

OPEN SOURCE AI DEFINITION

Online public townhall

April 19, 2024

last updated: April 18, 2024 (MJ)



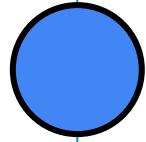
Community agreements

- **One Mic, One Speaker** -- Please allow one person to speak at a time.
- **Take Space, Make Space** -- If you tend to talk more, we invite you to make space for others to share. If you tend not to share, we invite you to speak up.
- **Kindness** -- This work is hard, but we don't have to be. Gentleness and curiosity help. Those who use insults or hate speech will need to leave the meeting.
- **Forward Motion** -- We advance by focusing on what is possible in the moment and doing it. Obstacles are marked for later discussion, not used to stop the process. If we hit a boulder, we note it on the map and keep walking. We'll come back and unearth it later on.
- **Solution-Seeking** -- This work is so complex that focusing on what won't work will stop it. Suggesting new ideas, options, and proposals is vulnerable, but crucial. All of us are needed to make this work.
- **Anything else?**

A wide-angle photograph of a large audience seated in rows of theater-style seating, facing a stage that is mostly out of frame. The seating is dark, and the audience members are diverse in age and attire. The background shows multiple levels of balconies filled with spectators. The lighting is dramatic, with spotlights visible on the ceiling and some audience members holding up phones to take pictures.

OSI's objective for 2024

Open Source AI Definition



Open Source AI Definition **Where Are We Now?**

Open Source AI Definition Elements

v.0.0.7

Preamble

Out of Scope Issues

4 Freedoms

Legal Checklist

Preamble

Why we need Open Source Artificial Intelligence (AI)

Open Source has demonstrated that massive benefits accrue to everyone when you remove the barriers to learning, using, sharing and improving software systems. These benefits are the result of the four freedoms that are granted by open source licenses. These benefits can be summarized as autonomy, transparency, and collaborative improvement.

Everyone needs these benefits in AI. We need essential freedoms to enable users to build and destroy AI systems that are reliable and transparent.

Out of scope issues

The Open Source AI Definition doesn't try to develop and design an AI system that is ethical, law-abiding or responsible, although it doesn't prevent it. The efforts to discuss the responsible development, deployment and use of AI systems, including through appropriate government regulation, are a separate conversation.

What is Open Source AI

An Open Source AI is an AI system made available under terms that grant the freedoms:

- **Use** the system for any purpose without having to ask for permission.
- **Study** the system's code and 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.

Precondition to exercise these freedoms is to have access to the preferred form to make modifications to the system.

Checklist to evaluate machine learning systems

This checklist is based on the paper "The Model Openness Framework: Promoting Completeness and Openness for Reproducibility, Transparency and Usability in AI" published Mar 21, 2024.

Preferred form to make modifications to machine-learning systems

The default set of components required for a machine learning Open Source AI are:

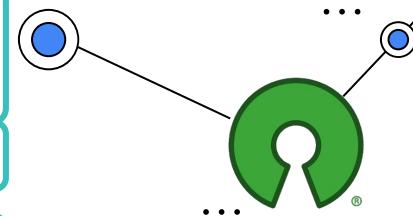
- **Data transparency**: Sufficiently detailed information on how the system was trained. This may include the training methodologies and techniques, the training data sets used, information about the provenance of those data sets, their scope and characteristics, how the data was obtained and selected, the labeling procedures and data cleaning methodologies.
- **Code**: The code used for pre-processing data, the code used for training, validation and testing, the supporting code like tokenizers and hyperparameters search (if used), the serving code, and any other artifacts.
- **Model**: The model parameters, including weights. Where applicable, these should include checkpoints from key intermediate stages of training as well as the final optimizer state.

Table of default required components

Required components	Legal frameworks
Code	Available under OSI-compliant license
- Data re-processing	Available under OSI-compliant license
- Training, validation and testing	Available under OSI-compliant license
- Inference	Available under OSI-compliant license
- Resources libraries and tools	Available under OSI-compliant license
Model	Available under OSI-compliant license
- Model architecture	Available under OSI-compliant license
- Model parameters (including weights)	Available under terms compatible with Open Source principles
Data transparency	
- Training methodologies and techniques	Available under open documentation license
- Training data scope and characteristics	Available under open documentation license
- Training data provenance (including how data was obtained and selected)	Available under open documentation license
- Training data labeling procedures, if used	Available under open documentation license
- Training data cleaning methodology	Available under open documentation license

The following components are not required, but their inclusion in releases is appreciated:

Optional components	
Code	
- Code used to perform inference for benchmark tests	
- Evaluation code	
Data	All data sets, including:
- Training data sets	
- Testing data sets	
- Validation data sets	
- Benchmarking data sets	
Data cards	
- Evaluation metrics and results	
- All other data documentation	
Model	All model elements, including
	Model architecture



Scan me!

Open Source AI Definition Elements

v.0.0.7

Preamble

Preamble

Why we need Open Source Artificial Intelligence (AI)

Open Source has demonstrated that massive benefits accrue to everyone when you remove the barriers to learning, using, sharing and improving software systems. These benefits are the result of the four freedoms that are granted under the OSI definition of open source. These can be summarized as autonomy, transparency, and collaborative improvement.

Everyone needs these benefits in AI. We need essential freedoms to enable users to build and destroy AI systems that are reliable and transparent.

Out of Scope Issues

Out of scope issues

The Open Source AI Definition doesn't say how to develop and design an AI system that is ethical, trustworthy or responsible, although it doesn't prevent it. The efforts to discuss the responsible development, deployment and use of AI systems, including through appropriate government regulation, are a separate conversation.

4 Freedoms

What is Open Source AI

An Open Source AI is an AI system made available under terms that grant the freedoms:

- **Use**: The system for any purpose without having to ask for permission.
- **Study**: The system's inner workings and 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.

Precondition to exercise these freedoms is to have access to the preferred form to make modifications to the system.

Checklist to evaluate machine learning systems

This checklist is based on the paper "The Model Openness Framework: Promoting Completeness and Openness for Reproducibility, Transparency and Usability in AI" published Mar 21, 2024.

Preferred form to make modifications to machine-learning systems

The default set of components required for a machine learning Open Source AI are:

- **Data transparency**: Sufficiently detailed information on how the system was trained. This may include the training methodologies, the training data sets used, information about the provenance of those data sets, their scope and characteristics, how the data was obtained and selected, the labeling procedures and data cleaning methodologies.
- **Code**: The code used for pre-processing data, the code used for training, validation and testing, the supporting code like tokenizers and hyperparameters search (if used), the inference code, and any other artifacts.
- **Model**: The model parameters, including weights. Where applicable, these should include checkpoints from key intermediate stages of training as well as the final optimizer state.

Table of default required components

Required components	Legal frameworks
Code	Available under OSI-compliant license
- Data re-processing	Available under OSI-compliant license
- Training, validation and testing	Available under OSI-compliant license
- Inference	Available under OSI-compliant license
- Resources libraries and tools	Available under OSI-compliant license
Model	Available under OSI-compliant license
- Model architecture	Available under OSI-compliant license
- Model parameters (including weights)	Available under terms compatible with Open Source principles
Data transparency	
- Training methodologies and techniques	Available under open documentation license
- Training data scope and characteristics	Available under open documentation license
- Training data provenance (including how data was obtained and selected)	Available under open documentation license
- Training data labeling procedures, if used	Available under open documentation license
- Training data cleaning methodology	Available under open documentation license

The following components are not required, but their inclusion in releases is appreciated:

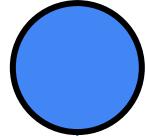
Optional components	
Code	
- Code used to perform inference for benchmark tests	
- Evaluation code	
Data	All data sets, including:
- Training data sets	
- Testing data sets	
- Validation data sets	
- Benchmarking data sets	
Data cards	
- Evaluation metrics and results	
- All other data documentation	
Model	All model elements, including:

Legal Checklist

Done ... ish?



Revising
draft



Open Source AI Definition

The Co-Design Process

Fall 2023: The 4 Freedoms

- The 4 Freedoms for AI
- Use • Study • Modify • Share

What should these
open source principles mean
for artificial intelligence?

Co-Design Workshop: Raleigh

All Things Open | October 2023



Co-Design Workshop: Monterey

Linux Foundation Member Summit | October 2023



open source
initiative®

Co-Design Workshop: Addis Ababa

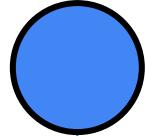


open source
initiative®

Open Source AI Definition

- **The 4 Freedoms for AI**

- - 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.



Open Source AI Definition **The Co-Design Process**

Winter 2023-24: Required Components

- Required Components for Open Source AI

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

Co-Design Workshop: San Jose

AI.dev | December 2023

Group Instructions

1: Introduce 10 minutes

- Name
- Pronouns
- “The way I interact with AI is...”

2: Brainstorm 30 minutes

- Prompt: *For your group’s AI system, how should the four freedoms apply to the components code, model, and data for the system to be licensed as open source?*
- Generate edit options without judgment.
- Share opinions and information with others in your group.

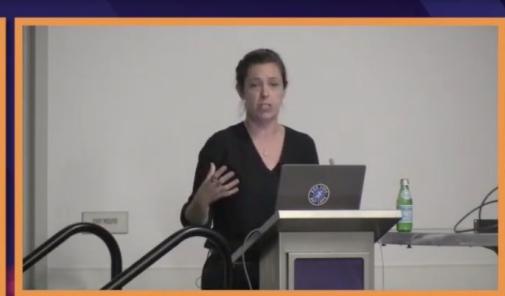
3: Write 10 minutes

- Write your conclusions on the butcher paper to document it.
- Decide how to summarize your recommendations in a few sentences.

Select Roles

- **Moderator:** Ensure your group moves through the steps on time
- **Spokesperson:** Report the group’s edits to the main group

28



open source
initiative®

● Virtual Workgroups

- Selected to represent a diversity of approaches to AI openness:
 1. **Llama 2**: commercial project, accompanied by limited amount of science and with a restrictive license
 2. **BLOOM**: open science project, with lots of details released but shared with a restrictive license
 3. **Pythia**: open science project, with a permissive license
 4. **OpenCV**: open source project, with ML components outside of the generative AI space

Workgroup Members

To achieve better global representation, we conducted outreach to Black, Indigenous, and other people of color, particularly women and individuals from the Global South.

Llama 2

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

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

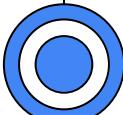
Pythia

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 all workgroup participants are people of color.

OpenCV

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. **Tlameko 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.



Workgroups: Required Component Selection

Component List

The Model Openness Framework: Promoting Completeness and Openness for Reproducibility, Transparency and Usability in AI

Matt White^{1,2}, Ibrahim Haddad², Cailean Osborne^{2,3},
Xiao-Yang (Yanglet) Liu^{1,4}, Ahmed Abdelmonsef¹, Sachin Varghese¹

¹LF AI & Data - Generative AI Commons, ²Linux Foundation,

³University of Oxford, ⁴Columbia University

matt.white@berkeley.edu, ibrahim@linuxfoundation.org,
cailean.osborne@oii.ox.ac.uk, x12427@columbia.edu,
{ahmed.abdelmonsef,sachin.varghese}@genaicommons.org

Abstract

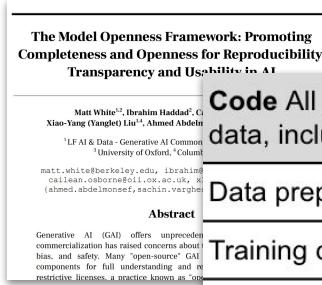
Generative AI (GAI) offers unprecedented possibilities but its commercialization has raised concerns about transparency, reproducibility, bias, and safety. Many "open-source" GAI models lack the necessary components for full understanding and reproduction, and some use restrictive licenses, a practice known as "openwashing." We propose the

Available on arXiv
CC BY-NC-SA 4.0



Workgroups: Required Component Selection

Component List



Component Votes

Code All code used to parse and process data, including:	Required to Use?	Required to Study?	Required to Modify?	Required to Share?
Data preprocessing code		SZ	SZ EL	
Training code		SZ	SZ	
Test code				
Code used to perform inference for benchmark tests				
Validation code			SZ	
Inference code	SM EL DT SM JT SZ		SZ	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	SZ	SZ	SZ

Example:
Llama 2
Workgroup



Workgroups: Required Component Selection

Component List

The Model Openness Framework: Promoting Completeness and Openness for Reproducibility.

Transparency and Usability in AI

Matt White^{1,2}, Ibrahim Haddad¹, Gaojun Xu¹, Xiao-Yang (Canyi) Liu^{1,4}, Ahmed Abdeltawab¹, Calliean Osborne^{1,5}, and Ahmed Alabdullah¹

¹LF AI & Data - Generative AI Committee, University of Oxford, ²Oxford, matt.white@berkeley.edu, ibrahim.haddad@ai.ox.ac.uk, xiaoyang.liu@ai.ox.ac.uk, ahmed.abdeltawab@ai.ox.ac.uk, calliean.osborne@ai.ox.ac.uk, ahmed.alabdullah@ai.ox.ac.uk

Abstract

Generative AI (GAI) offers unprecedented commercialization has raised concerns about bias, and safety. Many "open-source" GAI components for full understanding and responsible licenses, a practice known as "tagging".

Evaluation code

Other libraries or code artifacts that are part of the system, such as tokenizers and hyperparameter search code, if used.

Component Votes

Vote Compilation

OSI: AI Systems Review Workgroups

File Edit View Insert Format Data Tools Extensions Help

Components

A	B	C	D	E	F	G	H	I	J	K	L	M
Components	Recommendation	Rationale	Total	Vote Tally (MOF update)				Legend				
of an AI system	Should it be required?	Why should it be required?	Votes	Study	Use	Modify	Share					
■ Data preprocessing code	Lean yes	→ Likely required to study and run →	29	17	-2	13	1	Yes = Required (>2μ votes)				
■ Training, validation and testing code	Yes	→ Required to study →	39	24	2	13	0	Lean Yes = Likely required (<2μ-1.5μ votes)				
Test code	[combined into category]	→ Necessary for study, maybe run →	4	4	0	0	0	Maybe = Possibly required (<1.5μ-μ votes)				
Validation code	[combined into category]	→ Necessary for study, maybe run →	2	2	0	1	-1	Lean No = Likely not required (<μ-.5μ votes)				
■ Inference code	Yes	→ Possibly required to use and run →	39	11	13	7	8	No = Not required (<.5μ votes)				
■ Evaluation code	Lean no	→ Possibly required to study →	15	10	2	2	1					
Code used to perform inference for benchmark tests	No	→ Likely not required to study →	3	6	1	1	-5	μ = average votes per component				
Data												
■ Datasets	Maybe	→ Various datasets possibly required →	17	12	0	5	1					
▶ Training datasets	Maybe	→ Possibly required for study →	20	13	1	6	0					
▶ Testing datasets	Maybe	→ Possibly required for study →	19	14	1	4	0					
▶ Validation datasets	Lean no	→ Likely not required for study →	13	9	-1	5	0					
▶ Benchmarking datasets	Lean no	→ Likely not required for study →	15	10	-1	4	2					
■ Data card	No	→ Not required for study →	1	6	-3	-1	-1					
Evaluation metrics and results	[split into data and results]	→ Not required for study →	-1	4	-3	-1	-1					
■ Evaluation data	No	→ Not required for study →	3	7	-3	0	-1					
■ Evaluation results	No	→ Not required for study →	5	9	-3	0	-1					
All other data documentation	Lean no	→ Possibly required for study →	13	10	-1	4	0					

16

As of 2/29/24 μ =

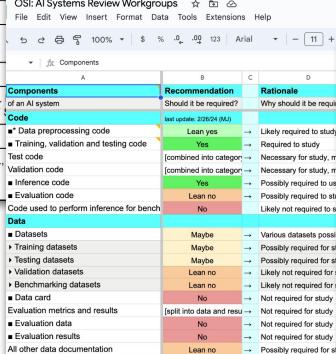
Workgroups: Required Component Selection

Component List

The Model Openness Framework: Promoting Completeness and Openness for Reproducibility.		
Transparency and Usability in AI		
Matt White ^{1,2} , Ibrahim Haddad ¹ , Gaurav Kumar ¹ , Xiao-Yang (Canyi) Liu ^{1,4} , Ahmed Abdelsayed ¹	LF AI & Data - Generative AI Committee	University of Oxford, 1 Columbia University, 2UC Berkeley, 3UCL, 4UConn
matt.white@berkeley.edu, ibrahim.haddad@csail.mit.edu, gaurav.kumar@csail.mit.edu, canyi.liu@csail.mit.edu, ahmed.abdelsayed@csail.mit.edu		
Abstract		
Generative AI (GAI) offers unprecedented commercialization has raised concerns about bias, and safety. Many "open-source" GAI components for full understanding and responsible licenses, a practice known as "tag		

Component Votes

Vote Compilation



Recommendation Report

Report on working group recommendations
Open Source AI process

Recommendations

The recommendations below respond to the question:

- Should X component be required for an AI system to be licensed as open?

Based on the number of votes for each component across all working groups, the follows:

Required

- Training, validation, and testing code
- Inference code
- Model architecture
- Model parameters
- Supporting libraries & tools*

Likely Required

- Data preprocessing code

Maybe Required

- Training datasets
- Testing datasets
- Usage documentation
- Research paper

Workgroups: Required Component Selection

Component List

Component Votes

Vote Compilation

The Model Openness Framework: Promoting Completeness and Openness for Reproducibility.

Transparency and Usability in AI

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 E SM JT	
Evaluation code	BG, EL SZ	
Other libraries or code artifacts that are part of the system, such as tokenizers and hyperparameter search code, if used.		

Abstract
Generative AI (GAI) offers unprecedented commercialization has raised concerns about bias, and safety. Many "open-source" GAI components for full understanding and re-use under restrictive licenses, a practice known as "toxic AI".

OSI: AI Systems Review Workgroups

Components	Recommendation	Rationale
Code	Should it be required? (last update: 2/20/24 (MS))	Why should it be required?
■ Data preprocessing code	Lean yes	→ Likely required to study
■ Training, validation and testing code	Yes	→ Required to study
■ Inference code	Yes	→ Necessary for study, must be open source
■ Evaluation code	Lean no	→ Possibly required to use
Code used to perform inference for benchmark tests	No	→ Possibly required to use
Data	Maybe	→ Various datasets pose challenges
■ Datasets	Maybe	→ Possibly required for study
■ Training datasets	Maybe	→ Possibly required for study
■ Testing datasets	Lean no	→ Likely not required for study
■ Validation datasets	Lean no	→ Likely not required for study
■ Benchmarking datasets	No	→ Not required for study
Data card	No	→ Not required for study
Evaluation metrics and results	(split into data and results)	Not required for study
■ Evaluation data	No	→ Not required for study
■ Evaluation results	No	→ Not required for study
All other data documentation	Lean no	→ Possibly required for study

Recommendation Report

Report on working group recommendations

Recommendations

The recommendations below respond to the question:

- Should X component be required for an AI system to be licensed as open?

Based on the number of votes for each component follows:

Required

- Training, validation, and testing code
- Inference code
- Model architecture
- Model parameters
- Supporting libraries & tools*

Likely Required

- Data preprocessing code

Maybe Required

- Training datasets
- Testing datasets
- Usage documentation
- Research paper

Definition Checklist

Checklist to evaluate legal documents

This table is work in progress. See slide 7 of Jan 26 town hall for more details.

Required components	Legal frameworks
Code	
- Data pre-processing	Available under OSI-compliant license
- Training, validation and testing	Available under OSI-compliant license
- Inference code	Available under OSI-compliant license
- Supporting libraries and tools	Available under OSI-compliant license
Model	
- Model architecture	Available under OSI-compliant license
- Model parameters (including weights)	To be defined in the next phase
Optional components	
- Code used to perform inference for benchmark tests	

v.0.0.6

Open Source AI Definition v.0.0.7

Required Components

Code

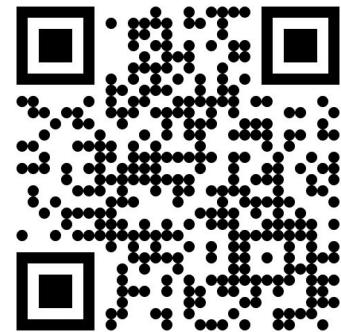
- Data pre-processing
- Training, validation and testing
- Inference
- Supporting libraries and tools

Model

- Model architecture
- Model parameters (including weights)

Data transparency

- Training methodologies and techniques
- Training data scope and characteristics
- Training data provenance (including how data was obtained and selected)
- Training data labeling procedures, if used
- Training data cleaning methodology



open source
initiative®

Open Source AI Definition v.0.0.7

Required Components: Legal Frameworks

Required components	Legal frameworks	Required components	Legal frameworks
Code <ul style="list-style-type: none">- Data pre-processing- Training, validation and testing- Inference- Supporting libraries and tools	Available under OSI-compliant license Available under OSI-compliant license Available under OSI-compliant license Available under OSI-compliant license	Data transparency <ul style="list-style-type: none">- Training methodologies and techniques- Training data scope and characteristics- Training data provenance (including how data was obtained and selected)- Training data labeling procedures, if used- Training data cleaning methodology	Available under OSI-compliant license Available under OSI-compliant license Available under OSI-compliant license Available under OSI-compliant license
Model <ul style="list-style-type: none">- Model architecture- Model parameters (including weights)	Available under OSI-compliant license Available under terms compatible with Open Source principles		Available under OSI-compliant license



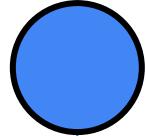
Open Source AI Definition v.0.0.7

Required Components: Legal Frameworks

Required components	Legal frameworks	Required components	Legal frameworks
Code		Data transparency	
- Data pre-processing	Available under OSI-compliant license	- Training methodologies and techniques	Available under OSI-compliant license
- Training, validation and testing	Available under OSI-compliant license	- Training data scope and characteristics	Available under OSI-compliant license
- Inference	Available under OSI-compliant license	- Training data provenance (including how data was obtained and selected)	Available under OSI-compliant license
- Supporting libraries and tools	Available under OSI-compliant license	- Training data labeling procedures, if used	Available under OSI-compliant license
Model		- Training data cleaning methodology	Available under OSI-compliant license
- Model architecture	Available under OSI-compliant license		
- Model parameters (including weights)	Available under terms compatible with Open Source principles		



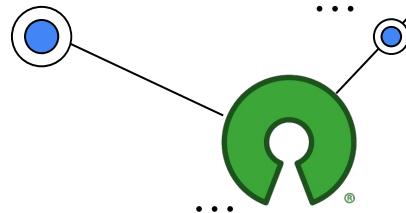
open source
initiative®



Open Source AI Definition

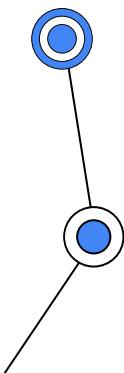
The Co-Design Process

Representation, Inclusion, and Equity



Equitable and inclusive stakeholder representation isn't only about for justice, it's about legitimacy.

A global definition requires global consultation.



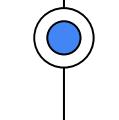


Stakeholder	Description	Example
1. System Creator	Makes AI system and/or component that will be studied, used, modified, or shared through an open source license	ML researcher in academia or industry
2. License Creator	Writes or edits the open source license to be applied to the AI system or component, includes compliance	IP lawyer
3. Regulator	Writes or edits rules governing licenses and systems	government policy-maker
4. Licensee	Seeks to study, use modify, or share an open source AI system	AI engineer in industry, health researcher in academia
5. 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
6. 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)





...



...

Most involved in current phase

Stakeholder	Description	Example
1. System Creator	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
2. License Creator	Writes or edits the open source license to be applied to the AI system or component, includes compliance	IP lawyer
3. Regulator	Writes or edits rules governing licenses and systems	government policy-maker
4. Licensee	Seeks to study, use modify, or share an open source AI system	AI engineer in industry, health researcher in academia
5. 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
6. 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)



open source
initiative®



open source
initiative®



Seeking document reviewers for Pythia and OpenCV

Open Source AI process



Mer

5d

TASK: As part of the systems review track, we're looking for volunteers to review licenses for the Pythia and OpenCV systems and fill out this [spreadsheet](#) 4 to check the compatibility of [version 0.0.6](#) 3 of our definition with current AI systems.

TIMELINE: Our goal was to complete this review by next Tuesday, April 2nd, though we'll likely extend the deadline in consultation with the volunteers who respond.

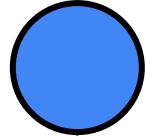
VOLUNTEERS: For transparency, reviewers will have their names and affiliations made public.

Black, Indigenous, Latine, and other people of color, women, queer, transgender, and non-binary people, people with disabilities, and people from poor and working class backgrounds are encouraged to respond.

LEARN MORE Reviewers are already assigned in the Llama 2 and BLOOM groups. We have two reviewers for Pythia and are seeking more. We have no reviewers yet for OpenCV. Further information on the workgroups and their past activities can be found [here](#) 2.



open source
initiative®



Open Source AI Definition

The Co-Design Process

Next Steps



open source
initiative®

2024 Timeline

OSAID v. 0.0.7
last Friday, April
12th

System testing work stream

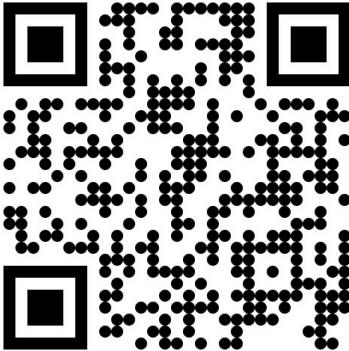
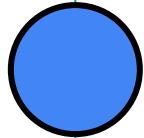
Stakeholder consultation work stream

Release schedule

February	March	April	May	June October
Call For Volunteers + Activity Feedback and Revision	Virtual System Review Meetings Begin	Virtual System Review Meetings Continue	Virtual System Review Meetings END	Feedback Informs Content of OSI In-Person Stakeholder Meeting	Monthly Virtual Meetings
Bi-Weekly Virtual Public Townhalls	Bi-Weekly Virtual Public Townhalls	Bi-Weekly Virtual Public Townhalls	Townhalls + PyCon Workshop (≈ May 17th, Pittsburgh)	Townhall + OSI In-Person Stakeholder Meeting (date + place TBD)	Release stable version
Draft 0.0.5	Draft 0.0.6	Draft 0.0.7	Draft 0.0.8	RC1	Stable Version

In-Person Meetings

Region	Country	City	Conference	Date
North America	United States	Pittsburgh	PyCon US	May 17
Europe	France	Paris	OW2	June
Africa	Nigeria	Lagos	Sustain Africa	June
North America	United States	New York	OSPOs for Good	July 9 - 11
Asia Pacific	China	Hong Kong	AI_dev	August 23
Latin America	Argentina	Buenos Aires	Nerdearla	September
Europe	France	Paris	(data governance)	September
North America	United States	Raleigh	All Things Open	Oct 27 - 29



Open Source AI Definition Stay Connected

- Public forum: discuss.opensource.org
- Become an OSI Member
 - Free or or full
 - SSO with other OSI websites
- Biweekly virtual town halls

The screenshot shows a web interface for the Open Source Initiative's discussion forum. At the top left is the OSI logo. The navigation bar includes 'Topics', 'More', 'Categories', 'Open Source AI' (which is highlighted in orange), and 'All categories'. Below the navigation is a search bar with the placeholder 'Topic'. A sidebar on the left lists 'Deep Dive: Artificial Intelligence' and 'Open Source AI Definition draft v. 0.0.4'. The main content area displays a post titled 'Deep Dive: Artificial Intelligence' with a sub-section for 'Open Source AI Definition Town Hall - Jan 26, 2024' and another for 'Open Source AI Definition Town Hall - Jan 12, 2024'. A message at the bottom states 'There are no more Open Source AI topics.' On the right side of the screen, there are several user profile icons and names.



Q & A



Thank you

We realize this is difficult work and we appreciate your help and openness in improving the definitional process.