Open Source Initiative

Defining Open Source AI

Mer Joyce | September 27, 2024

Supported by:

ALFRED P. SLOAN FOUNDATION

Why Define Open Source AI Now?

Frontier of OS

Defining Open Source AI is the most significant challenge facing the open source movement.

Shaping Regulation

Government regulations have begun in the EU, US, and elsewhere. We have the opportunity to share these new policies and laws by defining OSAI.

Combat Open-Washing

Companies are calling AI systems “open source” even though their licenses contain restrictions that go against the accepted principles and freedoms of open source.

Benefits of Open Source AI

by Lea Gimpel

Transparency + Safety

OSAI provides information essential for auditing systems and to mitigate bias, ensures accountability and transparency of data sources, and accelerates AI safety research

Market Deconcentration + AI Polyculture

OSAI makes more models available, spurs innovation and quality due to increased competition and tackles AI monoculture by providing more stakeholders access to foundational technology.

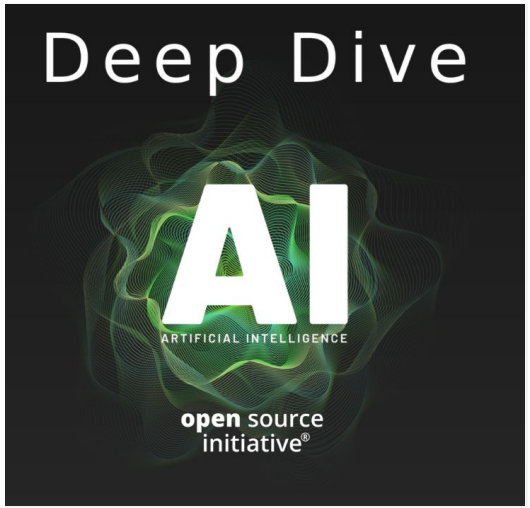
Diverse Applications

OSAI gives developers access to resources crucial for developing context-specific, localized applications that are representative of cultural and linguistic diversity and allow for model aligned with different value systems.

https://opensource.org/deepdive/webinars

2022-2023 research

Interviews (Podcast)



Panel discussions

Four panels with 4 experts covering 4 area:

- Business
- Society
- Legal
- Academia

Webinar series

18 experts of different disciplines from all over the world dissected issues from data governance, privacy, labor laws, software development and more.

https://deepdive.opensource.org/report/

https://opensource.org/deepdive/webinars

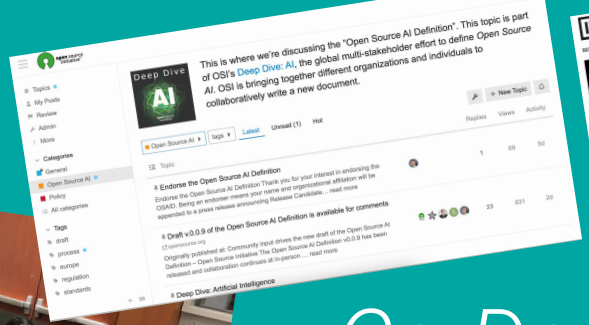
We've known that the availability of training data was THE ISSUE

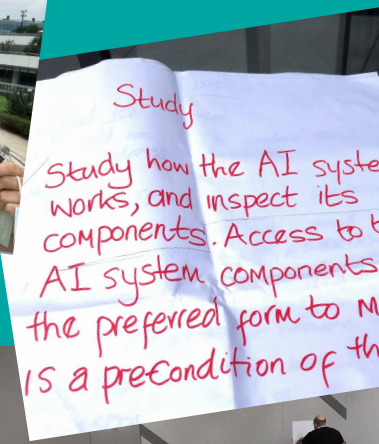
Co-Designing the OSAID

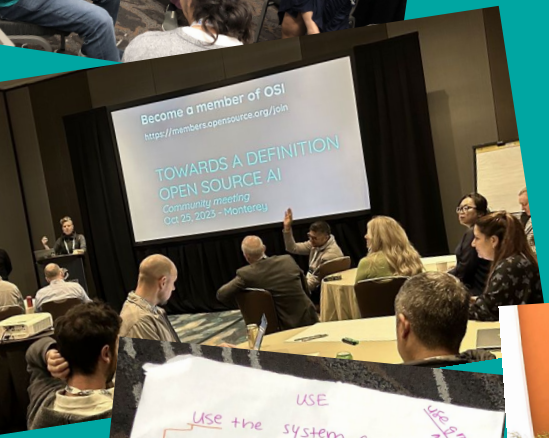
How we created the Open Source AI Definition through global consultation.

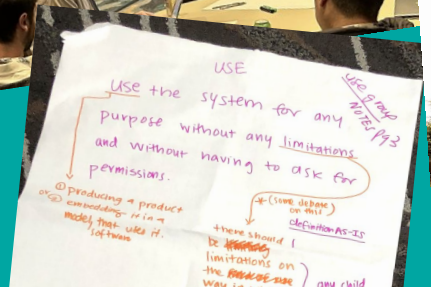
Co-Designing the OSAID
A Global Snapshot

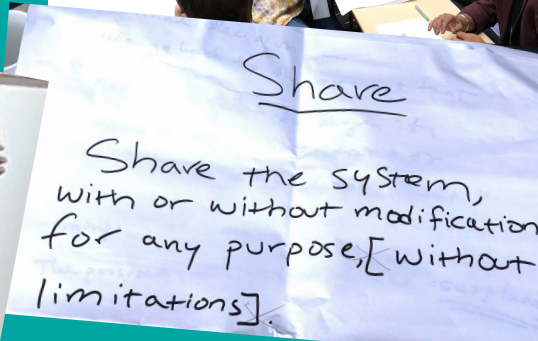
Our co-design process included in-person workshops on five continents – South America, North America, Africa, Europe, and Asia – and virtual participants from more than 35 countries.









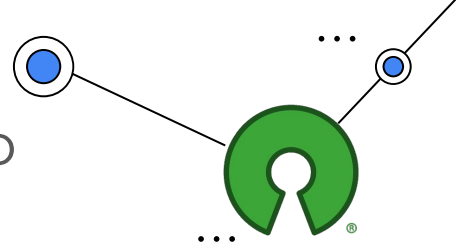


OSAID Co-Design Question

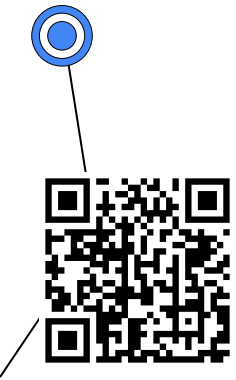
Use • Study • Modify • Share
What should these open source principles mean for artificial intelligence?

Open Source AI Definition Four Freedoms

v.0.0.9



1. **Use** the system for any purpose and without having to ask for permission.
2. **Study** how the system works and inspect its components.
3. **Modify** the system for any purpose, including to change its output.
4. **Share** the system for others to use with or without modifications, for any purpose.



OSAID Co-Design Question

What components must be open in order for an AI system to be used, studied, modified, and shared?

Virtual Workgroups

Members agreed to make their names and affiliations public to support the transparency of the co-design process.

Llama 2 Group

1. **Bastien Guerry** DINUM, French public administration
2. **Ezequiel Lanza** Intel
3. **Roman Shaposhnik** Apache Software Foundation
4. **Davide Testuggine** Meta
5. **Jonathan Torres** Meta
6. **Stefano Zacchiroli** Polytechnic Institute of Paris
7. **Mo Zhou** Debian, Johns Hopkins University
8. **Victor Lu** independent database consultant

BLOOM Group

1. **George C. G. Barbosa** Fundação Oswaldo Cruz
2. **Daniel Brumund** GIZ FAIR Forward - AI for all
3. **Danish Contractor** BLOOM Model Gov. WG
4. **Abdoulaye Diack** Google
5. **Jaan Li** University of Tartu, Phare Health
6. **Jean-Pierre Lorre** LINAGORA, OpenLLM-France
7. **Ofentse Phuti** WiMLDS Gaborone
8. **Caleb Fianku Quao** Kwame Nkrumah University of Science and Technology, Kumasi

Pythia Group

1. **Seo-Young Isabelle Hwang** Samsung
2. **Cailean Osborne** University of Oxford, Linux Foundation
3. **Stella Biderman** EleutherAI
4. **Justin Colannino** Microsoft
5. **Hailey Schoelkopf** EleutherAI
6. **Aviya Skowron** EleutherAI

Over 50% of **participants** are People of Color, 30% are Black, and 25% are women, trans, and nonbinary.

OpenCV Group

1. **Rahmat Akintola** Cubeseed Africa
2. **Ignatius Ezeani** Lancaster University
3. **Kevin Harerimana** CMU Africa
4. **Satya Mallick** OpenCV
5. **David Manset** ITU
6. **Phil Nelson** OpenCV
7. **Tlamelo Makati** WiMLDS Gaborone, Technological University Dublin
8. **Minyechil Alehegn Tefera** Mizan Tepi University
9. **Akosua Twumasi** Ghana Health Service
10. **Rasim Sen** Oasis Software Technology Ltd.

Voting for Requirements

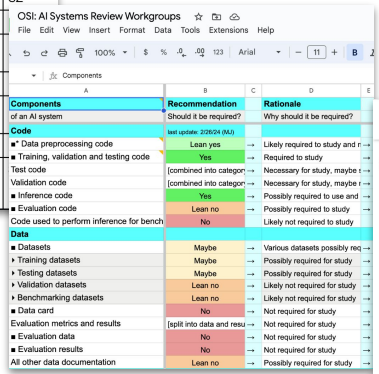
Selecting the required components described in the checklist and preferred form section

1. Workgroup Votes

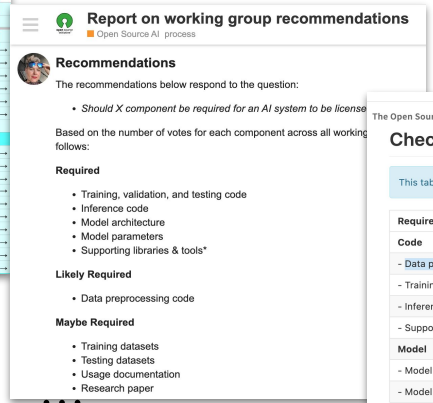
Code All code used to parse and process data, including:	Required to Use?	Required to Study?
Data preprocessing code		SZ
Training code		SZ
Test code		
Code used to perform inference for benchmark tests		
Validation code		
Inference code	SM EL DT SM JT SZ	
Evaluation code		
Other libraries or code artifacts that are part of the system, such as tokenizers and hyperparameter search code, if used.	BG,EL, SM, SZ	

components from Model Openness Framework (MOF)

2. Public Vote Compilation

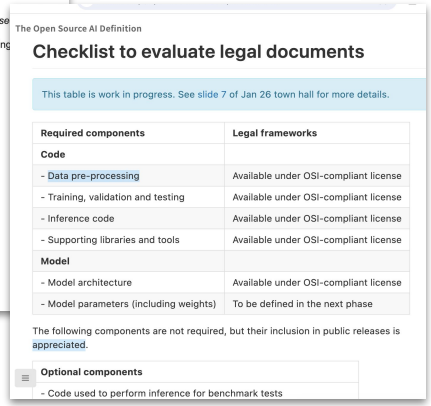


3. Public Results Report on Forum



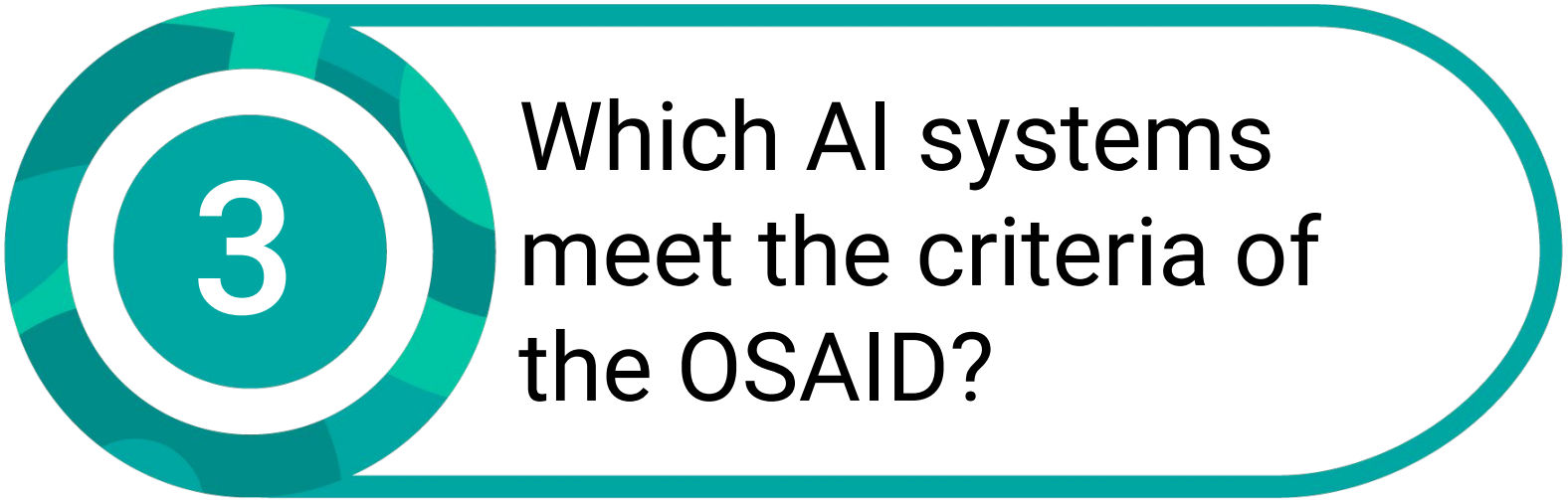
March 1, 2024

4. Definition v.0.0.6 + Checklist



March 10, 2024

OSAID Co-Design Question



Validation Reviewers

*We were interested in reviewing about 10 AI systems self-described as open to validate the definition. Work began **May 1, 2024.***

1. Arctic

1. **Jesús M. Gonzalez-Barahona** Universidad Rey Juan Carlos

2. BLOOM

2. **Danish Contractor** BLOOM Model Gov. Work Group
3. **Jaan Li** University of Tartu, One Fact Foundation

3. Falcon

1. **Casey Valk** Nutanix
2. **Jean-Pierre Lorre** LINAGORA, OpenLLM-France

4. Grok

1. **Victor Lu** independent database consultant
2. **Karsten Wade** Open Community Architects

5. Llama 2

1. **Davide Testuggine** Meta
2. **Jonathan Torres** Meta
3. **Stefano Zacchiroli** Polytechnic Institute of Paris
4. **Victor Lu** independent database consultant

6. LLM360

5. **[Team member TBD]** LLM360

7. Phi-2

6. **Seo-Young Isabelle Hwang** Samsung

8. Mistral

1. **Mark Collier** OpenInfra Foundation
2. **Jean-Pierre Lorre** LINAGORA, OpenLLM-France
3. **Cailean Osborne** University of Oxford, Linux Foundation

9. OLMo

4. **Amanda Casari** Google
5. **Abdoulaye Diack** Google

10. OpenCV

1. **Rasim Sen** Oasis Software Technology Ltd.

11. Pythia

1. **Seo-Young Isabelle Hwang** Samsung
2. **Stella Biderman** EleutherAI
3. **Hailey Schoelkopf** EleutherAI
4. **Aviya Skowron** EleutherAI

12. T5

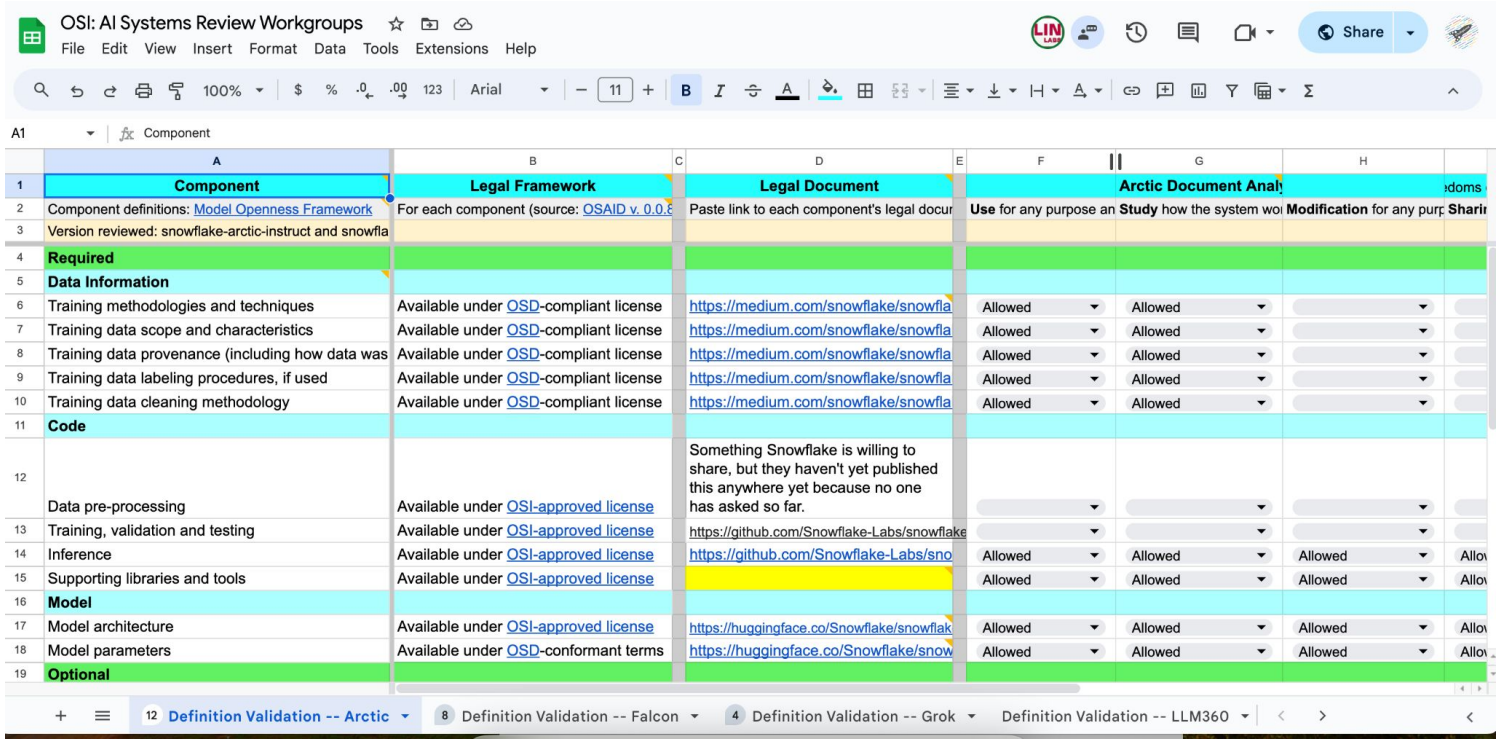
5. **Jaan Li** University of Tartu, One Fact Foundation

13. Viking

6. **Merlijn Sebrechts** Ghent University

Validation Review

Each system is reviewed on a public form, to maximize transparency.



OSI: AI Systems Review Workgroups

File Edit View Insert Format Data Tools Extensions Help

Share

A1 | Component

	A	B	D	F	G	H
1	**Component**	**Legal Framework**	**Legal Document**		**Arctic Document Analy**	
2	Component definitions: Model Openness Framework	For each component (source: OSAID v. 0.0.8	Paste link to each component's legal docur	**Use** for any purpose an	**Study** how the system wo	**Modification** for any purp
3	Version reviewed: snowflake-arctic-instruct and snowfla					
4	**Required**					
5	**Data Information**					
6	Training methodologies and techniques	Available under OSD-compliant license	https://medium.com/snowflake/snowfla	Allowed	Allowed	
7	Training data scope and characteristics	Available under OSD-compliant license	https://medium.com/snowflake/snowfla	Allowed	Allowed	
8	Training data provenance (including how data was	Available under OSD-compliant license	https://medium.com/snowflake/snowfla	Allowed	Allowed	
9	Training data labeling procedures, if used	Available under OSD-compliant license	https://medium.com/snowflake/snowfla	Allowed	Allowed	
10	Training data cleaning methodology	Available under OSD-compliant license	https://medium.com/snowflake/snowfla	Allowed	Allowed	
11	**Code**					
12	Data pre-processing	Available under OSI-approved license	Something Snowflake is willing to share, but they haven't yet published this anywhere yet because no one has asked so far.			
13	Training, validation and testing	Available under OSI-approved license	https://github.com/Snowflake-Labs/snowflake			
14	Inference	Available under OSI-approved license	https://github.com/Snowflake-Labs/sno	Allowed	Allowed	Allowed
15	Supporting libraries and tools	Available under OSI-approved license		Allowed	Allowed	Allowed
16	**Model**					
17	Model architecture	Available under OSI-approved license	https://huggingface.co/Snowflake/snowflak	Allowed	Allowed	Allowed
18	Model parameters	Available under OSD-conformant terms	https://huggingface.co/Snowflake/snow	Allowed	Allowed	Allowed
19	**Optional**					

12 Definition Validation -- Arctic | 8 Definition Validation -- Falcon | 4 Definition Validation -- Grok | Definition Validation -- LLM360

Validation Progress

Validation is ongoing. We have found that creator participation is in most cases necessary to identify all the legal documents needed to ascertain openness. Last updated in June, 2024.

AI System	Meets OSAID requirements?	Notes
Name of system with link to its review sheet	Based on OSAID v. 0.0.8 and/or v.0.0.6	Summary explanation of status (as of 6/11/24)
Arctic	Expect Yes	Verbal confirmation from Snowflake, which is adding legal documents to review sheet (6/3/24)
BLOOM	Confirmed No (license fails)	Usage restrictions in RAIL license
Falcon	Expect No	Documents on training methodologies and techniques and training, validation and testing are missing
Grok	Expect No	Very little public information on system
Llama 2	Confirmed No	Data pre-processing + training, validation and testing code are not available
LLM360	Expect Yes	Self-certified as compliant on the forum, awaiting addition of reviewable documents to their sheet
Mistral	Confirmed No	Some data information and code components missing, no training code available
OLMo	Expect Yes	Supporting libraries and tools unclear, but all other legal documentation is present
OpenCV	Unclear	Model requirement unclear because OpenCV does not store, but instead supports external deep learning frameworks
Phi-2	Unclear	Data information, code, and model information missing
Poro	Unclear	Most review documentation not yet located; Located documentation meets OSAID requirements
Pythia	Confirmed Yes	Only non-alignment was absence of labeling documentation, which was not created. v 0.0.8 adds "if used" to requirement, resolving this
T5	Expect Yes	Only possible restriction is in supporting libraries and tools because gcloud command requires special hardware. Hardware requirements are out of scope for the OSAID, so this is likely not a recognized restriction.

Current Version: 0.0.9

Open Source AI Definition

4 Freedoms	Open Weights	Open Code	Data Info
• Use • Study • Modify • Share	• Model weights and parameters	• Source code used to train and run the system	• Dataset or detailed information about the data used to train the system

The AI conundrums

> “If we assume, for example, that the definition requires full release of datasets, one thing is certain: in Julia’s words, it would be “a definition for which few existing systems qualify.”
> ***(OSI note: also less powerful and limited to specific domains***

https://redmonk.com/sograady/2024/07/03/ai-conundrums/

A long history of exceptions

> “The GNU C library uses a special kind of copyleft called the GNU Library General Public License, which gives permission to link proprietary software with the library. Why make this exception? [...] It is not a matter of principle;[...] but strategically it seems that disallowing them would do more to discourage use of the GNU system than to encourage development of free applications

Richard Stallman
https://www.gnu.org/philosophy/fsfs/rms-essays.pdf

Board Guidance

The OSI Board requires a definition that is:

Supported by diverse stakeholders

The definition needs to have approval by end users, developers, deployers and subjects of AI, globally.

Provides real-life examples

The definition must include relevant examples of AI systems that comply with it at the time of approval, so cannot have an empty set.

Ready by October 2024

A usable version of the definition needs to be ready for approval by the board at the October board meeting.

Approved June 21, 2024

Open Source AI Definition

The general structure of the document

Basic concepts

The Open Source AI Definition

The definition of preferred form to make modifications to machine learning

Clarifications

The Open Source AI Definition

version 1.0-RC1

Definitions

- AI system[1]: An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.
- Machine learning[2]: is a set of techniques that allows machines to improve their performance and usually generate models in an automated manner through exposure to training data, which can help identify patterns and regularities rather than through explicit instructions from a human. The process of improving a system's performance using machine learning techniques is known as "training".

Preamble

Why we need Open Source Artificial Intelligence (AI)

Open Source has demonstrated that massive benefits accrue to everyone after removing the barriers to learning, using, sharing and improving software systems. These benefits are the result of using licenses that adhere to the Open Source Definition. For AI, society needs at least the same essential freedoms of Open Source to enable AI developers, deployers and end users to enjoy those same benefits: autonomy, transparency, frictionless reuse and collaborative improvement.

What is Open Source AI

When we refer to a "system," we are speaking both broadly about a fully functional structure and its discrete structural elements. To be considered Open Source, the requirements are the same, whether applied to a **system**, a **model, weights and parameters**, or other structural elements.

An *Open Source AI* is an AI system made available under terms and in a way that grant the freedoms[3] 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 for any purpose, including to change its output.
- **Share** the system for others to use with or without modifications, for any purpose.

These freedoms apply both to a fully functional system and to discrete elements of a system. A precondition to exercising these freedoms is to have access to the preferred form to make modifications to the system.

Preferred form to make modifications to machine-learning systems

The preferred form of making modifications to a machine-learning system includes all the elements below:

- **Data information**: Sufficiently detailed information about the data used to train the system so that a skilled person can build a substantially equivalent system, together with the code requirements listed below. Data Information shall be made available under terms that allow the copying, modification, and redistribution of the information.
  - In particular, this must include: (1) a detailed description of all data used for training, including (if used) of unshareable data, disclosing the provenance of the data, its scope and characteristics, how the data was obtained and selected, the labeling procedures and data cleaning methodologies; (2) a listing of all publicly available training data and where to obtain it; (3) a listing of all training data obtainable from third parties and where to obtain it, including for a fee.
- **Code**: The complete source code used to train and run the system. The Code shall represent the full specification of how the Data Information was processed and how the training was done. Code shall be made available under OSI-approved licenses.
  - For example, if used, this must include code used for pre-processing data, code used for training, validation and testing, supporting libraries like tokenizers and hyperparameters search code, inference code, and model architecture.
- **Parameters**: The model parameters, such as weights or other configuration settings. Parameters shall be made available under OSI-approved terms[4].
  - For example, this might include checkpoints from key intermediate stages of training as well as the final optimizer state.

The licensing or other contractual terms applied to these elements and to any combination thereof may contain conditions that require any modified version to be released under the same license as the original.

Open Source models and Open Source weights

For machine learning systems,

- An **AI model** consists of the model architecture, model parameters (including weights) and inference code for running the model.
- **AI weights** are the set of learned parameters that overlay the model architecture to produce an output from a given input.

The preferred form to make modifications to machine learning systems also applies to these individual components. "Open Source models" and "Open Source weights" must include the data information and code used to derive those parameters.

1. Recommendation of the Council on Artificial Intelligence OECD/LEGAL/0449, Organization for Economic and Co-operation Development (OECD), 2024 ↩
2. Explanatory memorandum on the updated OECD definition of an AI system, OECD Artificial Intelligence Papers, No. 8, OECD Publishing, Paris ↩
3. These freedoms are derived from the Free Software Definition. ↩
4. The Open Source AI Definition does not take any stance as to whether model parameters require a license, or any other legal instruments, and whether they can be legally

Unlikely to change in the future

Most likely to change in the future

New FAQ entry (preview)

Open training data: data that can be copied, preserved and reshared. It provides the best way to enable users to study the system.

Public training data: data that others can inspect as long as it remains available. This also enables users to study the work. However, this data can degrade as links or references are lost or removed from network availability. To obviate this, different communities will have to work together to define standards, procedures, tools and governance models to overcome this risk, and Data Information is required in case the data becomes later unavailable..

Obtainable training data: data that can be obtained, including for a fee. This information provides transparency and is similar to a purchasable component in an open hardware system. The Data Information provides a means of understanding this data other than obtaining or purchasing it. This is an area that is likely to change rapidly and will need careful monitoring to protect Open Source AI developers.

Unshareable non-public training data: data that cannot be shared for explainable reasons, like Personally Identifiable Information (PII). For this class of data, the ability to study some of the system's biases demands a detailed description of the data – what it is, how it was collected, its characteristics, and so on – so that users can understand the biases and categorization underlying the system.

Stakeholder Groups

The Board requires that the definition have approval by end users, developers, deployers and subjects of AI, globally.

Stakeholder	Description	Example	Vols
1. **Developer**	Makes AI system and/or component that will be studied, used, modified, or shared through an open source license	ML researcher in academia or industry	**31%**
2. **Deployer**	Seeks to study, use modify, or share an open source AI system	AI engineer in industry, health researcher in academia	**46%**
3. **End User**	Consumes a system output, but does not seek to study, use, modify, or share the system	Student using a chatbot to write a report, artist creating an image	**≈ 90%**
4. **Subject**	Affected upstream or downstream by a system output without interacting with it intentionally + advocates for this group.	Photographer who finds their image in training dataset (upstream), mortgage applicant evaluated by a bank's AI system (downstream)	**≈ 100%**

2024 Timeline

System testing work stream

Stakeholder consultation work stream

Release schedule

February	April	July	August	September	October
Call For Volunteers + Activity Feedback and Revision	**Virtual System Review**	**Virtual System Review**	**Virtual System Review**	**Feedback and review Hardening WG**	**Feedback and review Hardening WG**
Bi-Weekly Virtual Public Townhalls	Bi-Weekly Virtual Public Townhalls	Townhalls + **- OSPOs for Good** (NYC) **- OS Community Africa** (virtual)	Townhalls + **- AI-dev** (Hong Kong) **- OSC** (Beijing)	Weekly townhalls **- DL Indaba** (Dakar) **- FOSSIndia** (Bangalore) **- RegenAI** (Ashland) **- LF Europe** (Vienna) **- Nerdearla** (Buenos Aires)	Weekly townhalls **- All Things Open** (Raleigh) **- Data Workshop** (Paris)
Draft 0.0.5	Draft 0.0.8	Draft 0.0.8	Draft 0.0.9	RC1	Version 1.0

Hola 👋

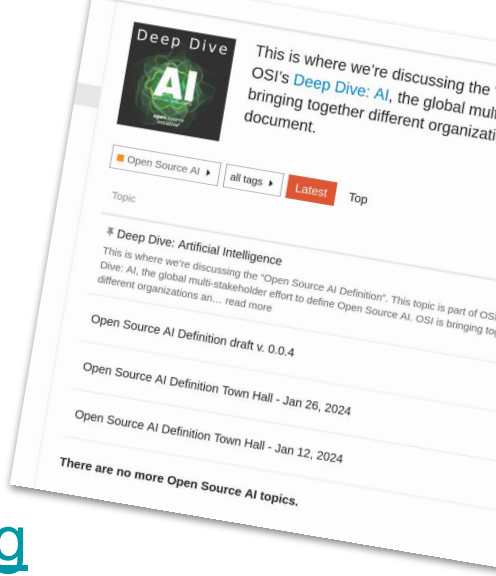

Call for Public Participation
Endorse the OSAID!
opensource.org/osaid-endorse
Mer Joyce
#NERDEARLA

Endorse the OSAID!

- In preparation for next month's launch, we are seeking both **individual** and **organizational** endorsements of the OSAID.
- **"Endorsement" means** your name and organizational affiliation will be appended to a press release announcing the OSAID.
- Use OSI's **webform** to give your endorsement.
- Email or DM me on the forum with any questions.

How to Participate

- Endorse the OSAID!
  - opensource.org/osaid-endorse
- Public forum: discuss.opensource.org
- Comment on RC1 next week
- Become an OSI Member (free + paid)
- Weekly virtual town halls



Thank You, Co-Designers!

*The following individuals volunteered as workgroup members or system reviewers during the co-design process. This group of **36 volunteers** represents **23 countries** by birth and residence. African participants are highlighted.*

- **Rahmat Akintola** Cubeseed Africa
- **George C. G. Barbosa** Fundação Oswaldo Cruz
- **Stella Biderman** EleutherAI
- **Amanda Casari** Google
- **Justin Colannino** Microsoft
- **Mark Collier** OpenInfra Foundation
- **Daniel Brumund** GIZ FAIR Forward - AI for All
- **Danish Contractor** BLOOM Model Gov. WG
- **Abdoulaye Diack** Google
- **Ignatius Ezeani** Lancaster University
- **Jesús M. Gonzalez-Barahona** Universidad Rey Juan Carlos
- **Bastien Guerry** DINUM, French public administration
- **Kevin Harerimana** CMU Africa
- **Seo-Young Isabelle Hwang** Samsung
- **Ezequiel Lanza** Intel
- **Jaan Li** University of Tartu, One Fact Foundation
- **Jean-Pierre Lorre** LINAGORA, OpenLLM-France
- **Victor Lu** independent database consultant
- **Tlamelo Makati** WiMLDS Gaborone, Technological University Dublin
- **Satya Mallick** OpenCV
- **David Manset** ITU
- **Phil Nelson** OpenCV
- **Cailean Osborne** University of Oxford, Linux Foundation
- **Ofentse Phuti** WiMLDS Gaborone
- **Caleb Fianku Quao** Kwame Nkrumah University of Science and Technology, Kumasi
- **Hailey Schoelkopf** EleutherAI
- **Rasim Sen** Oasis Software Technology Ltd.
- **Roman Shaposhnik** Apache Software Foundation
- **Aviya Skowron** EleutherAI
- **Minyechil Alehegn Tefera** Mizan Tepi University
- **Davide Testuggine** Meta
- **Jonathan Torres** Meta
- **Akosua Twumasi** Ghana Health Service
- **Casey Valk** Nutanix
- **Stefano Zacchiroli** Polytechnic Institute of Paris
- **Mo Zhou** Debian, Johns Hopkins University

THANKS!

Q+A