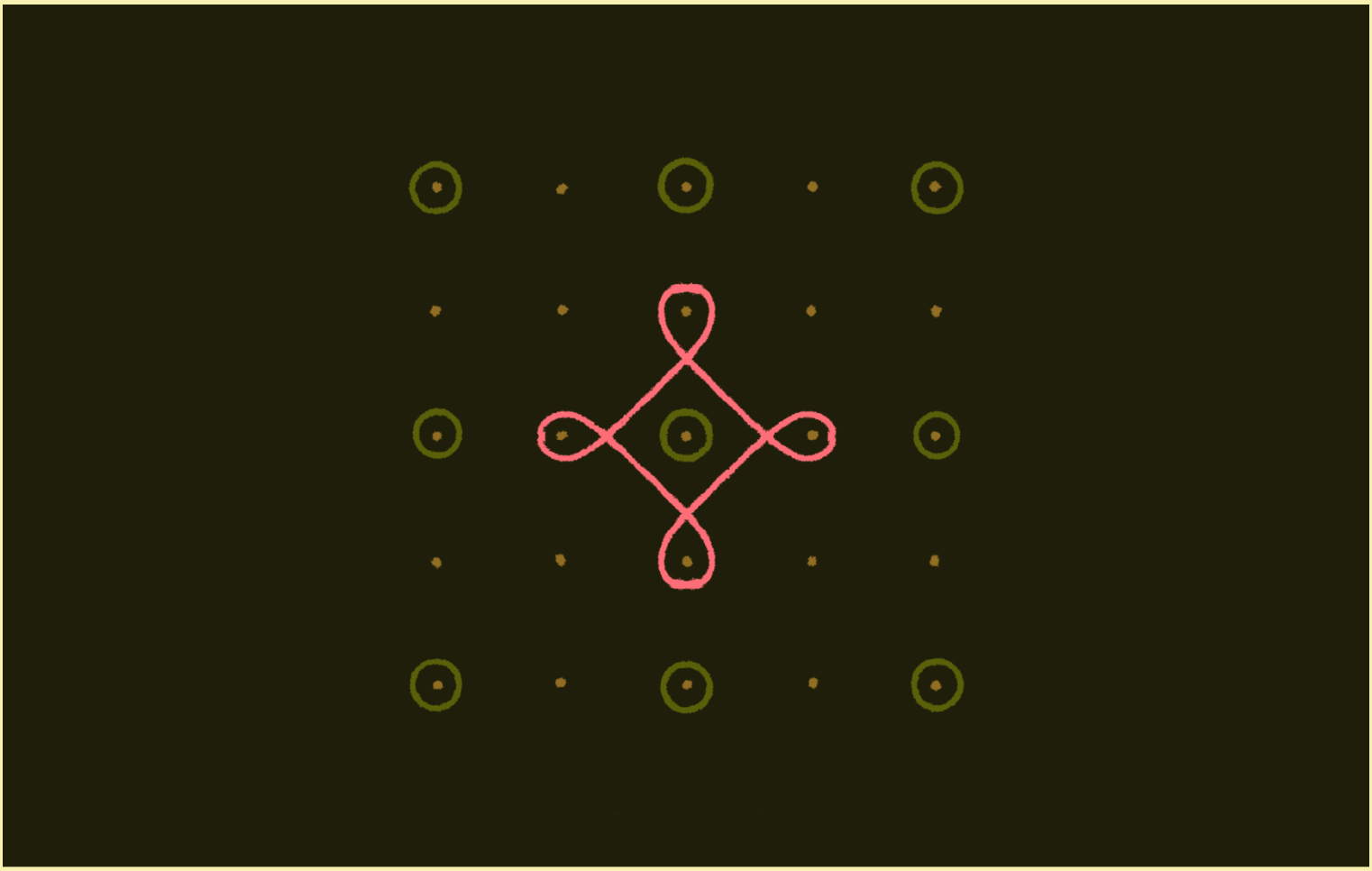
Reframing Impact:
AI Summit 2026

# Frugal AI

Timnit Gebru

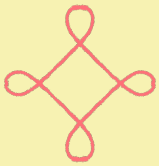
February 2026



AINOW aapti institute THE MAYBE

This piece is part of Reframing Impact, a collaboration between AI Now Institute, Aapti Institute, and The Maybe. In this series we bring together a wide network of advocates, builders, and thinkers from around the world to draw attention to the limitations of the current discourse around AI, and to forge the conversations we want to have.

In the run-up to the 2026 India AI Impact Summit, each piece addresses a field-defining topic in AI and governance. Composed of interview excerpts, the pieces are organized around a frame (analysis and critique of dominant narratives) and a reframe (provocations toward alternative, people-centered futures).



*Timnit Gebru is the founder and executive director of the Distributed AI Research (DAIR) Institute, an independent organization conducting community-rooted research. She was fired by Google in December 2020 for raising issues of discrimination in the workplace.*

*In this interview, Gebru critiques the dominant "one giant model" paradigm in AI, which prioritizes massive, ill-defined models and creates new problems in systems that previously served more specific purposes. She argues that this approach stifles innovation around potential resource-efficient approaches and results in subpar tools due to poorly defined tasks. The approach is in direct opposition to the engineering principle of building specific tools for specific contexts. Gebru advocates for stitching together "frugal AI" efforts, highlighting localized, community-rooted organizations—like those focused on low-resource languages—that curate data and use smaller models. She proposes that these organizations federate their resources and tools to collectively challenge the monopolistic power and resource-intensive practices of Big Tech.*

*Following is a lightly edited transcript of the conversation.*

FRAME: The “bigger is better” or “machine god” paradigm is creating new problems we haven’t seen before, and stifling our collective imagination of what more efficient models could look like and do.

**New problems are created by models that have ill-defined tasks and goals.**

We’re in the age of: Use as much data as possible, use as big models as possible, and don’t try to have one task that you work on—try to do everything all at once, try to create a digital machine god. That’s where we are.

I want to give an example of what this has done for speech recognition. Sure, there were issues with it. Sometimes the output would be erroneous. There would be some mistakes. But this idea of so-called hallucinations—I know not a good term—was not a problem with speech recognition. You didn’t have an issue where the output was a whole bunch of “made up” text.

OpenAI has something called Whisper, which is supposedly a speech-recognition model and translation between different languages. It’s one giant model that does a bunch of different things. Now we hear that doctors are using Whisper to transcribe patient notes, and it has been found to make up a whole bunch of things. Some examples were bizarre, like a speech that said “I think he was wearing a necklace,” and it’s transcribed to “He was holding a terror knife and he killed a bunch of people.” You’re not saving the data, so you don’t even know what the [original] speech is. This is an example of this “one giant model for everything approach” creating problems in things that we didn’t have before.

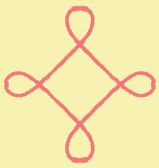
We’ve now been pushed to a paradigm that is ridiculous, that is never going to be safe because you don’t have a well-defined task. Like [with] speech recognition—the input is well constrained, the output is defined. We know what the relationship between the input and the output should be. With ChatGPT, we have no idea. It can produce any kind of output. It can take any kind of input.

This idea that we can just use one way of doing things for everything in the world, one giant model for everything has introduced problems we didn't even have before, and also results in subpar tools for many people around the world.

**Our vision for what's possible with resource-efficient models is constrained by dominant paradigms.**

I'm also not pro [the idea] that we need to do ChatGPT, but with fewer resources—hat's just not a good paradigm. Why do we need chatbots? We need to both build other things and do it with other resources. When DeepSeek came out, I had two conflicting feelings. One of them was, when you're constrained, that's when you innovate. When you feel like you have everything and you don't need to think about being efficient or clever, you're not going to innovate. So what's happening in the US is not innovation. On the other hand, DeepSeek was still this large language model paradigm.

Why is everybody's imagination hijacked?



REFRAME: We need localized efforts to be stitched together to have a chance at competing with dominant players, and a return to basic principles of building tools for a specific purpose.

**Smaller efforts are creating useful tools, despite incentives stacked against these efforts.**

Industry has absolutely no incentive to look at less resource-intensive things because they view their stealing of data as a competitive advantage, and the fact that they can outdo anyone with GPU [spend] or how big their data centers are as a competitive advantage. So there is not going to be any research in these kinds of places on how to do something efficiently—and then produce and give that research away to other people. That's not going to happen.

On the other hand, there is a parallel movement of different small organizations to do this kind of thing. One example that I think many of us talk about is Te Hiku Media in New Zealand. They gather data with Māori language revitalization in mind, and people are very excited to participate in that data gathering process.

This is very inspiring for me because one, they're not trying to build the one machine god that rules them all. They're only concerned about their own language and they're supporting other people in building their other tools. An American company wanted to license the data and they said, "Absolutely not, because everything we do has to serve the Māori people first." And they also said, "You guys beat this language out of our grandparents and now you want to sell it to us as a service."

I collaborate with a number of small language tech organizations—Te Hiku Media is just one of them. There is one called Lesan that focuses on Ethiopian languages. And there is another one called Ghana NLP that focuses on Ghanaian languages. And from speaking to them, I found that they had similar complaints.

One is when OpenAI or Meta or something comes with an announcement of a big model, a number of potential investors in these smaller organizations literally told them to close up shop.

Meta came out with an announcement of a model called No Language Left Behind and claimed there is one model that performs automatic translation across two hundred languages, including fifty-five African languages, for which there was before no state-of-the-art translation system. [Investors] were like, "Facebook has solved it, so your little puny startup is not going to be able to do anything."

Similarly, when they speak to people at OpenAI and other places, they basically threaten them by saying, "OpenAI is going to put you out of business soon because we're going to make our models better in your language. You're better off collaborating with us and supplying us data for which we're going to pay you peanuts."

Another thing that was happening was that potential customers would go to them and say, "Come back to us when you have more linguistic coverage." So your contextual knowledge is not about South Africa, it's about Ethiopia and that political situation and how different language groups are treated and what kind of politics is associated with which kind of ethnicity. But then what you're being pushed to do by both clients and investors is claim that you are either producing language technology for all African languages and even scaling globally.

My idea was to have a federation of these small organizations where they share data amongst themselves, but externally, there's like a kind of an easier interaction between them. So if you want language support in both Ghanaian languages and Ethiopian languages, you can interface with one application interface. You don't have to go to all these different organizations.

And perhaps they could also band together and present themselves as a really good competition to these bigger organizations. At the same time, each of them knows their context so well and they curate data so well, they're not just guzzling everything on the internet. They're really curating data and using low resource models.

**We need to focus on creating specific tools for specific contexts.**

I [also] don't think this is a language-specific problem. I think this is a paradigm that needs to exist for everything. Not even just so-called AI. When I'm building a transportation device, I'm not saying, "This thing is going to be used by one person, a hundred people, it's gonna

transport helicopters. And it's going to go by sea, by land, by any way you can imagine." We don't do things this way.

It is the same case with my specialty, computer vision. Unfortunately, a lot of bad things are part of computer vision. A lot of surveillance, face recognition, gait recognition, action recognition, etc. Which is unfortunate, but also things like plant recognition or even medical imaging, trying to analyze images and see what the likelihood of having a tumor might be. This is also part of computer vision.

You want to create a specific tool for a specific context. If you're interested in radiology, you should have data and figure out what specific thing you're trying to build the model for. You're not trying to just build something that you claim will be like a superhuman doctor. That doesn't make any sense.

To me, this idea of a task-specific tool is a concept in anything we build in engineering—it's not new. It's just that people came along and decided that they want to build a machine god and then claimed that they are doing it. And then they end up stealing data, killing the environment, exploiting labor in that process.

**There is power in pooling smaller resources together, and challenging the dominant paradigms.**

At DAIR, we're building our own cluster, which is like a small data center, because we don't want to use cloud computing resources. And we want to support our peers who don't want to use AWS or Google Cloud. We found that from our analysis, a one-time investment of $400,000, which is a lot, would give us an equivalent cluster that would have cost us almost $2 million per year to use in cloud computing services.

So the question that we want to answer—and some other people are thinking about it [too]—is, "How can we share resources across all of us and pool our resources to support each other?" This is a very different kind of paradigm than whatever Big Tech is pushing. And I'm not looking to Big Tech to push to a different paradigm, but there is an alternate ecosystem that's building around this idea that we don't have to do the same things that these people are pushing.

It’s not the frugal AI thing that’s new. It’s more the flip side of what we’ve been seeing that’s ridiculous and new. Let’s go back to basic principles.