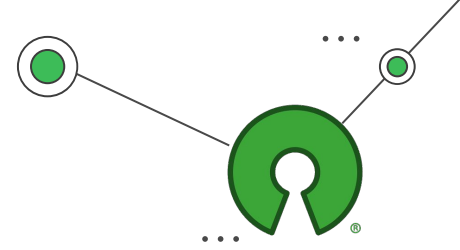


Open Source and AI

Drawing the line between freedom and chains in the age of machine learning

Stefano Maffulli | OSI Executive director

The foundation of the Open Source ecosystem



Globally recognized

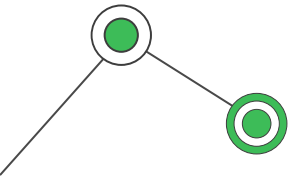
OSI is recognized by individuals, companies, and by public institutions from all over the world as **the authority** that defines Open Source.

Neutral partner

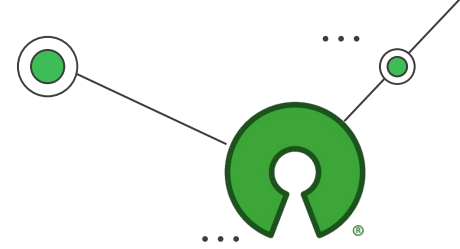
OSI represents equally the interests of all Open Source projects, allowing us to participate in complex policy discussions **without prejudice**. The board only acts for the public interest, not those of sponsors or any special interest group.

Trusted independent

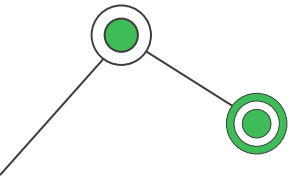
OSI is a public charity with an **open governance**. Two thirds of our board is elected by individual members and affiliate organizations.



Our projects



License and Legal	Policy & Standards	Advocacy & Outreach
<p>An anchor for open community consensus on what constitutes Open Source. We protect the Open Source principles, enforcing the marks “Certified Open Source” and “Open Source Approved License”.</p>	<p>Monitoring policy and standards setting organizations. Supporting legislators and policy makers educating them about the Open Source ecosystem, its role in innovation and its value for an open future.</p>	<p>Leading global conversations with non-profits, developers and lawyers to improve the understanding of Open Source. OSI investigates the impacts of ongoing debates from artificial intelligence to security.</p>



What is Open Source AI

To be Open Source, an AI system needs to be available under legal terms that grant the freedoms to:

- **Use** the system for any purpose and without having to ask for permission.
- **Study** how the system works and inspect its components.
- **Modify** the system to change its recommendations, predictions or decisions to adapt to your needs.
- **Share** the system with or without modifications, for any purpose.

**Access to source code is
a precondition for this.**

```
20 glBindBuffer(GL_ELEMENT_ARRAY_BUFFER, indexBufferObj);
21 glBufferData(GL_ELEMENT_ARRAY_BUFFER, 6 * sizeof(unsigned int), indices, GL_STATIC_DRAW);
22
23 // shader
24 unsigned int shader = createShader(readFile("res/shaders/vertex.shader"),
25   readFile("res/shaders/fragment.shader"));
26 glUseProgram(shader);
27 int location1 = glGetUniformLocation(shader, "u_color");
28
29 while(!glfwWindowShouldClose(window)) {
30     // clear the buffer
31     glClear(GL_COLOR_BUFFER_BIT);
32
33     // sets the background color
34     glClearColor(1.0f, 1.0f, 1.0f, 1.0f);
35
36     // draw
37     glUseProgram(shader);
38     glUniform4f(location1, 85.0f*INV_255, 184.0f*INV_255, 237.0f*INV_255, 1.0f);
39     glBindVertexArray(vertexArrayObj);
40     glDrawElements(GL_TRIANGLES, 6, GL_UNSIGNED_INT, NULL);
41
42     // display bound buffer
43     glfwSwapBuffers(window);
44 }
```

The black box

How does it do that?

Can we fix it?



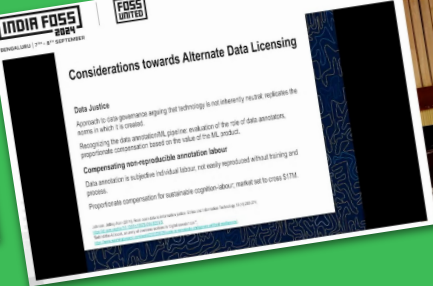
The source code must be the preferred form in which a programmer would modify the program.

The Open Source Definition #2

... the preferred form ...

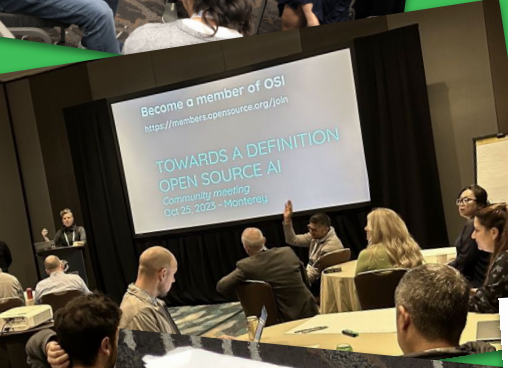
Co-Designing the OSAID A Global Snapshot

Our co-design process included in-person workshops on five continents – South America, North America, Africa, Europe, and Asia – and virtual participants from more than 35 countries.



Study

Study how the AI system works, and inspect its components. Access to the AI system components is the preferred form to maintain. This is a precondition of the



Share

Share the system, with or without modification for any purpose, [without limitations].

USE

Use the system for any purpose without any limitations and without having to ask for permissions.

USE

Use the system for any purpose without any limitations and without having to ask for permissions.

there should be no limitations on the use of the system

any child



The co-design process



As defined by the OECD.

What elements are necessary to:

- use
- study
- modify
- share

an AI system?

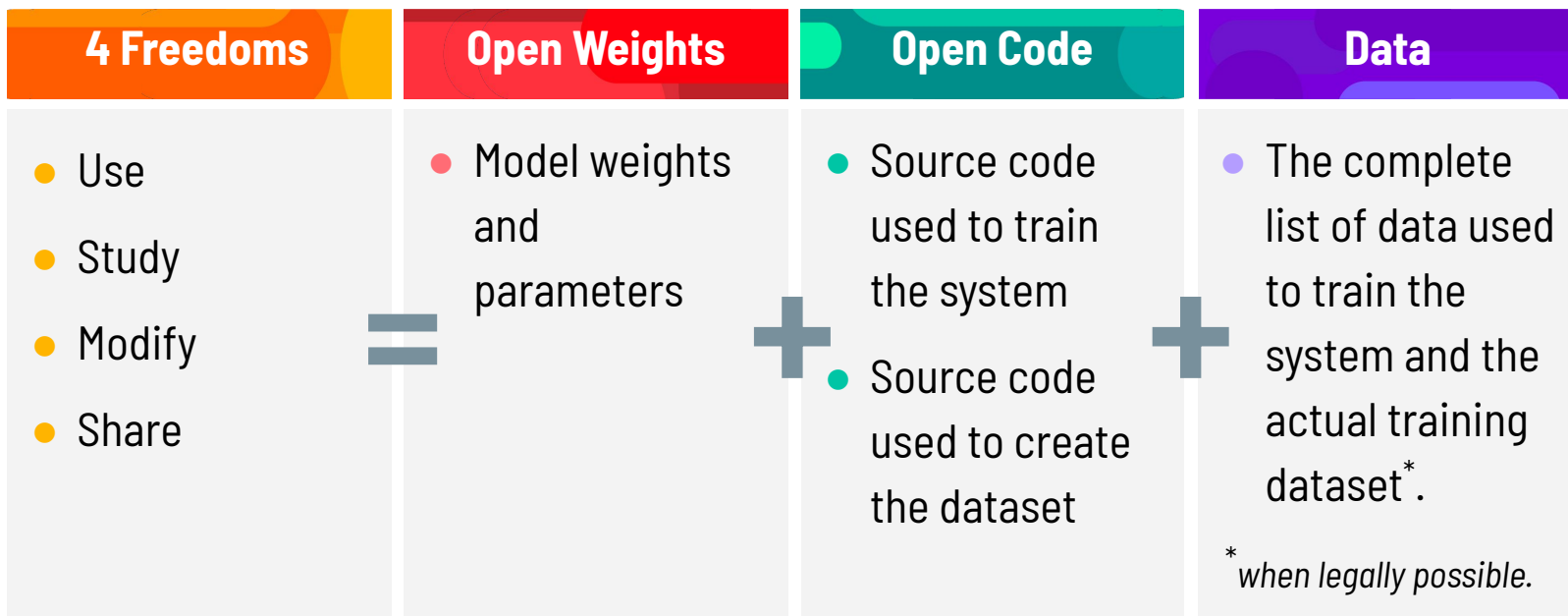
Which existing AI systems meet the criteria of the Open Source AI Definition draft?

Promote and ask communities for feedback in worldwide tour of the validated draft. Iterate versions based on feedback.

Formally release a stable version. Continue gathering global feedback, release new versions as necessary.



Open Source AI Definition



All under OSD-compatible/OSI-approved terms.

Meta's comment

“We agree with our partner the OSI on many things, but we, like others across the industry, disagree with their new definition,” the spokesperson said.

“There is no single open source AI definition, and defining it is a challenge because previous open source definitions do not encompass the complexities of today's rapidly advancing AI models. We make Llama free and openly available, and our license and acceptable use Policy help keep people safe by having some restrictions in place. We will continue working with the OSI and other industry groups to make AI more accessible and free responsibly, regardless of technical definitions.”

“There is no single open source AI definition”

**Have you received all the tools
and knowledge you need
without any restrictions?**



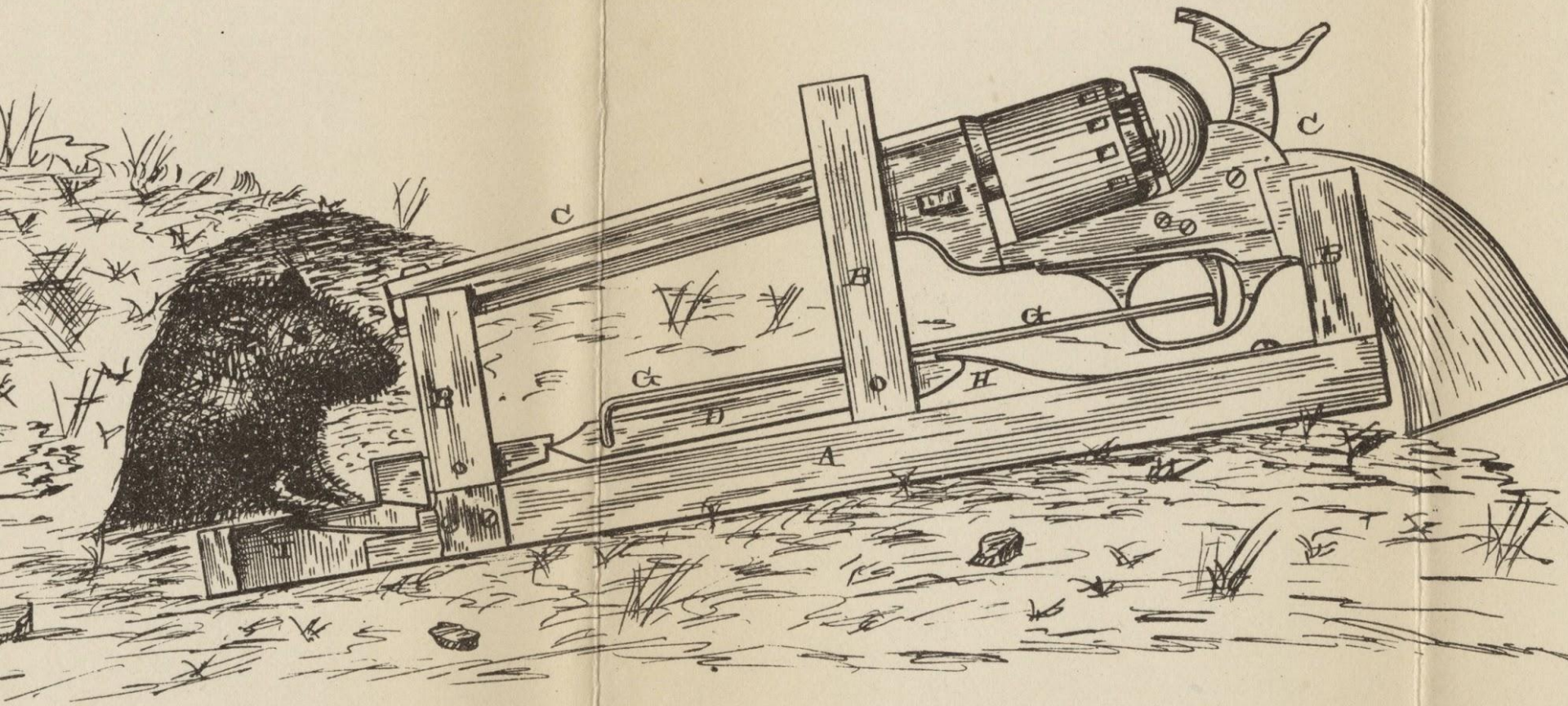
Acceptable Use Policy

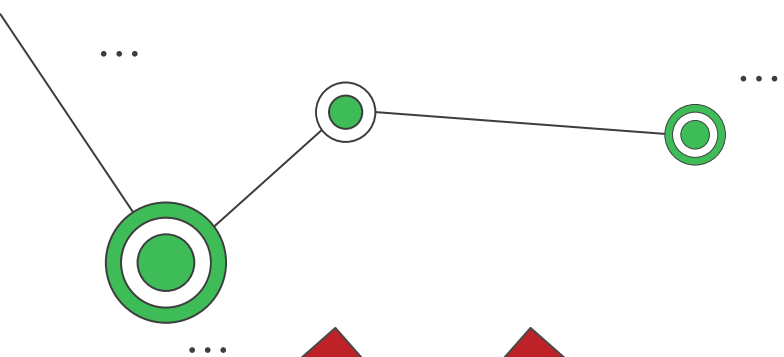


1. Violate the law or others' rights [...]
2. Engage in, promote, incite, facilitate, or assist in the planning or development of activities that present a risk of death or bodily harm to individuals [...]
3. Intentionally deceive or mislead others [...]

Additional Commercial Terms. If, on the Meta Llama 3 version release date, the monthly active users of the products or services made available by or for Licensee, or Licensee's affiliates, is **greater than 700 million monthly active users** in the preceding calendar month, you must request a license from Meta, which Meta may grant to you in its sole discretion, and you are not authorized to exercise any of the rights under this Agreement unless or until Meta otherwise expressly grants you such rights.

With respect to any multimodal models included in Llama 3.2, the rights granted under Section 1(a) of the Llama 3.2 Community License Agreement are **not being granted to you if you are an individual domiciled in, or a company with a principal place of business in, the European Union.**

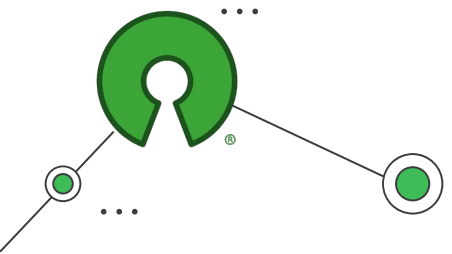




Accountable
Use of Data



Responsible
Deployment





Degrees of deprivation



The gate of Open Source

Explosion of freedom



openstack



SQLite

Digital
Public
Goods
Alliance

DPG



Criticisms



Wrong term!

There is no “source” therefore Open Source AI is the wrong term.

Open Source allows reproducibility but this is not possible, so call it something else.

OSI stretched its mission

OSI should stick to software, let others deal with AI/ML.

The process was unfair

Some people felt excluded. Companies couldn't engage due to the speed and public nature.

Open data must be required

...

Not solved by others

FSF

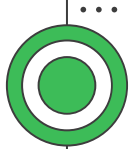
A new dilemma

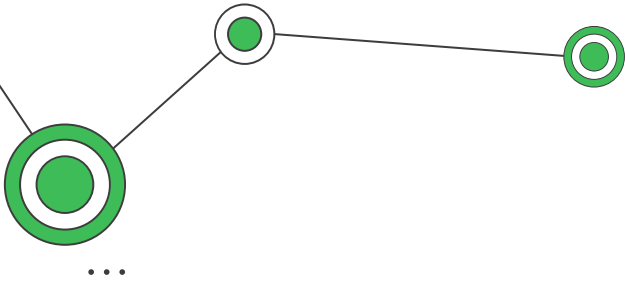
“It may be that some nonfree ML have **valid moral reasons** for not releasing training data, such as personal medical data. In that case, we would describe the application as a whole as nonfree. But **using it could be ethically excusable** [...]”

SFC

“Aspirational” vision

B. Kuhn: “I truly don't know for sure (yet) if the only way to respect user rights[...] is to only use training sets that are publicly available[...]. [I] believe that **it will take at least a decade** for our best minds to find a reasonable answer [...]”



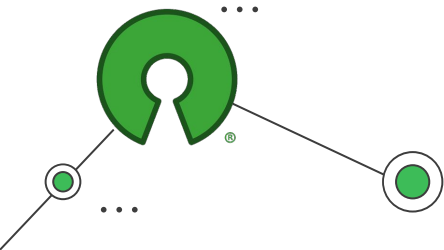


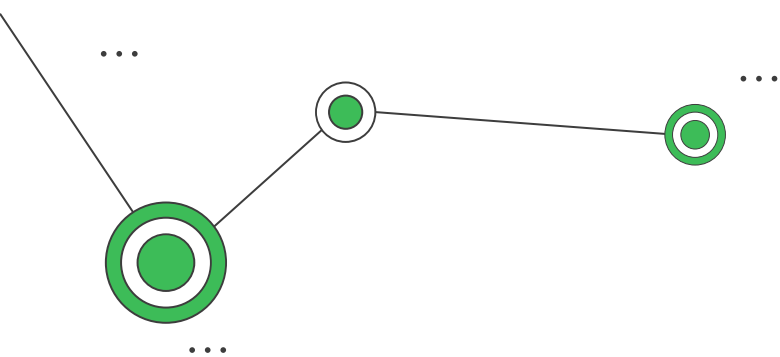
Other criticism

Too restrictive?

Training code and data processing code are not required to “fork”, as demonstrated by thousands of derivatives of Llama with new capabilities.

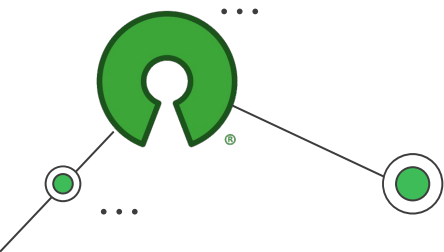
None of the top commercial releases fit the Definition.





What's next

OSI current work plan



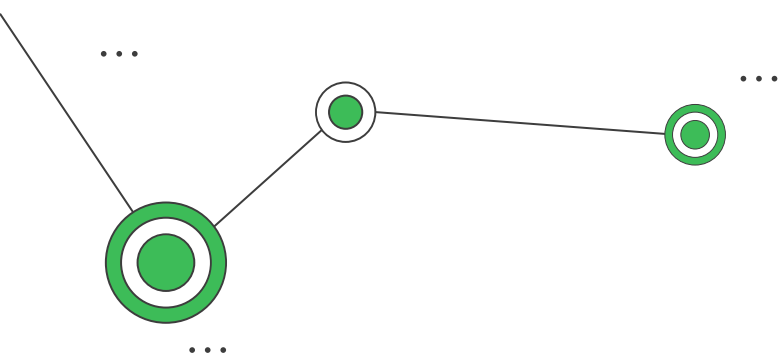
v. 1.0 is a conversation starter!

2025:

- Establish a multilateral working group to record major issues
- Monitor the AI space, see how the practice evolves
 - who releases what, under what conditions?
 - what's happening outside of LLM/GenAI? (biotech, medical, vision, etc) What's up with federated learning? What's the role of synthetic data?
- Solicit public feedback to v.1.0
 - Use the public OSI forums
 - Hold BoFs at popular conferences

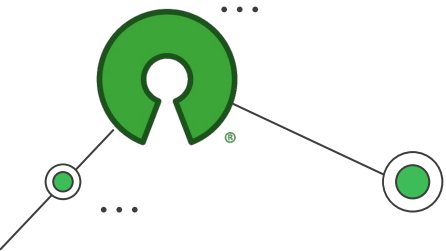
2026:

- Run a process to address major issues
- Fix major issues by Q4 2026



State of AI policy discussion in Europe

Jordan



Questions?

Endorse the OSAID

<https://opensource.org/ai/endorsements>

stefano@opensource.org

