

The Future of IaC

Join us for this live webinar to explore the state of IaC tools & future-proof your DevOps practices

LARGE LANGUAGE MODELS / OPEN SOURCE

Why Open Source AI Has No Meaning

As the community struggles to define open source AI, it battles unprecedented complexity.

Sep 9th, 2024 9:47am by [Alex Williams](#)



Image by Finn Hackshaw from Unsplash.

Meta is winning the **open source AI** story by using the term to push a large language model (LLM) that is not open source.

But really, the **Meta** camp can call **open source** whatever it wants.

AI creates a paradox for the open source community. On one side are the open pragmatists and on the other are the ones who want open source AI to be inspirational and true to its principles. The problem: open source is diminishing in

Is Terraform Dead?

Join us for this live webinar to explore the state of IaC tools & future-proof your DevOps practices

But underlying the arguments is a complexity not seen in the open source community since the **Open Source Initiative** defined open source more than two decades ago.

What we see: a stewardship question that allows for the exploitation of open source while the parties all push their own version of truth. **OSI is working on a definition**. But will it have the gravitas and the backing of the community? Right now, that's an open question.

Who Owns the Language of Open Source?

Open source has succumbed to linguistic drift partly due to a lack of stewardship of the language that defines open source, said **Ashley Williams** in an interview with The New Stack.

Williams is the founder and CEO of **Axo**, a packaging and distribution platform for developers building portable and secure software; she co-founded the **Rust Foundation** and is recognized as an open source strategist.

TRENDING STORIES

1. [Why Open Source AI Has No Meaning](#)
2. [What's Behind Elastic's Unexpected Return to Open Source?](#)
3. [An LLM-Turbocharged Community Calendar Reboot](#)
4. [Valkey: A Redis Fork With a Future](#)
5. [How to Explain the Security Advantages of Open Source](#)

Stewardship is more than the OSI managing the definitions and keeping course, she said Language changes over time. It becomes a matter of how people use the language.

"I really mean stewardship of what the colloquial definition of open source means," Williams said. "And I think that there are a couple of dimensions to how that stewardship got dropped."

roles shaping commercial open source — people who had the role of director or other titles.

A shift happened. Open source served as a way to decrease costs for technology development. It served commercial interests by relying on volunteers to manage open source projects.

Open source continued to shift in meaning, especially in the cloud native era, Williams said. It broadened with the advent of openness in open source with such efforts as open governance.

So, when you look at open source, you have to look at who is using the words and why they are using them. The term “open source,” is now used by Meta. It owns the conversation.

Meta can call its LLM **Llama** open source because there is so much confusion about how to define open source AI in the first place.

Even the leader of OSI cites the challenge of using the term “open source AI,” pointing to the name of **OpenAI**, the organization behind ChatGPT.

“If it wasn’t taken as a name, that would be ideal because there is technically no ‘source’ in it,” said **Stefano Maffulli**, executive director of OSI, in an interview with The New Stack. “So using the term ‘open source AI’ is a little bit of a misnomer, but it is what it is, right? It’s already out there. We have to deal with it.”

Maffulli said it didn’t help when the European Union said that open source AI systems, or AI systems that are open source, have special advantages and are spared from some requirements.

“It’s another driver for [Mark] Zuckerberg to push to be affiliated with the term ‘open source, AI.’”

OSI has **posted the draft definition** for people to comment on. It’s like a mire — bogged down and muddy.



"There can be no restriction of **commercialization**, and we have fought tooth and nail amongst ourselves as a community for years when anybody tries to restrict commercialization or even to put ethical provisions in there, because the free flow for us really matters," said Brock, who resigned from the OSI board in July 2023, after serving for 21 months.

"And that's what allows open source to be used, reused without concern about restriction, and it allows adoption to happen at scale, and it's really essential to the model."

It's hard enough to manage one definition for open source. When a second definition of open source is created, said Brock, "you run the risk of confusion and undermining ... the very heart of what open source software is."

Pragmatic vs. Aspirational

A pragmatic definition versus an aspirational one has kept the community thinking through the implications of **training data**. In the meantime, the confusion will rest on which checkboxes an LLM provider can check to comply with an open source definition. If a provider can restrict modification, does that mean that LLM is still open source?

OSI cares about a definition but maintains on its website that "defining training data as a benefit, not a requirement, is the best way to go."

It's the OSI's position about training data that poses the most consternation.

The source of the model comes from the data and the code, wrote **Steve Pousty**, a developer advocacy consultant, in a comment on the open source AI draft definition on the **OSI site**.

"This definition does not grant the freedom to modify and is unacceptable as an Open Source Definition," wrote Pousty. "With AI models, the weights are the user interface. I can use them directly as a user. They are what is typically distributed to everyone.

The actual source of the model comes from the data AND the code. The weights are built using the code and the data. Together, they make up the ability to

Maffulli takes some exception. He described how murky this subject gets. The training data may be a big bucket of private information, copyrighted material and factual information. How the data gets distributed can lead to legal complexities, such as what **EleutherAI** faces.

EleutherAI is a nonprofit AI research lab. The data it uses comes from the internet and other sources. It opens the weights, the code and the training data. Still, rights activists don't like it, so they targeted Books3, a book data set by EleutherAI.

In August 2023, the Rights Alliance, an antipiracy group in Denmark, filed to have the Books3 data set removed from Pile, its data set for training large language models.

It's an example of why open source gets so tangled in AI's chaos. It's not just the data, like Pousty noted, it's the weights and the code that make the system.

The aspirational look at the mess will tell you that the AI system is not open source without data transparency. Open the training data to continue the trajectory for aspirational approaches to AI systems development.

Seeking 'a Spectrum of Openness'

In August at **FOSSY24** in Portland, Ore., a keynote panel reviewed the state of open source and AI.

"What really people want is a spectrum of openness," said **Julia Ferraioli**, an open source strategist, researcher, and practitioner at **Amazon Web Services**.

"There's the ability to make the infrastructure, the software around a model, open source through software licenses. But for the data, for the model itself, it gets a little bit more complicated. What people tend to want is a binary: 'Is this open source or not?'"

"So while a spectrum of openness can be useful, it's hard to implement it in a practical manner."

"I think for the long term, we need to hold that line — have a clear, aspirational point," Randall said. "And I don't care if the OSI defines their trademark open source definition of AI as this or not. But we need to define a clear aspirational point and recognize that is where we get the full benefits of software freedom."

Anything less than that is not open source, Randall said, and that's fine. If there are things out there that don't meet that aspirational point and we still recognize them as open source, then that's when things get muddy.

It's about recognizing that the large corporations already have the advantage, Maffulli said. If we take a purely aspirational approach, no one can meet the definition, which will cause trouble for the small players. Big tech companies have a lot of data that they can use. The rest of us don't.

"Open source is going to be in a corner, playing with toys," he told TNS, "while the big guys are playing with cars and machine guns."

TNS



Alex Williams is founder and publisher of The New Stack. He's a longtime technology journalist who did stints at TechCrunch, SiliconAngle and what is now known as ReadWrite. Alex has been a journalist since the late 1980s, starting at the...

[Read more from Alex Williams →](#)

Amazon Web Services is a sponsor of The New Stack.

SHARE THIS STORY



TRENDING STORIES

1. Why Open Source AI Has No Meaning

[What's Behind Elastic's Unexpected Return to Open Source?](#)

5. How to Explain the Security Advantages of Open Source

INSIGHTS FROM OUR SPONSORS



6 top incident management use cases for AI copilots

10 September 2024

5 ways teams used BigPanda during the CrowdStrike outage

5 September 2024

Alert noise reduction: How to cut through the noise

5 September 2024



liblab SDKs Boost Doppler's Developer Adoption by 178% in 6 Months

9 September 2024

Using SDKs to Reduce API Integration Time

12 August 2024

Introducing liblab hub: SDK Collection for Effortless API Integration

25 July 2024



How to install plugins in Kubernetes and essential plugins to get started

10 September 2024



Spacestronaut in Orbit: Daiki Takasao

September 2024

TNS DAILY NEWSLETTER

Receive a free roundup of the most recent TNS articles in your inbox each day.

EMAIL ADDRESS

SUBSCRIBE

The New Stack does not sell your information or share it with unaffiliated third parties. By continuing, you agree to our [Terms of Use](#) and [Privacy Policy](#).

ARCHITECTURE

- Open Source
- Cloud Native Ecosystem
- Containers
- Edge Computing
- Microservices
- Networking
- Serverless
- Storage

ENGINEERING

- AI
- Large Language Models
- Frontend Development
- Software Development
- API Management
- Python
- JavaScript
- TypeScript
- WebAssembly

OPERATIONS

Platform Engineering
Operations
CI/CD
Tech Careers
Tech Culture
DevOps
Kubernetes
Observability
Service Mesh

CHANNELS

Podcasts
Ebooks
Events
Newsletter
TNS RSS Feeds

THE NEW STACK

About / Contact
Sponsors
Advertise With Us
Contributions



Community created roadmaps, articles, resources and journeys for developers to help you choose your path and grow in your career.

Frontend Developer Roadmap
Backend Developer Roadmap
Devops Roadmap

FOLLOW TNS

© The New Stack 2024

Disclosures
Terms of Use
Advertising Terms & Conditions
Privacy Policy
Cookie Policy



