

DEC 11 | 8am PDT/11am EDT - Cloud DevSecOps Workshop  
with Terraform Cloud, Prisma Cloud by Palo Alto

F

AI / OPEN SOURCE

# The Open Source AI Definition: What the Critics Say

The Open Source Initiative aimed to achieve consensus when it defined open source AI. Instead, it's laid bare community concerns in an anxious time for open source overall.

Nov 18th, 2024 8:30am by [Richard Gall](#)

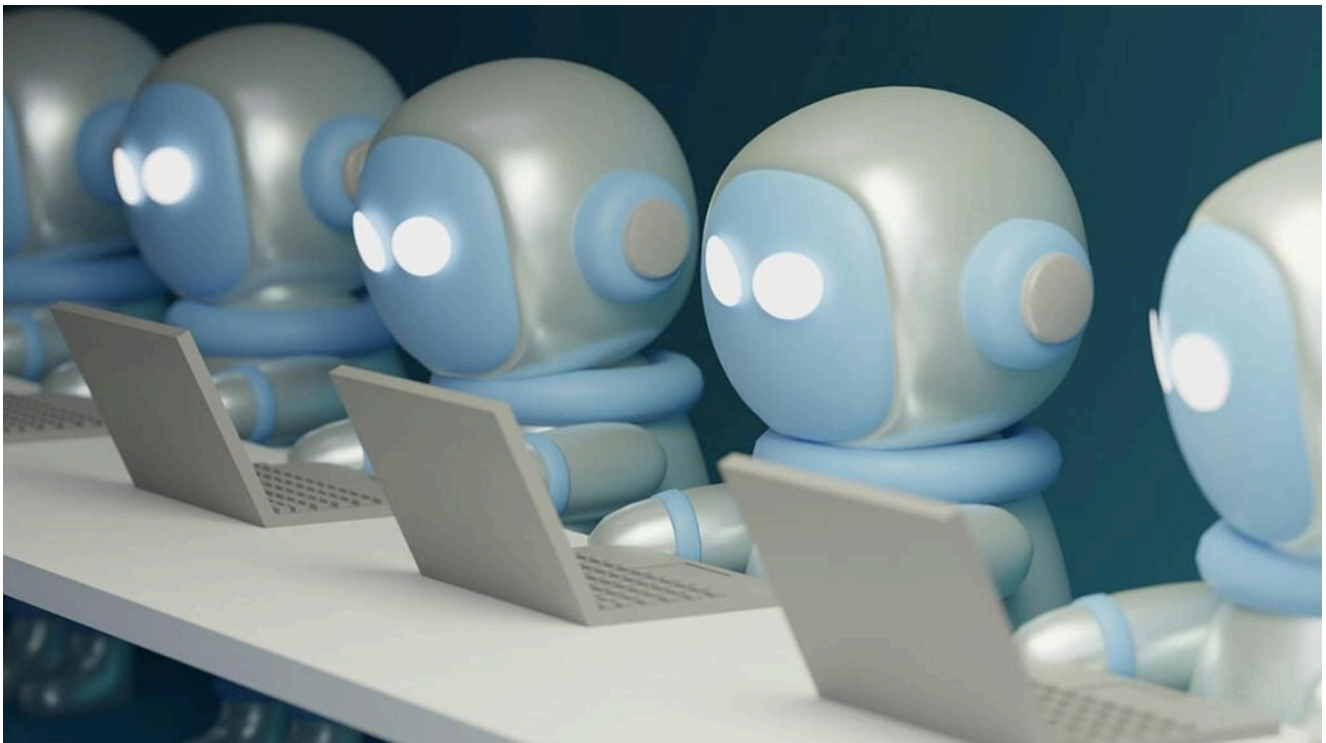


Image by Valeria Nikitina from Unplash+.

The **release of the Open Source Initiative's definition of open source AI** at the end of October was supposed to be a milestone of consensus building. It was, OSI chair **Tracy Hinds** told The New Stack over email, an attempt to "drive a stake in the ground so that the industry, academia, civil society and policymakers have



Persistent Storage Solutions

Open Source SDS and world-class support for your most critical and intensive business applications

Le

## THE NEW STACK

---

a particularly fortuitous essay — titled **The OSI lacks competence to define open source AI** — that he posted on his blog, **Sam Johnston**, an AI leader at **Kwai**, described the whole process as “[ramming] through a redefinition of a quarter century of Open Source that everybody else hates.”

Far from settling an issue in open source, then, the OSI’s definition appears to have further stirred up ongoing tensions in the world of open source. That this has all happened in the same news cycle as the **ongoing WordPress debacle** only emphasizes that this is symptomatic of a bigger problem no one has quite been able to get to grips with.

So, what are critics saying are the main problems with what the OSI has published? Do they think the definition simply needs to be discarded? Or does the OSI deserve more credit than many are currently granting it? Can some **patience and flexibility** help us reach a greater consensus?

### What Kind of ‘Thing’ Is AI, Anyway?

One of the most fundamental criticisms of the OSI’s definition is that it simply doesn’t make sense to talk about open source AI.

“I think it’s a mistake because you’re defining what I would call a market that will consistently shift and change,” **Amanda Brock**, chair of OpenUK told The New Stack. (Brock emphasized that her views on this subject are hers alone, not reflective of OpenUK.)

To illustrate her point, she highlighted earlier conversations about the notion of there being “open source mobile” or “open source cloud.”

#### TRENDING STORIES

1. [RIP Open Core — Long Live Open Source](#)
2. [CNCF Sics Developers on Kubernetes Patent Trolls](#)
3. [What Comes After Open Source? Bruce Perens Has Some Ideas](#)

## THE NEW STACK

---

"I don't think it was a mistake not to define 'cloud' or 'mobile.' I think it would have been a very inappropriate thing to do," she noted. This, Brock said, is because these are categories of products or services rather than a distinct "thing." In other words, **AI is a complex and multifaceted thing** made up of different parts — open source software, on the other hand, is ultimately all source code.

This is partly why **Julien Sobrier**, senior product manager at **Endor Labs**, has reservations about the definition.

"An AI model is made of many components: The training set, the weights and programs to train and test the model," he told TNS over email. "It is important to make the whole chain available as open source to call the model 'open.'"

On this, the OSI is well aware of the unique challenge of thinking about AI as a single "thing."

"It was evident early in the two-year process [of developing the definition] that AI systems are not programmed; they're totally different from software," Hinds told TNS. "The core of the work that OSI led was to discover what is the preferred form of making modifications to an AI/ML system: it's the model parameters, the code used to train the system, the code used to process the training data and the data itself, unless it's illegal to distribute."

Hinds went on to say that this "is counter-intuitive for some software developers, used to having complete access to source code and build scripts of open source software."

It's clear that the OSI is keen to reflect the nuances of AI systems in its definition. But maybe the attempt to do so has created something altogether more confusing. This is particularly an issue when it comes to differences between data and software, one of the most contentious parts of the definition — which

arguably set the OSI up against all sides, from corporate giants to open source purists.

“Combining training data with open source software in this way is confusing,” he wrote, “as traditionally we’ve often been encouraged to consider licensing for data and code separately.”

Brock echoed Barker. “Any lawyer who’s worked in the space has been telling everybody for decades not to do that because it is incorrect ... data and software are very different intellectual property. They have totally different concerns around them, so you should not be mixing those licenses.”

However, the OSI doesn’t regard this as a concern. As Hinds put it, for the OSI, AI systems are a distinct “thing,” separate from software and AI.


“It was evident early in the two-year process that AI systems are not programmed; they’re totally different from software,” she wrote to TNS. “AI (machine learning, in particular) relies on data, but it’s not data either.”

The difference in perspective should be clear: for someone like Brock, AI *isn’t* a distinct entity separate from the software or data on which it depends; for the OSI, it *is* — and ultimately, the definition wouldn’t really work if it wasn’t treated as such.

Brock believes there is an alternative approach. “If you were going to [define] anything and you thought it wasn’t software or data, you would look at a component part and define the components.”

## Is the OSAID Even Workable?

Beyond the existential question of what AI actually is, there are also questions about whether the definition is useful as a tool for policymakers, researchers and technologists.

Although **the version released in October** isn’t a beta, Hinds stresses that the OSI  isn’t see this definition as something static. This version is what she describes as “a workable standard.”

## THE NEW STACK

---

However, while such flexibility may be well-intentioned and a strength in a field that is changing rapidly, from a policy perspective, its seemingly provisional quality removes the weight and force that you'd typically expect from a definition with legal implications.

For Brock, the idea of a definition that evolves seems almost pointless. "I don't think it is helpful to create a definition where it's going to constantly ebb and flow," she said. "When you get to policy and legislation and [you're] assessing risk, if you look at a definition that isn't fixed, that's a risk in itself that you're not going to adopt."

But maybe the definition's clout and punch aren't *everything* at this stage. It takes time for new ideas to take root in a community.

Despite his reservations and questions about the definition, Barker sees the debate — as chaotic as it may seem to us now — as productive and an integral part of the whole process.

"This is part of the broader push-and-pull we see whenever a disruptive paradigm emerges, as we saw with 'the cloud,'" he told TNS. "This creative conflict, while messy, ultimately leads to a clearer understanding and stronger standards. It's just going to be murky for a while."

For the OSI, what is critical above all else is, according to Hinds, that it "meaningfully [gives] developers the right to fork." At a fundamental level, Hinds and the OSI believe their definition does this.

"Some people believe that more components are required to guarantee more transparency," she told TNS. "Other groups instead believe that model parameters and architecture are enough to study and modify."

The OSAID, Hinds added, "found that while those approaches are legitimate, whether is optimal to promote meaningful collaboration and innovation. The OSAID grants users the rights (with licenses) and the tools (with the list of required

## THE NEW STACK

---

possible impact on open source more generally. Johnston raised this point in his essay, writing that “an ineffective and unimplementable open source AI definition risks our being crowded out by opaque open-washed offerings from commercial vendors.”

If he’s point is correct, then it would be highly ironic. The context in which the OSAID has emerged is one where terms like “open source” have been attached to AI-related projects and products in ways that seem to push the boundaries of the open source ethos — code that can be accessed distributed and modified (i.e. “forked”) freely.

Although the OSI has not named this as motivation for the OSAID, the organization’s explicit criticism of Meta’s use of “open source” to describe its Llama model means it isn’t a stretch to regard this as an important subtext to the definition. (In an interview with the Financial Times — sorry, it’s paywalled— the OSI’s Executive Director Stefano Maffulli accused the company of polluting the term.)

How this argument concludes is difficult to predict. As Barker put it: “I doubt many could tell you the exact definition of open source on the spot, aside from experts – and even they don’t always agree.”

Despite his reservations about the OSAID, he’s also circumspect. “Having worked in this space for nearly 20 years, I understand the importance of these discussions — but I also try to stay pragmatic.”

The details of the definition aside, if the OSI has made a mistake with this definition, it’s in seeing open source as a purely legal term, an abstract idea or tool that lives separately from the community that has used and stewarded it over decades.

Corresponding with TNS over email, Aeva Black — section chief for open source security at the U.S. Cybersecurity and Infrastructure Security Agency, and a former OSI board member — made the point that while the original open source definition ultimately just “refers to a specific class of software licenses,” its growth “seems to have outpaced the transmission of the culture that led to the



## THE NEW STACK

---

insular as people believe in it and participate in it. People will only really continue to believe in it if the concept has power and meaning in the industry.

This was a point made by **Ashley Williams**, founder and CEO of **axo**, in a TNS piece **published in September**. She highlighted that while open source has undergone linguistic drift over the decades, this isn't a policy or even directly an institutional one: it's a question of stewardship and community.

The author of that piece, TNS Publisher **Alex Williams** (no relation to Ashley) puts the point neatly: "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."

The community has become disparate and fragmented by competing interests (typically commercial) — so much so that committed ownership of the term "open source" has all but disappeared

### Should the OSI Have Even Taken This On?

Given the OSAID only seems to have exacerbated confusion and controversy, the major question that needs to be answered is whether this should have been taken on by the organization in the first place.

Brock echoed the likes of Johnston in claiming the OSAID poses a "real risk of undermining open source software." But she also sees much of this risk as stemming from the way resources have been sucked up by the project.

For her, the key issue is what can be done to ensure the longevity of open source, **especially at a time when it seems so vulnerable**. She told TNS that she believes the OSAID "undermines... longevity by creating confusion and by sort of dispersing resources."

Hinds offered a robust rebuttal of this argument in defense of the OSI. "We're here to focus our resources on the projects and priorities that the community-directed board directs us to pursue," she said.

"If you look at our activities in the past two years, the OSI has done a lot more

Even if the resources are there and the OSI has the confidence to take on the many challenges facing open source, there’s no doubt that questions about its approach will remain. It’s possible that the biggest threat to open source is status anxiety on the part of the major players in the open source world and, rather than addressing the needs of open source in a more holistic or aligned manner, we’re seeing fragmentation and in-fighting.

“As a recognized steward of open-source principles, the OSI’s mission has always been to uphold standards and emphasize the importance of non-proprietary software,” Barker told TNS. “Whether they strictly need a specific AI definition to do that is debatable. But with AI emerging as both fundamental and disruptive, the OSI may feel it’s essential to position themselves as a relevant authority here.”

The real challenge for the broader community — if it wants to tackle many of open source’s current issues, not just AI — is to disentangle authority and stewardship. While both require leadership and vision, there’s undoubtedly some tension between them; good stewardship may require giving up some authority or distributing power more widely. Until that’s addressed, we may remain in murky waters for some time.

TNS



Richard Gall is a writer based in the northeast of England. He's interested in the intersection of technology, culture and politics and likes talking to technologists about their work. Find him on X @richggall or on BlueSky @richgall.bsky.social.

[Read more from Richard Gall →](#)

## SHARE THIS STORY



## TRENDING STORIES