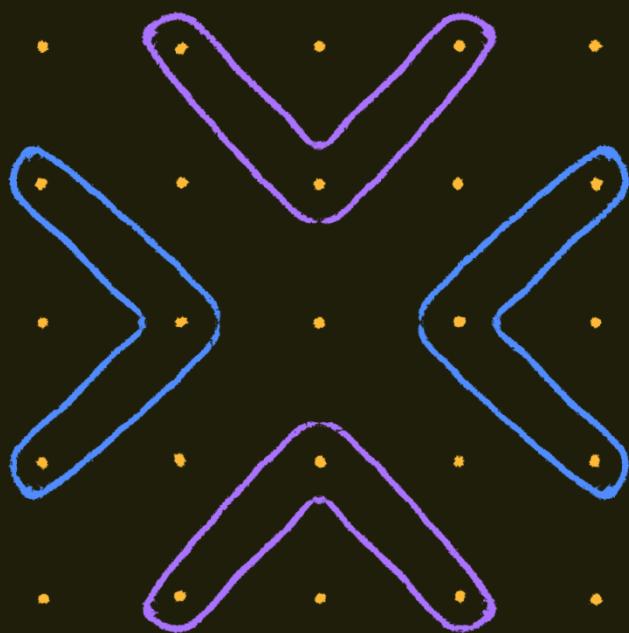


Reframing Impact:  
AI Summit 2026

# Open Source

Meredith Whittaker

February 2026



AI NOW



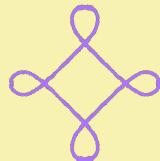
apti institute



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).



*Meredith Whittaker is the President of Signal. Her research and advocacy focus on the social implications of artificial intelligence and the tech industry responsible for it, with a particular emphasis on power and the political economy driving the commercialization of computational technology.*

*In this conversation, Whittaker unpacks “open-source AI.” In the context of software, “open source” referred to a set of precise technical protocols and processes that were arguably decentralizing. With AI, Whittaker says, this is not the case. Open-source AI, she argues, is not technical so much as vibes-based. The vast infrastructural consolidation of AI capabilities in the US and China is driving countries to embrace open-source AI—yet it is precisely the scale of this consolidation that renders open-source AI a false promise. In this context, Whittaker calls for a pragmatic return to a more technical understanding of open source—and questions power structures that sacrifice social life for short-term goals.*

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

**FRAME:** Open-source AI borrows and twists a technical concept developed in the context of software. It provides vibes-based reassurance in the face of industry domination—but abandons material reality.

**Open source AI doesn't do what its boosters claim – we have to disentangle promises from material reality.**

In the last few years the legacy understanding of the value of openness for technology and technological progress has been twisted. The term “open source” was applied specifically to software. Different licenses were created around what you can do with open-source software: You can reuse it, tweak it, modify it. These very precise and specific and endlessly pedantically debated protocols played a crucial role in establishing modern technical practices, and they arguably decentralized power.

What you see with open-source AI, which became a thing in 2023, is narrative arbitrage: The halo of democratizing AI, reducing concentration of power, increasing scrutability is assumed to apply to AI when in fact the capabilities, affordances, and virtues of open source in software do not cleanly map on to AI.

The models and documentation surrounding the most open forms of AI are not bad in and of themselves. Being able to reuse a model can be useful. Being able to use it on-prem in an industrial setting that isn't sending data to a cloud provider may be necessary to maintain confidentiality. Being able to look at an open-weights model and understand a little bit better what might need to be tweaked for our use case is useful. Being able to examine the dataset used for training, being able to play with that and extend it is also very useful.

All of those things are great—but they don't do the things many of their boosters or their adherents seem to think they do. The key novelty of the current AI moment is the presence of concentrated amounts of data that had not been available before, and powerful distributed computational systems to process that data to train and perform inference on AI models. Even with open-source AI, you still need huge amounts of data, labor, and infrastructure. They don't

challenge the concentration that includes distribution networks, economies of scale, entrenched reach, the ability to define the tooling and the standards, and so on. Claiming they do these things confuses and distracts us from the type of solutions we need.

As people who are technologically responsible and grounded in a material understanding, we have to disentangle the rhetorical halo, which is misplaced from what these things actually do; what the positive benefits might be in different contexts; and what they fail to do and we thus must solve otherwise.

### **The misuse of technical concepts in AI abandons precision in favor of vibes.**

Half the technical terms I came up using in very precise ways are now just vibes-based evocations of abstractions that the user seems to have either forgotten or never learned. We are living in a world where vibes and ideology and even a theological concept of what technology is seems to have inoculated the population with a zombie virus. Half of tech is running around praying to a sky god that lives on an Nvidia chip.

I think we have to take this seriously because what I've seen happen in my industry—the detachment from technical reality, invoking abstractions as if they were the ground truth—is pretty stunning. And it's coming from a lot of the people who fifteen years ago were very pedantically teaching me the brass tacks of that material reality.

What we have now with the vibe-coding-crypto-influencer-turned-AI-influencer cohort is a bunch of people who have no idea about tech, but are being taken seriously because they can spin up a proof of concept via Claude code in three minutes. And a bunch of septuagenarian CEOs are taking their word for it.

There's probably some honest misunderstanding here—no one's asking questions and there's a lack of grounded material basis for a lot of the claims that are being made about AI in government and elsewhere.

I think that is matched with intellectual shame: The same powerful men who would rip apart a bad P&L document in a board meeting fail to ask a single question when presented with fantastical claims about AI and its capabilities. There is a fear of looking nontechnical, looking ignorant, looking behind the curve. A FOMO-driven juggernaut is dictating the need to adopt

AI. We don't know where, we don't know how, we're not sure how we measure it, but if we don't have an AI strategy, we are behind.

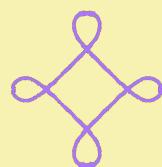
There is also a weak political opportunism where you don't want to look like a loser who doesn't have a solution to this. You'd like to spend the remainder of your time in power being a person who won. So quickly trading on a bit of rhetorical exaggeration may be more politically advantageous in the short term, allowing you to sidle up next to those in power and pretend you're with the party, as opposed to going direct with them and saying, "This form of openness is not actually providing the affordances that we need. This form of sovereignty is actually not sovereign. This form of independence is actually not independent."

**In the face of AI-industry concentration, open-source AI provides rhetorical reassurance—but it is incapable of delivering the promised goods.**

We're in a situation where you have two poles of power in AI. The infrastructural dominance of platform companies in the US emerged out of the commercialization of the internet. China resisted the market reach of the US firms and contained their market, built their own platforms, and had a large enough population to support top-down state control to mandate it.

So there is an anxiety—rightly so—among those who are not a major AI company or the US or Chinese state. This is an anxiety about sovereignty. Where do you stand in relation to this powerful set of technologies and how to figure out how to have a piece of it? What is a state if these things are ceded to centralized corporate actors? How do we maintain our status, our parity with the two states that do have these?

Terms like "openness" sound pretty appealing in this context.



**REFRAME:** Our tech ecosystem needs private and non-extractive technologies that uphold technical understandings of open source and scrutability.

**Small, focused deployments of open-source AI can be useful—but they should be pragmatic and mission-aligned.**

In the context of Signal, I am always pragmatic. I don't have the privilege of living in a counterfactual universe. This is not an academic exercise. We have to build something that is useful or we don't acquit ourselves honorably of our mission.

The problem with large language models is that they aren't small enough to run on-device. Then we get into what WhatsApp's doing where you split the baby around—sending data off-site to be processed, but saying that it's end-to-end encrypted because they are counting their server as an end. That gets messy and dishonest. I also don't think people want AI in their clean, crisp, beautiful, cool messaging app. There is not a need for this.

We do use an open-source AI model—not an LLM, it's small—that is meant for face detection to enable our “blur faces” feature. When you take a picture of crowds, you can block out people's faces. So if you post on social media, it's not scraped up by a Clearview database. That is very clearly a privacy-preserving use. Everything stays on-device. It's not sending things out to inference via some LLM on a cloud server, etc. Of course we'd use that. That helps people and is totally aligned with our mission.

**In this context, a precise technical understanding of open source and scrutability are the floor for establishing trust and building democratic technologies.**

I think openness should be the floor. As software, [Signal is] open source in the more clearly defined way. We're open source because we are a core infrastructure for the human right to private communication and free expression. We are relied on in life-or-death scenarios by people who really trust us. We don't want them to trust us because Meredith is a good talker or because the marketing team is excellent.

---

We don't think that there is an ability to be trustworthy without being able to scrutinize what we actually do. Does the code map to my rhetoric? Is there an issue that needs to be patched? We have a huge number of the best people in the security research community, the cryptography world, the hacker community, all scrutinizing our code regularly. I think of it as white blood cells looking for issues, writing into our security mailing list, finding things that we can test and patch to make sure we are actually running this core infrastructure with integrity.

We also raise the bar for the ecosystem. We are developing more than just Signal, and we think there should be a lot more giving back to the ecosystem to enable secure private and non-extractive technologies, given the fact that they are now the nervous system of our institutions and societies.

We're all in different places on the same boat. [Our] model is a kind of integrity that we shouldn't think of as exceptional. We should think of it as normal and ask everyone else, "Why aren't you following us?" When a CrowdStrike outage takes down half the internet, when it turns out a company was sharing data with governments or undermining their ethics principles by participating in military campaigns, we should really be questioning: "How did our formation of power enable a handful of people in service of short-term outputs to make socially significant decisions on behalf of everyone else without scrutability, without clarity, and without some form of democratic governance and oversight?"