OPEN SOURCE AI DEFINITION

Online public townhall

June 14, 2024

last updated: June 11, 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?**

OSI’s objective for 2024

Open Source AI Definition

Open Source AI Definition

Current Version

OSAID v.0.0.8

Open Source AI Definition

v.0.0.8

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 using licenses that adhere to the Open Source Definition. The benefits can be summarized as autonomy, transparency, frictionless reuse, and collaborative improvement.

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

4 Freedoms

What is Open Source AI

- **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.

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

systems

The preferred form of making modifications for a machine-learning Open Source AI must include:

- **Data information**: Sufficiently detailed information about the data used to train the system, so that a skilled person can recreate a substantially equivalent system using the same or similar data.
  - For example, if used, this would 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 source code used to train and run the system.
  - For example, if used, this would 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.
- **Model**: The model parameters.
  - For example, this might include checkpoints from key intermediate stages of training as well as the final optimizer state.

Legal Checklist

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.

Table of default required components

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

The following components are not required as the preferred form of making modifications, but their inclusion in releases is appreciated.

Optional components	Legal frameworks
Data information All data sets, including:	
- Training data sets	Available under OSD-compliant license
- Testing data sets	Available under OSD-compliant license
- Validation data sets	Available under OSD-compliant license
- Benchmarking data sets	Available under OSD-compliant license
- Data card	Available under OSD-compliant license
- Evaluation data	Available under OSD-compliant license

Open Source AI Definition

Key Feedback

OSAID v.0.0.8

Open Source AI Definition **Data Information**

v.0.0.8

Data information
- 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

- Training data sets

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 using licenses that adhere to the Open Source Definition. The benefits can be summarized as autonomy, transparency, frictionless reuse, and collaborative improvement.

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

What is Open Source AI

An Open Source AI is an AI system made available under 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 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.

Preferred form to make modifications to machine-learning systems

The preferred form of making modifications for a machine-learning Open Source AI must include:

- **Data information**: Sufficiently detailed information about the data used to train the system, so that a skilled person can recreate a substantially equivalent system using the same or similar data.
  - For example, if used, this would 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 source code used to train and run the system.
  - For example, if used, this would 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.
- **Model**: The model parameters.
  - For example, this might include checkpoints from key intermediate stages of training as well as the final optimizer state.

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.

Table of default required components

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

The following components are not required as the preferred form of making modifications, but their inclusion in releases is appreciated.

Optional components	Legal frameworks
- Training data sets	Available under OSD-compliant license
- Testing data sets	Available under OSD-compliant license
- Validation data sets	Available under OSD-compliant license
- Benchmarking data sets	Available under OSD-compliant license
- Data card	Available under OSD-compliant license
- Evaluation data	Available under OSD-compliant license

Requiring only **data information**…

…instead of **training datasets** is the greatest point of debate now.

Open Source AI Definition **Other Components**

v.0.0.8

Code	
- Data pre-processing	Available under OSI-approved license

- Data card	Available under OSD-compliant license

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 using licenses that adhere to the Open Source Definition. The benefits can be summarized as autonomy, transparency, frictionless reuse, and collaborative improvement.

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

What is Open Source AI

An Open Source AI is an AI system made available under 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 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.

Preferred form to make modifications to machine-learning systems

The preferred form of making modifications for a machine-learning Open Source AI must include:

- **Data information**: Sufficiently detailed information about the data used to train the system, so that a skilled person can recreate a substantially equivalent system using the same or similar data.
  - For example, if used, this would 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 source code used to train and run the system.
  - For example, if used, this would 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.
- **Model**: The model parameters.
  - For example, this might include checkpoints from key intermediate stages of training as well as the final optimizer state.

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.

Table of default required components

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

The following components are not required as the preferred form of making modifications, but their inclusion in releases is appreciated.

Optional components	Legal frameworks
Data information All data sets, including:	
- Training data sets	Available under OSD-compliant license
- Testing data sets	Available under OSD-compliant license
- Validation data sets	Available under OSD-compliant license
- Benchmarking data sets	Available under OSD-compliant license
- Data card	Available under OSD-compliant license
- Evaluation data	license

Others have proposed…

… removing **data pre-processing code** requirement if training data is not required.

…requiring a **model card**

… and **data card** to standardize system documentation.

Open Source AI Definition **Describing Legal Requirements**

v.0.0.8

Data Information	
- Training methodologies and techniques	Available under OSD-compliant license
- Training data scope and characteristics	Available under OSD-compliant license
- Training data provenance (including how data was obtained and selected)	Available under OSD-compliant license
- Training data labeling procedures, if used	Available under OSD-compliant license
- Training data cleaning methodology	Available under OSD-compliant license

- Model parameters	Available under OSD-conformant terms

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 using licenses that adhere to the Open Source Definition. The benefits can be summarized as autonomy, transparency, frictionless reuse, and collaborative improvement.

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

What is Open Source AI

An Open Source AI is an AI system made available under 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 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.

Preferred form to make modifications to machine-learning systems

The preferred form of making modifications for a machine-learning Open Source AI must include:

- **Data information**: Sufficiently detailed information about the data used to train the system, so that a skilled person can recreate a substantially equivalent system using the same or similar data.
  - For example, if used, this would 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 source code used to train and run the system.
  - For example, if used, this would 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.
- **Model**: The model parameters.
  - For example, this might include checkpoints from key intermediate stages of training as well as the final optimizer state.

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.

Table of default required components

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

The following components are not required as the preferred form of making modifications, but their inclusion in releases is appreciated.

Optional components	Legal frameworks
Data information All data sets, including:	
- Training data sets	Available under OSD-compliant license
- Testing data sets	Available under OSD-compliant license
- Validation data sets	Available under OSD-compliant license
- Benchmarking data sets	Available under OSD-compliant license
- Data card	Available under OSD-compliant license
- Evaluation data	Available under OSD-compliant license

In contrast to the clear "OSI-approved" licenses available for code components…

… the "OSD **compliant**" requirement for data information...

…and "OSD **conformant**" requirement for model parameters have been challenging for reviewers to interpret.

Preferred form to make modifications

Data information

Sufficiently **detailed information** about the data used to train the system, so that a skilled person can **recreate** a substantially **equivalent system** using the **same or similar** data.

Code

The source code used to train and run the system.

Model

The model parameters *(weights and biases)*

Data Information Explained

- The intention of *Data information* is to allow developers to **recreate** a substantially **equivalent system** using **the same or similar** data.
- Came out of the systems review process, with votes by volunteers.

Zooming in on the issues with datasets

- The Pile taken down after an alleged copyright infringement in the US. But legal in Japan. Maybe legal in EU
- DOLMA, initially had a restrictive license. Later switched to a permissive one. Suffers from the same legal uncertainties of the Pile, however the Allen Institute has not been sued, yet.
- Training techniques that preserve privacy like federated learning don’t create datasets.

Alternative proposals

- Use synthetic data: Experimental, unproven technology, limited to corner cases
- All their components must be “open source”: This integralism ignores that even the GNU project accepts system library exceptions and other compromises.

Open Source AI Definition

System Validation

OSAID v.0.0.8

Validation Reviewers

We're interested in reviewing about 10 AI systems self-described as open as part of this definition validation phase. Those marked (*) have were reviewed in previous phases.

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

9. LLM360

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

We will need an independent reviewer for LLM360

8. Mistral

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

7. OLMo

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

8. OpenCV*

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

9. Phi-2

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

10. Pythia*

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

11. T5

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

Viking

6. **Merlijn Sebrechts** Ghent University

Validation Challenges

It is hard for volunteer reviewers to find required documents independently..

OSI: AI Systems Review Workgroups

File Edit View Insert Format Data Tools Extensions Help

	A	B	D	F	G	H
1	**Component**	**Legal Framework**	**Legal Documen**		**Arctic Document Analysis**	Freedoms
2	Component definitions: Model Openness Framework	For each component (source: OSAID v. 0.0.8	Paste link to each componer document below	**Use** for any purpose and without having to ask for permission	**Study** how the system works and inspect its components	**Modification** for any purpose, including to change its output
3	Version reviewed: snowflake-arctic-instruct and snowfla	ke-arctic-base (both seems to have the same characteristics) https://huggin	.o/Snowflake/snowflake-arctic-instruct https://huggingface.co/Snowflake/snowflake-arctic-base			
19	**Optional**					
20	**Data Information** All data sets, including:					
21	Training data sets	Available under OSD-compliant license				
22	Testing data sets	Available under OSD-compliant license				
23	Validation data sets	Available under OSD-compliant license				
24	Benchmarking data sets	Available under OSD-compliant license				
25	Data card	Available under OSD-compliant license				
26	Evaluation data	Available under OSD-compliant license		Allowed	Allowed	Allowed
27	Evaluation results	Available under OSD-compliant license				
28	Other data documentation	Available under OSD-compliant license				
29	**Code**					
30	Code used to perform inference for benchmark tes	Available under OSI-approved license	https://github.com/Snowflake-Labs/snowflake			
31	Evaluation code	Available under OSI-approved license				
32	**Model** All model elements, including:					
33	Model card	Available under OSD-compliant license	https://huggingface.co/Snowflake/snowflake-	Allowed	Allowed	Allowed
34	Sample model outputs	Available under OSD-compliant license				

Definition Validation -- OLMo | Definition Validation -- Phi-2 | Definition Validation -- T5 | Definition Validation -- TBD | Form

Validation Challenges

This meant a lot of the review analysis has been incomplete.

OSI: AI Systems Review Workgroups

File Edit View Insert Format Data Tools Extensions Help

	A	B	C	D	E	F	G	H	
1	**Component**	**Legal Framework**		**Legal Document**			**Arctic Document**	Freedoms	
2	Component definitions: Model Openness Framework	For each component (source: OSAID v. 0.0.8		Paste link to each component's legal document below		**Use** for any purpose and without having to ask for permission	**Study** how the syste works and inspect its components	**odification** for any urpose, including to hange its output	**Sharir** with o for any
3	Version reviewed: snowflake-arctic-instruct and snowfla	ke-arctic-base (both seems to have the same characteristics) https://huggingface.co/Snowflake/snowflake-arctic-instruct https://huggingface						owflake/snowflake-arctic-base	
19	**Optional**								
20	**Data Information** All data sets, including:								
21	Training data sets	Available under OSD-compliant license							
22	Testing data sets	Available under OSD-compliant license							
23	Validation data sets	Available under OSD-compliant license							
24	Benchmarking data sets	Available under OSD-compliant license							
25	Data card	Available under OSD-compliant license							
26	Evaluation data	Available under OSD-compliant license				Allowed	Allowed	Allowed	Allo
27	Evaluation results	Available under OSD-compliant license							
28	Other data documentation	Available under OSD-compliant license							
29	**Code**								
30	Code used to perform inference for benchmark tes	Available under OSI-approved license		https://github.com/Snowflake-Labs/snowflake					
31	Evaluation code	Available under OSI-approved license							
32	**Model** All model elements, including:								
33	Model card	Available under OSD-compliant license		https://huggingface.co/Snowflake/snowflake-		Allowed	Allowed	Allowed	Allo
34	Sample model outputs	Available under OSD-compliant license							

Definition Validation -- OLMo | Definition Validation -- Phi-2 | Definition Validation -- T5 | Definition Validation -- TBD | Form F

Validation Solutions

Having the help of system creators to locate documents has been crucial.

Thank you, Arctic!

OSI: AI Systems Review Workgroups

File Edit View Insert Format Data Tools Extensions Help

A1 | Component

	A	B	C	D	F	G	H	
1	**Component**	**Legal Framework**		**Legal Document**		**Arctic Document Analy**		dom
2	Component definitions: Model Openness Framework	For each component (source: OSAID v. 0.0.8		Paste link to each component's legal	**Use** for any purpose an	**Study** how the system wo	**Modification** for any purp	**Sharin**
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/s ia	Allowed	Allowed		
7	Training data scope and characteristics	Available under OSD-compliant license		https://medium.com/snowflake/ fla	Allowed	Allowed		
8	Training data provenance (including how data was	Available under OSD-compliant license		https://medium.com/snowflake/ wfla	Allowed	Allowed		
9	Training data labeling procedures, if used	Available under OSD-compliant license		https://medium.com/snowflake owfla	Allowed	Allowed		
10	Training data cleaning methodology	Available under OSD-compliant license		https://medium.com/snowflake owfla	Allowed	Allowed		
11	**Code**							
12	Data pre-processing	Available under OSI-approved license		Something Snowflake is willin to share, but they haven't yet pu lished this anywhere yet because n one has asked so far.				
13	Training, validation and testing	Available under OSI-approved license		https://github.com/Snowflake-Labs/snowf ake				
14	Inference	Available under OSI-approved license		https://github.com/Snowflake-Labs/sn	Allowed	Allowed	Allowed	Allo
15	Supporting libraries and tools	Available under OSI-approved license			Allowed	Allowed	Allowed	Allo
16	**Model**							
17	Model architecture	Available under OSI-approved license		https://huggingface.co/Snowflake/snowflak	Allowed	Allowed	Allowed	Allo
18	Model parameters	Available under OSD-conformant terms		https://huggingface.co/Snowflake/snow	Allowed	Allowed	Allowed	Allo
19	**Optional**							

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

Validation Expectations

Given current system information, our expected review results are as follows. If we are missing information, please let us know.

Confirmed Yes	Expect Yes	Unclear	Expect No	Confirmed No
Review complete + all required components present	Review incomplete + all required components expected to be present	Incomplete data, cannot ascertain oppeness	Review incomplete + some or all required components expected to be absent	Review complete + some/all required components absent

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	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
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.

Open Source AI Definition

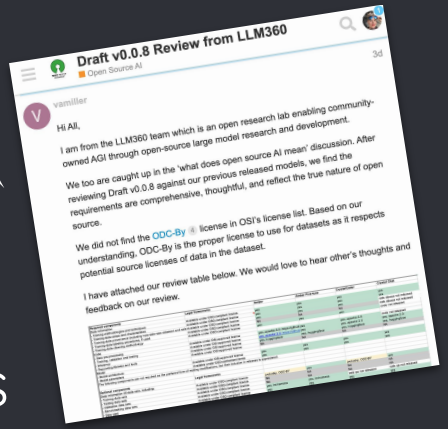
What's Next?

June - October 2024

- Complete validation phase
- Resolve comments, release v. 0.0.9 after validation
- Cut the release candidate with sufficient endorsement

Complete the Validation Phase

Thanks, LLM360!


Draft v0.0.8 Review from LLM360

Open Source AI

3d

vamiller

Hi All,

I am from the LLM360 team which is an open research lab enabling community-owned AGI through open-source large model research and development.

We too are caught up in the 'what does open source AI mean' discussion. After reviewing Draft v0.0.8 against our previous released models, we find the requirements are comprehensive, thoughtful, and reflect the true nature of open source.

We did not find the ODC-By license in OSI's license list. Based on our understanding, ODC-By is the proper license to use for datasets as it respects potential source licenses of data in the dataset.

I have attached our review table below. We would love to hear other's thoughts and feedback on our review.

1. Reach out to **AI system creators** to fill in the blanks on their own systems by pointing us to correct documentation
2. Invite **volunteers** to also help us fill in these blanks

2024 Timeline

System testing work stream
Stakeholder consultation work stream
Release schedule

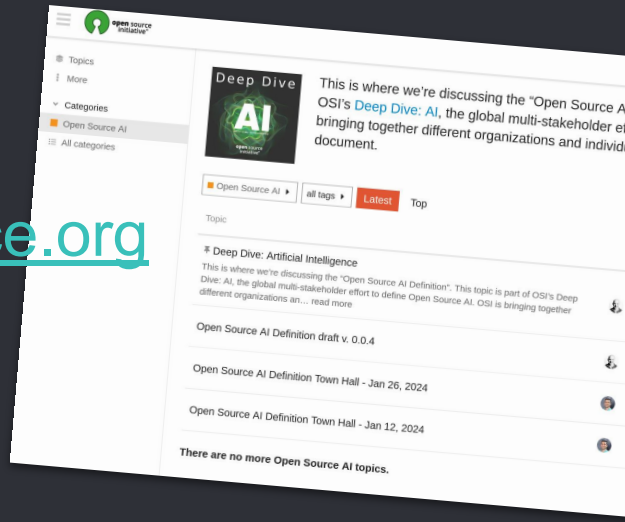
February	March	April	May	June ...	... October
Call For Volunteers + Activity Feedback and Revision	**Virtual System Review Meetings Begin**	**Virtual System Review Meetings Continue**	**Virtual System Review Meetings END**	**Feedback Informs Content of OSI In-Person Stakeholder Meeting**	**Monthly Virtual Meetings**
Bi-Weekly Virtual Public Townhalls	Bi-Weekly Virtual Public Townhalls	Bi-Weekly Virtual Public Townhalls	Townhalls + **PyCon Workshop** (≈ May 17th, Pittsburgh)	Townhall + **Virtual Launch Event** (date TBD)	**Release stable version**
Draft 0.0.5	Draft 0.0.6	Draft 0.0.7 and 8	Draft 0.0.9	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 11 - 12
North America	United States	New York	**OSPOs for Good**	July 9 - 11
Africa	Virtual	Virtual	**Sustain Africa**	July
Asia Pacific	China	Hong Kong	**AI_dev**	August 23
Latin America	Argentina	Buenos Aires	**Nerdearla**	September
Europe	TBD	TBD	**(data governance)**	October
North America	United States	Raleigh	**All Things Open**	Oct 27 - 29

Participation Options

- Public **forum**: discuss.opensource.org
- Become an OSI **member**
  - Free or or full
  - SSO with other OSI websites
- Biweekly virtual **townhalls**… like this one!
- **Volunteer** to help with validation (email or DM Mer Joyce)



Q & A

Thank you

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