. Public reviews and technical coverage point in the same direction. Sesame’s own research page also frames the project around what it calls “voice presence,” the attempt to make spoken interaction feel genuinely natural rather than merely intelligible.

But there is a simpler question sitting behind all of that:

What is doing the thinking?

Because the voice is not the brain.

The voice is the output layer.

Something behind it is deciding what to say.

What Sesame Says Publicly

On its official research page, published February 27, 2025, Sesame describes CSM, the Conversational Speech Model, as the system responsible for the speech side of the experience. It explains that the model jointly handles text and speech tokens and that the company trained three sizes: `1B + 100M`, `3B + 250M`, and `8B + 300M`.

That same official material also makes an important limitation clear. Sesame writes that while CSM generates high-quality conversational prosody, it can only model the text and speech content in a conversation, not the full structure of conversation itself. The company points toward future “fully duplex models” that would learn turn-taking, timing, and pacing more deeply.

Sesame’s GitHub page is even more direct. It states that CSM is an audio generation model, not a general-purpose multimodal LLM, that it cannot generate text, and that users should pair it with a separate LLM for text generation.

That matters.

It means the part people are reacting to most strongly is not, by Sesame’s own description, a standalone conversational intelligence system.

It is the speech layer.

What Appears to Sit Behind It

The public evidence suggests that the text model behind the demo is separate from CSM.

Firsthand observation: across repeated direct conversations with the public Sesame demo from its release through March 24, 2026, I have asked the system about the model behind the conversation and received answers identifying it as Gemma 27B. I treat that as product-surface evidence, not as formal technical documentation.

Secondary reporting points in the same direction. The Decoder reported in March 2025 that Sesame’s demo used a 27B version of Google’s Gemma as the text model behind the experience.

That does not amount to a full official architectural disclosure from Sesame itself.

But it is enough to support a narrower and more defensible conclusion:

the visible evidence points toward a system in which Sesame's distinctive contribution is the conversational speech layer, while the text-generation layer appears to come from a separate model that has not been publicly documented as a uniquely proprietary frontier breakthrough.

That last clause is important.

This article is not claiming that Gemma 27B is weak.

It is not weak.

It is also not claiming that Sesame's full stack has been formally documented in public.

It has not.

The narrower claim is that the available public record does not show that Sesame's effect depends on a uniquely documented text-model breakthrough comparable to the strongest closed-model systems.

That is a much more interesting finding anyway.

What Sesame Actually Opened

Sesame open-sourced `csm-1b`, the smallest public CSM variant, through GitHub and Hugging Face.

What it did not open-source includes the larger `3B + 250M` and `8B + 300M` versions described on its research page, the production-quality voice fine-tuning that appears to shape the public demo, and the full text-generation stack behind the conversational experience.

That distinction matters.

The public release gives outsiders a meaningful look at the speech layer.

It does not give them the whole deployed system.

Why This Matters

This matters because it shifts where the breakthrough seems to be.

If the strongest public AI conversation experience I have used does not come with clear public evidence of a uniquely documented frontier text-model breakthrough, then the effect is probably being driven disproportionately by the speech layer and the system integration around it.

That is the real surprise.

Not that the demo sounds good.

But that it sounds that good without public evidence that the underlying text layer is the main source of the leap.

Inference: the current Sesame experience suggests that conversational quality may depend less on having the single best underlying model than many people assume, and more on how voice, prompting, latency, and orchestration are assembled into one surface.

That does not prove model quality does not matter.

It clearly does.

But it suggests the bottleneck may not be where many people think it is.

The Bigger Point

Most AI discussion still treats progress as if it lives mainly inside a single model.

Which company has the smartest model.

Which benchmark moved.

Which release is best.

The Sesame demo suggests a different lesson.

The most important thing may not be any one component.

It may be the assembly.

A strong speech layer.

A competent text layer.

Good prompting.

Careful integration.

A system that makes the pieces feel more unified than they really are.

That is not a small detail.

That is a different theory of where the value is.

And if that reading is correct, then Sesame’s importance is not just that it built a memorable demo.

It is that it exposed how much can already be done by combining existing parts well.

What Comes Next

Sesame’s own research points toward larger and more capable multimodal systems in the future. If the company eventually closes the remaining gap between voice quality and the intelligence behind it, the result may be much stronger still.

But Sesame is not the only group that can see the shape of that gap now.

The architecture is partially visible.

The speech layer is partially open.

The integration pattern is more legible than it was before.

So the question is no longer whether this kind of conversational experience is possible.

Sesame already answered that.

The harder question is who understands the stack well enough to close the remaining gap first.

Follow the data, not the narrative.

— Dennis Hedegreen, follow the data

Method note

This article uses three evidence levels.

Verified public sources: Sesame research page for CSM architecture, stated limits, and model sizes; Sesame GitHub and Hugging Face pages for what was actually open-sourced and the statement that CSM is not a general-purpose text-generating model.

Firsthand observation: repeated direct use of Sesame’s public demo from release through March 24, 2026, including direct questions to the system about the model behind the conversation.

Secondary reporting: reporting that supports the Gemma 27B identification, treated as supportive rather than conclusive.

Inference: conclusions about where the main bottleneck currently appears to be, and what the Sesame demo implies about assembly versus raw model prestige.

— Dennis Hedegreen, follow the data

CANONICAL URL: HTTPS://HEDEGREENRESEARCH.COM/ARTICLES/THE-BEST-AI-CONVERSATION-I-HAVE-HEARD-IS-NOT-POWERED-BY-THE-BEST-MODEL/
PDF VIEW GENERATED: 2026-07-11T15:35:45.956Z