OPEN SOURCE AI DEFINITION

Online public townhall

March 22, 2024

last updated: March 21, 2024 (SM)

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?**

The objective for 2024

Open Source AI Definition version 1.0

hackmd.io/@opensourceinitiative/osaid-0-0-5

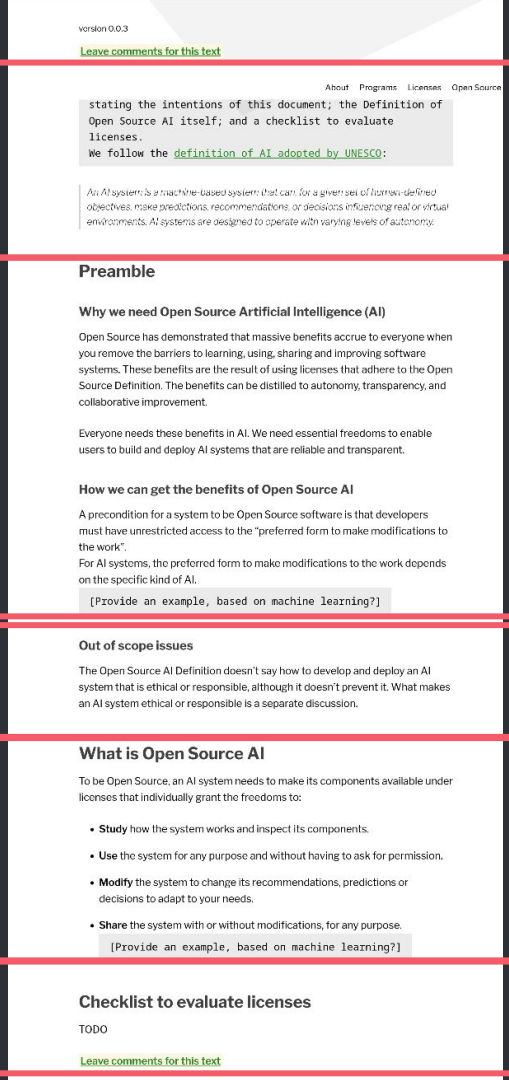
Definition of AI system

Preamble

Out of scope issues

4 freedoms

Legal checklist



version 0.0.3

Leave comments for this text

About Programs Licenses Open Source

stating the intentions of this document; the Definition of Open Source AI itself; and a checklist to evaluate licenses.
We follow the definition of AI adopted by UNESCO:

> *An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.*

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 distilled to autonomy, transparency, 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.

How we can get the benefits of Open Source AI

A precondition for a system to be Open Source software is that developers must have unrestricted access to the "preferred form to make modifications to the work".
For AI systems, the preferred form to make modifications to the work depends on the specific kind of AI.

[Provide an example, based on machine learning?]

Out of scope issues

The Open Source AI Definition doesn't say how to develop and deploy an AI system that is ethical or responsible, although it doesn't prevent it. What makes an AI system ethical or responsible is a separate discussion.

What is Open Source AI

To be Open Source, an AI system needs to make its components available under licenses that individually grant the freedoms to:

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

[Provide an example, based on machine learning?]

Checklist to evaluate licenses

TODO

Leave comments for this text

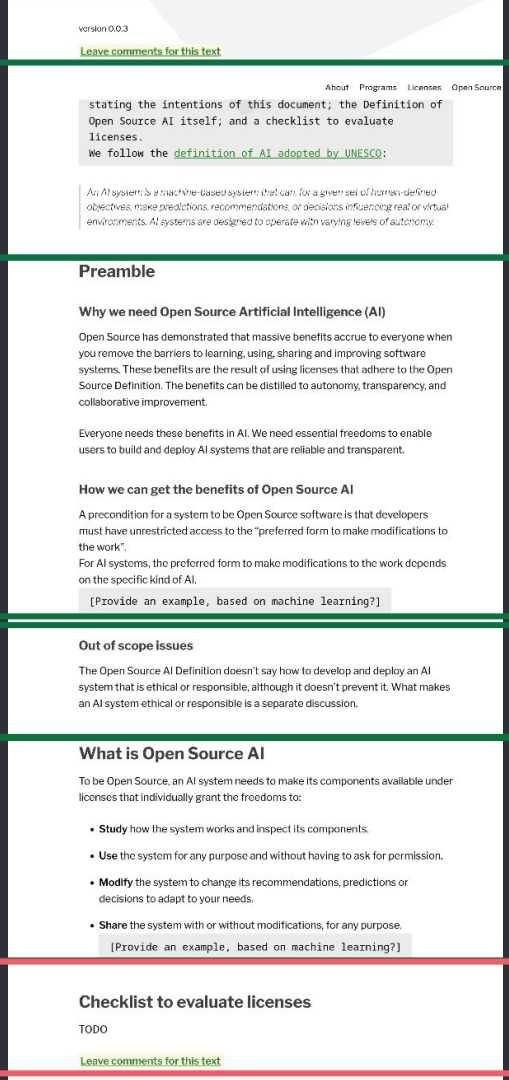
Definition of AI system

Preamble

Out of scope issues

4 freedoms

Legal terms checklist



version 0.0.3

Leave comments for this text

About Programs Licenses Open Source

stating the intentions of this document; the Definition of Open Source AI itself; and a checklist to evaluate licenses.
We follow the definition of AI adopted by UNESCO:

> An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.

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 distilled to autonomy, transparency, 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.

How we can get the benefits of Open Source AI

A precondition for a system to be Open Source software is that developers must have unrestricted access to the "preferred form to make modifications to the work".
For AI systems, the preferred form to make modifications to the work depends on the specific kind of AI.

[Provide an example, based on machine learning?]

Out of scope issues

The Open Source AI Definition doesn't say how to develop and deploy an AI system that is ethical or responsible, although it doesn't prevent it. What makes an AI system ethical or responsible is a separate discussion.

What is Open Source AI

To be Open Source, an AI system needs to make its components available under licenses that individually grant the freedoms to:

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

[Provide an example, based on machine learning?]

Checklist to evaluate licenses

TODO

Leave comments for this text

Done … ish?

Working on

What is Open Source AI

An Open Source AI is an AI system made available to the public 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.

Data: transparency requirements only

Precondition to exercise these freedoms is to have access to the preferred form to make modifications to the system. For machine learning systems that means having public access to:

- **Data**: Sufficiently detailed information on how the system was trained, including 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 libraries like tokenizers and hyperparameters search code (if used), the inference code, and the model architecture.
- **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.

We need to talk about “systems” because openness is a combination of availability of multiple artifacts

Alt: we talk about “open weights” only

What phase 2 will look like

For each AI system, build a table like:

Required component	Link to resource	Legal framework
Data pre-processing code	URL	OSI-approved license
Training, validation and testing code	URL	…
Inference code	URL	…
Supporting libraries and tools	URL	…
Model architecture	URL	…
Model parameters	URL	???

Getting the specifications

AI systems	List of components	Legal frameworks	Legal documents	**Checklist**
Active working groups: - Llama2 - Pythia Setting up: - BLOOM - OpenCV	What elements are necessary to: - use - study - modify - share an AI system?	For each artifact, evaluate which laws apply. Some will be under “Intellectual Property” regimes, some will be under other regimes.	We’ll match the components and the identified legal frameworks with the terms of the legal documents already in use, where available.	After repeating this exercise enough times, we’ll be able to generalize the outcomes and write the specs to evaluate the freedoms granted.

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	Bi-Weekly Virtual Public Townhalls	Townhall + **OSI In-Person Stakeholder Meeting** (date + place TBD)	**Release version 1.0**
Draft 0.0.5	Draft 0.0.6	Draft 0.0.7	Draft 0.0.8	RC1	v. 1.0

What phase 2 will look like

For each AI system, build a table like:

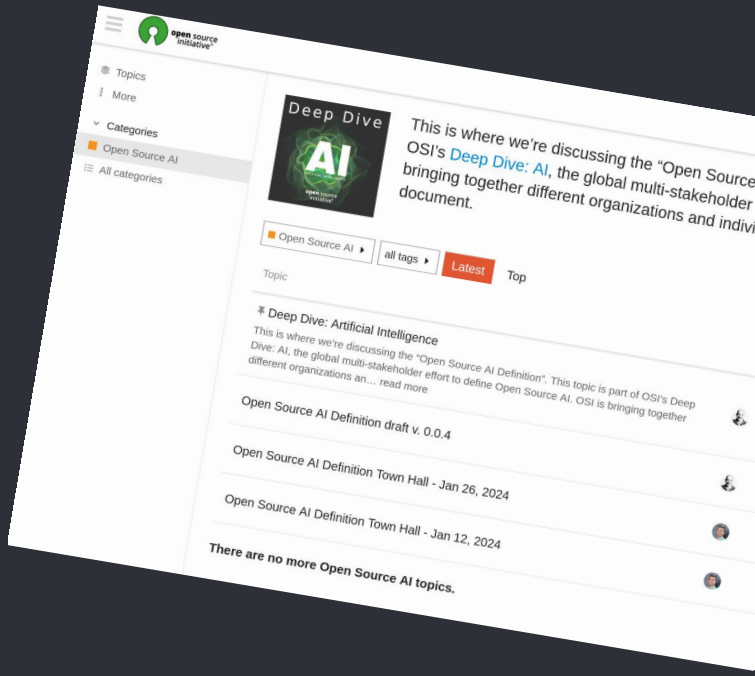
Required component	Link to resource	Legal framework
Data pre-processing code	URL	OSI-approved license
Training, validation and testing code	URL	…
Inference code	URL	…
Supporting libraries and tools	URL	…
Model architecture	URL	…
Model parameters	URL	???

Deep Dive AI in-person meetings

Region	Country	City	Conference	Date
North America	United States	Pittsburgh	PyCon US	May 20 - 23
Europe	?			May ?
Africa	Nigeria	Abuja	OSCA	June 6 - 8
Latin America	Mexico	Mexico D.F.	Latam OSS	July 19 - 20
Asia Pacific	Hong Kong	Hong Kong	AI_dev	August 23
North America	United States	Raleigh	All Things Open	Oct 27 - 29

Join the conversation

- discuss.opensource.org
- Public forum
- Join as OSI member
  - Free or full
  - SSO with other OSI websites



Q & A

Thank you

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

Criteria for RC1 and v. 1.0

RC1

- Expected outcome of in-person meeting end May/early June!
- The draft is completed in all its parts
- The draft is supported by at least 2 representatives for each of the 6 stakeholder groups

version 1

- Expected outcome of in-person and online meetings through the summer/early autumn
- The draft is endorsed by at least 5 reps for each of the stakeholder groups
- Announced in late October

Help us find stakeholders

System Creator	License Creator	Regulator	Licensee	End User	Subject
Makes AI system and/or component that will be studied, used, modified, or shared through an open source license (e.g., ML researcher in academia or industry)	Writes or edits the open source license to be applied to the AI system or component; includes compliance (e.g., IP lawyer)	Writes or edits rules governing licenses and systems (e.g. government policy-maker)	Seeks to study, use modify, or share an open source AI system (e.g. AI engineer, health researcher, education researcher)	Consumes a system output, but does not seek to study, use, modify, or share the system (e.g., student using a chatbot to write a report, artist creating an image)	Affected upstream or downstream by a system output without interacting with it intentionally; includes advocates for this group (e.g. people with loan denied, or content creators)
✅	✅	⚠️	✅	⚠️	⚠️
Enough to start	Enough to start	Leads to US, EU, Singapore, no commitment yet	Enough to start	Which org is squarely in this space?	ACLU, Algorithmic Justice League