

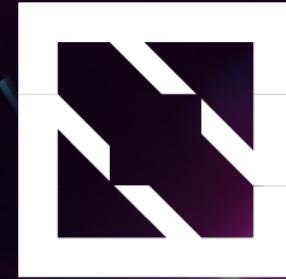


# KubeCon

THE LINUX FOUNDATION



China 2024



# CloudNativeCon





KubeCon



CloudNativeCon



China 2024



# Unveiling the Future: **Nurturing Openness in AI Development**

## 揭示未来: 培育人工智能开放性发展

Anni Lai

Chair, Generative AI Commons, LF  
AI & Data  
Head of Open Source Operations,  
Futurewei

Mer Joyce

Founder  
Do Big Good LLC  
Co-Design Facilitator, Open Source AI Definition  
Open Source Initiative

## Who We Are

200+  
Active Members

80+  
Organizations

The Generative AI Commons is:

- A community-driven open membership initiative represented by non-profits, academia, and industry in a neutral forum within LF AI & Data
- Shaping the future of Generative AI through open-source and open-science
- Embracing responsible and trustworthy AI, fostering collaboration, promoting AI literacy, and advocating for our community

# Generative AI Commons: It's All about the People !



KubeCon



CloudNativeCon



China 2024



## Workstream Leaders



Anni Lai  
Generative AI Commons Chair  
[LinkedIn](#)



Arnaud Le Hors  
Generative AI Commons Vice Chair  
[LinkedIn](#)



Ahmed Abdelmonsef  
Framework Workstream Lead  
[LinkedIn](#)



Alexy Khrabrov  
Framework Workstream Lead  
[LinkedIn](#)



Nick Chase  
Model and Data Workstream Lead  
[LinkedIn](#)



Sriram Easwaran  
Model and Data Workstream Lead  
[LinkedIn](#)



Sachin Varghese  
Applications Workstream Lead  
[LinkedIn](#)



Raghavan Muthuregundanathan  
Applications Workstream Lead & Education and Outreach Workstream Lead  
[LinkedIn](#)



Susan Malaika  
Responsible AI Workstream Lead  
[LinkedIn](#)



Suparna Bhattacharya  
Responsible AI Workstream Lead  
[LinkedIn](#)



Ofer Hermoni  
Education and Outreach Workstream Lead  
[LinkedIn](#)

A community built and supported by "**volunteers**" who are passionate about GenAI

- a. Build **thought leadership** in GenA via hosting guest speakers and webinars, publishing white papers/blogs/research papers/education materials, etc
- b. Uncover **open source projects** in the GenAI area
- c. Build **GenAI technology landscape**, solution stacks/reference architecture, best practices, governance guideline, etc.
- d. Build a **GenAI community** (which is different from the traditional open source community)

OLF AI & DATA

# Generative AI Commons: Workstreams and Deliverables



KubeCon



CloudNativeCon



China 2024



## MAD (Models, Application, Data)

- Building a GenAI landscape
- Build GenAI solution stack/reference architecture/ecosystem
- Identify open source projects in the MAD area for LF AI & Data
- Host open discussions and build thought leadership in the MAD area
- Support LF AI & Data GenAI projects/help grow ecosystem if needed

[Generative AI Commons  
\(gенаicommmons.org\)](https://gенаicommmons.org)

## Frameworks

- Model Openness Framework
- Model Openness Tool ([isitopen.ai](https://isitopen.ai))
- MOF 1.0 spec

## Edu. & Outreach

- Education and training
- Thought leadership (blogs, white papers, webinars, etc)
- Open strategies in GenAI Research
- Glossary posted
- AI\_Dev, OSS AI events

## Responsible AI

- Responsible AI Framework (RAF)
- Identify open source projects in R. AI area
- NIST engagement/AI policy update

# The Importance of Openness in AI

- Fosters Innovation
- Democratizes Access
- Enhances Transparency
- Accelerates Learning
- Promotes Ethical AI
- Encourages Standardization
- Mitigates Risks of Bias
- Supports Sustainability
- Increases Trust and Adoption



# Efforts for Model Transparency



China 2024

crfm.stanford.edu/fmti/May-2024/index.html

## The Foundation Model Transparency Index

A comprehensive assessment of the transparency of foundation model developers

Paper (May 2024) Transparency Reports Data Board Press Past versions

Center for Research on Foundation Models

### Foundation Model Transparency Index Scores by Major Dimensions of Transparency, May 2024

Source: May 2024 Foundation Model Transparency Index

Major Dimensions of Transparency	A D E P T	AI21labs	ALPHALPHA	amazon	ANTHROPIC	servicenow	Google	IBM	Meta	Microsoft	MISTRAIL	OpenAI	stability.ai	WRITER	Average
	Fuyu-8B	60%	40%	0%	10%	100%	0%	60%	40%	40%	20%	20%	40%	50%	34%
	Data	0%	60%	40%	0%	10%	100%	0%	60%	40%	40%	20%	20%	40%	50%
	Labor	0%	43%	71%	14%	14%	100%	29%	43%	29%	100%	100%	14%	100%	43%
	Compute	14%	86%	100%	0%	14%	100%	14%	100%	71%	57%	14%	14%	43%	86%
	Methods	0%	100%	100%	50%	75%	100%	75%	100%	75%	100%	100%	50%	75%	100%
	Model Basics	83%	100%	100%	83%	50%	100%	83%	100%	100%	100%	100%	50%	100%	100%
	Model Access	100%	67%	100%	67%	67%	100%	67%	67%	100%	100%	100%	67%	100%	33%
	Capabilities	80%	80%	100%	80%	100%	100%	80%	60%	100%	100%	100%	100%	60%	100%
	Risks	0%	57%	57%	43%	86%	100%	43%	71%	71%	29%	14%	57%	14%	14%
	Mitigations	0%	40%	20%	20%	40%	0%	40%	80%	60%	0%	60%	60%	0%	20%
	Distribution	57%	86%	100%	57%	86%	100%	57%	86%	71%	71%	71%	86%	71%	77%
	Usage Policy	40%	100%	100%	80%	100%	100%	100%	40%	40%	100%	40%	80%	60%	80%
	Feedback	67%	100%	67%	33%	33%	100%	67%	67%	33%	67%	33%	67%	33%	60%
	Impact	29%	29%	29%	0%	14%	14%	29%	0%	14%	0%	14%	14%	14%	14%
	Average	36%	73%	76%	41%	53%	86%	53%	67%	62%	66%	62%	49%	58%	57%

Source: Stanford Center for Research on Foundation Models

# Challenges in Open Source AI

- Rapidly growing # of available models
- A lot of “Open Washing”
- Many Licenses with restrictions
- No consistent definition of “open” in AI
- Lack of understanding of license implications
- Many components not released (datasets)
- Illegally converted licenses
- Open source license used on non-software



# Model Openness Framework (MOF)



arXiv > cs > arXiv:2403.13784

Search... All fields Search  
Help | Advanced Search

Computer Science > Machine Learning

[Submitted on 20 Mar 2024 (v1), last revised 3 Jun 2024 (this version, v3)]

## The Model Openness Framework: Promoting Completeness and Openness for Reproducibility, Transparency, and Usability in Artificial Intelligence

Matt White, Ibrahim Haddad, Cailean Osborne, Xiao-Yang Liu Yanglet, Ahmed Abdelmonsef, Sachin Varghese

Generative AI (GAI) offers unprecedented opportunities for research and innovation, but its commercialization has raised concerns about transparency, reproducibility, and safety. Many open GAI models lack the necessary components for full understanding and reproducibility, and some use restrictive licenses whilst claiming to be ``open-source''. To address these concerns, we propose the Model Openness Framework (MOF), a ranked classification system that rates machine learning models based on their completeness and openness, following principles of open science, open source, open data, and open access. The MOF requires specific components of the model development lifecycle to be included and released under appropriate open licenses. This framework aims to prevent misrepresentation of models claiming to be open, guide researchers and developers in providing all model components under permissive licenses, and help individuals and organizations identify models that can be safely adopted without restrictions. By promoting transparency and reproducibility, the MOF combats ``openwashing'' practices and establishes completeness and openness as primary criteria alongside the core tenets of responsible AI. Wide adoption of the MOF will foster a more open AI ecosystem, benefiting research, innovation, and adoption of state-of-the-art models.

Comments: 22 pages

Subjects: Machine Learning (cs.LG); Artificial Intelligence (cs.AI); Computers and Society (cs.CY); Software Engineering (cs.SE)

Cite as: arXiv:2403.13784 [cs.LG]  
(or arXiv:2403.13784v3 [cs.LG] for this version)  
<https://doi.org/10.48550/arXiv.2403.13784>

**Access Paper:**

- View PDF
- HTML (experimental)
- TeX Source
- Other Formats

[view license](#)

Current browse context:  
cs.LG  
< prev | next >  
new | recent | 2403

Change to browse by:  
cs  
  cs.AI  
  cs.CY  
  cs.SE

References & Citations

- NASA ADS
- Google Scholar
- Semantic Scholar

Export BibTeX Citation

Bookmark

# MOF Components



KubeCon



CloudNativeCon



THE LINUX FOUNDATION  
OPEN SOURCE SUMMIT



China 2024



Evaluation Code



Preprocessing Code



Model Architecture



Libraries & Tools



Training Code



Inference Code



Datasets



Evaluation Data



Sample Model Outputs\*



Model Weights & Parameters



Model Metadata



Configuration File



Data Card



Research Paper



Evaluation Results



Model Card



Technical Report

# MOF Acceptable Licenses



KubeCon



CloudNativeCon

THE LINUX FOUNDATION  
OPEN SOURCE SUMMIT

China 2024

Component	Domain	Content Type	Accepted Open License
Datasets	Data	Data	Preferred: CDLA-Permissive-2.0, CC-BY-4.0 Acceptable: Any including unlicensed
Data Preprocessing Code	Data	Code	Acceptable: OSI-approved
Model Architecture	Model	Code	Acceptable: OSI-approved
Model Parameters	Model	Data	Preferred: CDLA-Permissive-2.0 Acceptable: OSI-Approved, Permissive Open Data Licenses
Model Metadata	Model	Data	Preferred: CDLA-Permissive-2.0 Acceptable: CC-BY-4.0, Permissive Open Data Licenses
Training Code	Model	Code	Acceptable: OSI-approved
Inference Code	Model	Code	Acceptable: OSI-approved
Evaluation Code	Model	Code	Acceptable: OSI-approved
Evaluation Data	Model	Data	Preferred: CDLA-Permissive-2.0 Acceptable: CC-BY-4.0, Permissive Open Data Licenses
Evaluation Results	Model	Documentation	Preferred: CC-BY-4.0 Acceptable: Permissive Open Content Licenses
Supporting libraries and Tools	Model	Code	Acceptable: OSI-approved
Model Card	Model	Documentation	Preferred: CC-BY-4.0 Acceptable: Permissive Open Content Licenses
Data Card	Data	Documentation	Preferred: CC-BY-4.0 Acceptable: Permissive Open Content Licenses
Technical Report	Model & Data	Documentation	Preferred: CC-BY-4.0 Acceptable: Permissive Open Content Licenses
Research Paper	Model & Data	Documentation	Preferred: CC-BY-4.0 Acceptable: Permissive Open Content Licenses
Sample Model Outputs	Model	Data or Code	Unlicensed

<u>MOF Class</u>	<u>Components Included</u>
<b>Class I - Open Science</b>	<ul style="list-style-type: none"><li>• Research Paper</li><li>• Datasets (any license or unlicensed)</li><li>• Data Preprocessing Code</li><li>• Model Parameters (intermediate checkpoints)</li><li>• Model Metadata (optional)</li><li>• And all Class II Components</li></ul>
<b>Class II - Open Tooling</b>	<ul style="list-style-type: none"><li>• Training Code</li><li>• Inference Code</li><li>• Evaluation Code</li><li>• Evaluation Data</li><li>• Supporting Libraries &amp; Tools (optional)</li><li>• And all Class III Components</li></ul>
<b>Class III - Open Model</b>	<ul style="list-style-type: none"><li>• Model Architecture</li><li>• Model Parameters (final checkpoint)</li><li>• Technical Report</li><li>• Evaluation Results</li><li>• Model Card</li><li>• Data Card</li><li>• Sample Model Outputs (optional)</li></ul>

# Model Openness Tool (MOT)

isitopen.ai

MODEL OPENNESS FRAMEWORK

Models Evaluate Model Submit Model Logout ( [REDACTED] )

## Models

**Disclaimer**

This is beta software and may contain inaccurate information. If you experience errors with the tool or discover inaccurate information, please use the "report" feature in the model details to report inaccuracies or for site issues please [contact us](#).

▼ Search filters

Name	Organization	Classification	Last updated	Badge
Baichuan-1-13B	Baichuan AI	Class III - Open Model	2024-06-19	<div><span>Class I - Open Science</span> Fail</div> <div><span>Class II - Open Tooling</span> In progress (40%)</div> <div><span>Class III - Open Model</span> Conditional Pass</div>
Arctic-Base	Snowflake	Class III - Open Model	2024-06-19	<div><span>Class I - Open Science</span> Fail</div> <div><span>Class II - Open Tooling</span> In progress (20%)</div> <div><span>Class III - Open Model</span> Conditional Pass</div>
Arctic-Instruct	Snowflake	Class III - Open Model	2024-06-19	<div><span>Class I - Open Science</span> Fail</div> <div><span>Class II - Open Tooling</span> In progress (20%)</div> <div><span>Class III - Open Model</span> Conditional Pass</div>
XGen	Salesforce	Class III - Open Model	2024-06-19	<div><span>Class I - Open Science</span> Fail</div> <div><span>Class II - Open Tooling</span> Fail</div> <div><span>Class III - Open Model</span> Conditional Pass</div>
XGLM	FAIR (Fundamental Artificial Intelligence Research)	Class III - Open Model	2024-06-19	<div><span>Class I - Open Science</span> Fail</div> <div><span>Class II - Open Tooling</span> Fail</div> <div><span>Class III - Open Model</span> Conditional Pass</div>



KubeCon



CloudNativeCon



THE LINUX FOUNDATION  
OPEN SOURCE SUMMIT



AI Dev  
Open Source Dev & ML Summit

China 2024



View  
Model  
Catalogue

DLF AI & DATA

# Please Fill Out the Survey



KubeCon



CloudNativeCon



THE LINUX FOUNDATION  
OPEN SOURCE SUMMIT



China 2024

## 2024 Generative AI Survey



This survey aims to understand the deployment, use, and challenges of generative AI technologies in organizations and the role of open source in this domain. It should take about 10 minutes to answer. Your participation is greatly appreciated.



<https://www.research.net/r/QWTLLV7>



🎉 New Today - v.0.0.9 🎉

# The Open Source AI Definition (OSAID)

Mer Joyce

15

# Open Source Initiative (OSI)



KubeCon



CloudNativeCon



China 2024



The OSI is recognized by individuals, companies, and by public institutions as **the authority that defines Open Source** and stewards the Open Source Definition for software.

A non-profit corporation with global scope formed to educate and advocate for the benefits of open source, the OSI builds bridges among different constituencies in the open source community.



# Open Source AI Definition (OSAID)



KubeCon



CloudNativeCon



THE LINUX FOUNDATION  
OPEN SOURCE SUMMIT



AI\_dev  
Open Source AI & ML Summit

China 2024

For the past **two years**, the OSI has been engaged in a **public process** to create the Open Source AI Definition (OSAID).

I'll introduce you to our  **new** version today, **v.0.0.9**, and the months-long co-design process to create it.



# Co-Designing the OSAID

OSAID co-design  
workshop @ All  
Things Open  
(October, 2023)



KubeCon



CloudNativeCon



THE LINUX FOUNDATION

OPEN SOURCE SUMMIT



Open Source AI & ML Summit

China 2024

## Co-Design

A set of creative methods for  
making public decisions  
**with, not for**, directly  
impacted stakeholders by  
sharing knowledge and  
power.

## Use • Study • Modify • Share

What should these  
open source principles  
mean for artificial  
intelligence?

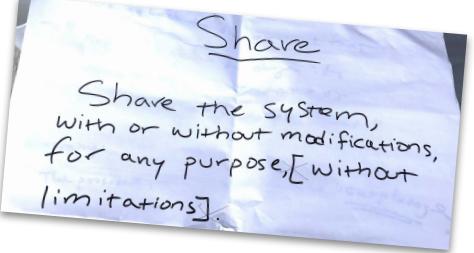


# 4 Freedoms of Open Source AI



China 2024

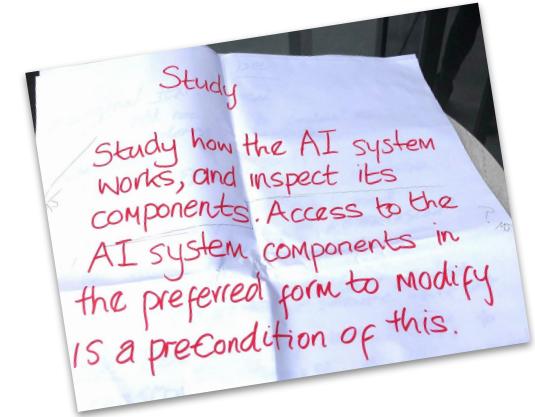
## In-Person Co-Design Workshops



All Things Open Conference  
Raleigh, USA  
October, 2023



Linux Foundation Member Summit  
Monterey, USA  
October, 2023



# 4 Freedoms of Open Source AI



KubeCon



CloudNativeCon



THE LINUX FOUNDATION  
OPEN SOURCE SUMMIT



Open Source Dev & ML Summit

China 2024

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.



# Preferred Form of an OSAI System



KubeCon



CloudNativeCon



THE LINUX FOUNDATION  
OPEN SOURCE  
SUMMIT

China 2024



AI\_dev  
Open Source Dev & ML Summit

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



# Preferred Form of an OSAI System



KubeCon



CloudNativeCon

THE LINUX FOUNDATION  
OPEN SOURCE SUMMITAI\_dev  
Open Source AI&ML Summit

China 2024

## Virtual Co-Design Workgroups

### 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. **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



# Preferred Form of an OSAI System



China 2024

## Virtual Co-Design Process

### 1. Workgroup Votes

Code All code used to parse and process data, including:	Required to Use?	Required to Study?
Data preprocessing code		
Training code		
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	

Component list is from the Model Openness Framework (MOF)

### 2. Public Vote Compilation

The screenshot shows a Google Sheets document titled "OSI: AI Systems Review Workgroups". It contains a survey table with columns for "Components", "Recommendation", and "Rationale". The "Components" column lists various AI system components. The "Recommendation" column has cells colored green (Lean yes), yellow (Yes), orange (Lean no), and red (No). The "Rationale" column provides reasons for each recommendation. A sidebar on the left shows a navigation menu for the workgroups.

Components	Recommendation	Rationale
■ Data preprocessing code	Lean yes	Likely required to study and run
■ Training, validation and testing code	Yes	Required to study
Test code	[combined into category]	Necessary for study, maybe run
Validation code	[combined into category]	Necessary for study, maybe run
■ Inference code	Yes	Possibly required to use and run
■ Evaluation code	Lean no	Possibly required to study
Code used to perform inference for bench	No	Likely not required to study
■ Datasets	Maybe	Various datasets possibly req
▶ Training datasets	Maybe	Possibly required for study
▶ Testing datasets	Maybe	Possibly required for study
▶ Validation datasets	Lean no	Likely not required for study
▶ Benchmarking datasets	Lean no	Likely not required for study
■ Data card	No	Not required for study
Evaluation metrics and results	[split into data and resu]	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

### 3. Public Results Report on Forum

The screenshot shows a forum post titled "Report on working group recommendations" under the "Open Source AI process" category. The post includes a "Recommendations" section with a photo of a user and a "Checklist to evaluate legal documents" section. The "Recommendations" section lists requirements for training, inference, and supporting tools. The "Checklist" section lists legal frameworks for code, model, and data components.

**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?

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

**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
<b>Code</b>	
- 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
<b>Model</b>	
- Model architecture	Available under OSI-compliant license
- Model parameters (including weights)	To be defined in the next phase

### 4. Public Definition Update

v.0.0 6 below



# New Today: OSAlD v.0.0.9



KubeCon



CloudNativeCon



China 2024



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

## Open Weights + Open Code +

Model weights and parameters

Source code used to train and run the system

## Data Information

The dataset or detailed information about the data used to train the system



open source  
initiative®



# New Version: OSAID v.0.0.9



KubeCon



CloudNativeCon



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

# Weights

The model weights and parameters, made available under OSI-approved terms

Examples: checkpoints from key intermediate stages of training as well as the final optimizer state



# New Version: OSAID v.0.0.9



KubeCon



CloudNativeCon



China 2024



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

## Code

The source code used to train and run the system, made available with OSI-approved licenses

Examples: code used for pre-processing data, training, validation and testing, supporting libraries like tokenizers and hyperparameters search code, inference code, and model architecture.



# New Version: OSAID v.0.0.9



KubeCon



CloudNativeCon



The preferred form of making modifications for a machine-learning Open Source AI System 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. Data information shall be made available with licenses that comply with the Open Source Definition.



Examples: training methodologies and techniques, 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.

# Training Data in the OSAID



KubeCon



CloudNativeCon



China 2024



*OSI affirms the benefits of full access to training data while acknowledging it is not always possible for reasons of law, privacy norms, technical feasibility, and cultural practice.*

- **Training data is valuable to study AI systems:** to understand the biases that have been learned, which can impact system behavior. But training data is **not part of the preferred form** for making modifications to an existing AI system
- Data can be hard to share. Laws that permit training on data often **limit the resharing** of that same data to protect copyright or other interests.
- **Privacy rules** also give a person the rightful ability to control their most sensitive information, such as decisions about their health.
- Similarly, much of the world's **Indigenous knowledge** is protected through mechanisms that are not compatible with later-developed frameworks for rights exclusivity and sharing.
- **Open training data** (data that can be reshared) provides the **best way** to enable users to study the system, along with the preferred form of making modifications.
- **Public training data** (data others can inspect as long as it remains available) **also enables users** to study the work, along with the preferred form.



# OSAID Approval Criteria



KubeCon



CloudNativeCon



THE LINUX FOUNDATION  
OPEN SOURCE SUMMIT



Open Source AI & ML Summit

China 2024

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.

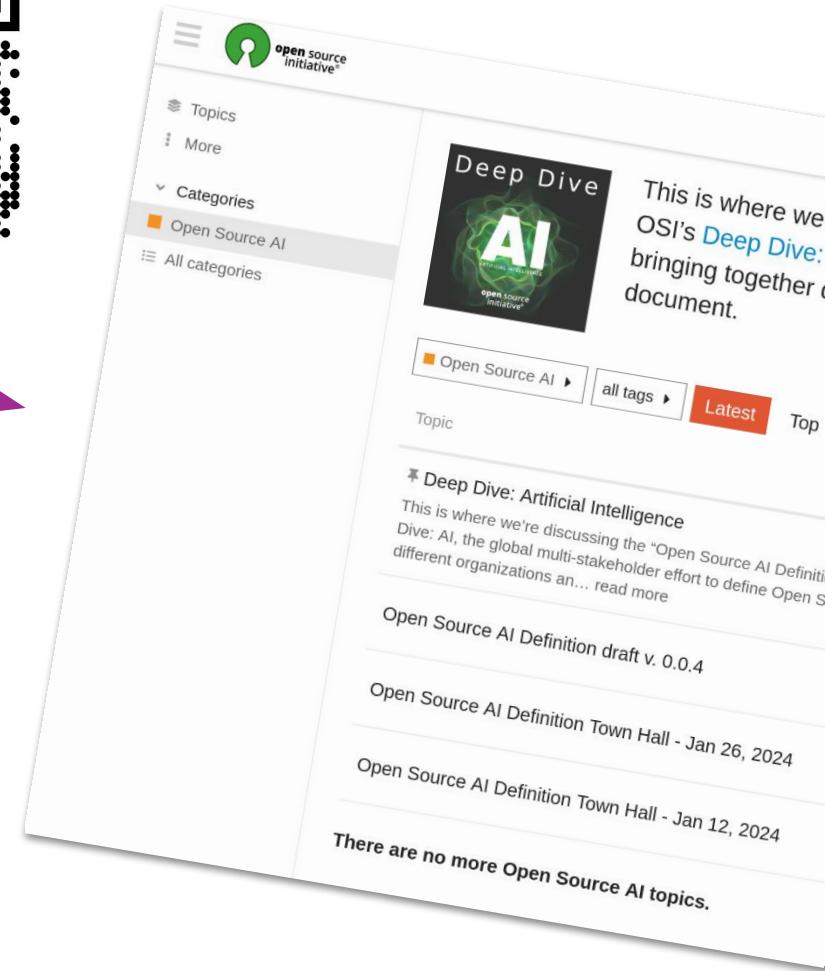
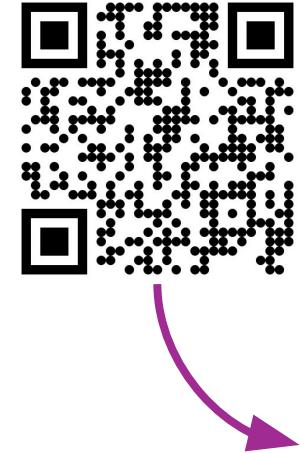
## Ready by October 2024

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



# Join Us!

- Public **forum**:  
[discuss.opensource.org](https://discuss.opensource.org)
- Become an **OSI member**
  - Free or or full
  - SSO with other OSI websites
- Weekly **virtual townhalls**
  - Announced on forum and social media



# Open Source AI Definition v.0.0.9



KubeCon



CloudNativeCon



China 2024



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

**Open  
Weights** +

Model weights and parameters

**Open  
Code** +

Source code used to train and run the system

**Data  
Information**

The dataset or detailed information about the data used to train the system



open source  
initiative®



# Thank You!

谢谢

# Q+A

Anni Lai  
Mer Joyce