

# Meeting of the Technical Advisory Council (TAC)

June 3rd, 2021

 **DLF** AI & DATA

# Antitrust Policy

- › Linux Foundation meetings involve participation by industry competitors, and it is the intention of the Linux Foundation to conduct all of its activities in accordance with applicable antitrust and competition laws. It is therefore extremely important that attendees adhere to meeting agendas, and be aware of, and not participate in, any activities that are prohibited under applicable US state, federal or foreign antitrust and competition laws.
- › Examples of types of actions that are prohibited at Linux Foundation meetings and in connection with Linux Foundation activities are described in the Linux Foundation Antitrust Policy available at <http://www.linuxfoundation.org/antitrust-policy>. If you have questions about these matters, please contact your company counsel, or if you are a member of the Linux Foundation, feel free to contact Andrew Updegrave of the firm of Gesmer Undergone LLP, which provides legal counsel to the Linux Foundation.

# Recording of Calls

## **Reminder:**

TAC calls are recorded and available for viewing on the [TAC Wiki](#)

# Reminder: LF AI & Data Useful Links

- › Web site: [lfaidata.foundation](https://lfaidata.foundation)
- › Wiki: [wiki.lfaidata.foundation](https://wiki.lfaidata.foundation)
- › GitHub: [github.com/lfaidata](https://github.com/lfaidata)
- › Landscape: <https://landscape.lfaidata.foundation> or <https://l.lfaidata.foundation>
- › Mail Lists: <https://lists.lfaidata.foundation>
- › Slack: <https://slack.lfaidata.foundation>
- › Youtube: <https://www.youtube.com/channel/UCfasaeqXJBCAJMNO9HcHfbA>
- › LF AI Logos: <https://github.com/lfaidata/artwork/tree/master/lfaidata>
- › LF AI Presentation Template:  
[https://drive.google.com/file/d/1eiDNJvXCqSZHT4Zk\\_-czASlz2GTBRZk2/view?usp=sharing](https://drive.google.com/file/d/1eiDNJvXCqSZHT4Zk_-czASlz2GTBRZk2/view?usp=sharing)
- ›
- › Events Page on LF AI Website: <https://lfaidata.foundation/events/>
- › Events Calendar on LF AI Wiki (subscribe available):  
<https://wiki.lfaidata.foundation/pages/viewpage.action?pageId=12091544>
- › Event Wiki Pages: <https://wiki.lfaidata.foundation/display/DL/LF+AI+Data+Foundation+Events>

# Agenda

- › Roll Call (1 min)
- › Welcome new General Member - GSI Technologies (1 min)
- › Approval of Minutes from May 20th (3 mins)
- › Graduation Proposal (25 minutes)
  - › Milvus (Jun Gu <jun.gu@zilliz.com>)
- › Invited Presentation (15 minutes)
  - › KOSA.AI (Layla Li <layla@kosa.ai>, Sonali Sanghrakja <sonali@kosa.ai>)
- › Electing New TAC Chair (Jacqueline Cardoso ) (5 minutes)
- › Badging for Projects (Andre de Waal, Mandy Chessell) (5 minutes)
- › Annual Project Review Schedule (Jim Spohrer) (2 minutes)
- › LF AI General Updates (1 minutes)
- › Open Discussion (1 minutes)

# TAC Voting Members

\* = still need backup specified on [wiki](#)

Board Member	Contact Person	Email
AT&T	Anwar Atfab*	<a href="mailto:anwar@research.att.com">anwar@research.att.com</a>
Baidu	Ti Zhou	<a href="mailto:zhouti@baidu.com">zhouti@baidu.com</a>
Ericsson	Rani Yadav-Ranjan*	<a href="mailto:rani.yadav-ranjan@ericsson.com">rani.yadav-ranjan@ericsson.com</a>
Huawei	Huang Zhipeng	<a href="mailto:huangzhipeng@huawei.com">huangzhipeng@huawei.com</a>
IBM	Susan Malaika	<a href="mailto:malaika@us.ibm.com">malaika@us.ibm.com</a>
Nokia	Jonne Soininen	<a href="mailto:jonne.soininen@nokia.com">jonne.soininen@nokia.com</a>
OPPO	Jimin Jia*	<a href="mailto:jjajimin@oppo.com">jjajimin@oppo.com</a>
SAS	Nancy Rausch	<a href="mailto:nancy.rausch@sas.com">nancy.rausch@sas.com</a>
Tech Mahindra	Amit Kumar	<a href="mailto:Kumar_Amit@techmahindra.com">Kumar_Amit@techmahindra.com</a>
Tencent	Bruce Tao	<a href="mailto:brucetao@tencent.com">brucetao@tencent.com</a>
Zilliz	Jun Gu*	<a href="mailto:jun.gu@zilliz.com">jun.gu@zilliz.com</a>
ZTE	Wei Meng	<a href="mailto:meng.wei2@zte.com.cn">meng.wei2@zte.com.cn</a>
Graduate Project	Contact Person	Email
Acumos	Amit Kumar	<a href="mailto:Kumar_Amit@techmahindra.com">Kumar_Amit@techmahindra.com</a>
Angel	Bruce Tao	<a href="mailto:brucetao@tencent.com">brucetao@tencent.com</a>
Egeria	Mandy Chessell	<a href="mailto:mandy_chessell@uk.ibm.com">mandy_chessell@uk.ibm.com</a>
Horovod	Travis Addair*	<a href="mailto:taddair@uber.com">taddair@uber.com</a>
ONNX	Jim Spohrer (Chair of TAC)	<a href="mailto:spohrer@us.ibm.com">spohrer@us.ibm.com</a>
Pyro	Fritz Obermeyer*	<a href="mailto:fritz.obermeyer@gmail.com">fritz.obermeyer@gmail.com</a>

# Welcome new member

LF AI & Data Member Company - General



# Approval of May 6th, 2021 Minutes

Draft minutes from the May 20<sup>th</sup> TAC call were previously distributed to the TAC members via the mailing list

## **Proposed Resolution:**

- › That the minutes of the May 20<sup>th</sup> meeting of the Technical Advisory Council of the LF AI & Data Foundation are hereby approved.



# Graduation Proposal - Milvus

Jun Gu (Zilliz) <jun.gu@zilliz.com>

# Project Graduation Proposal Review/Discussion/Note: Milvus

*Milvus is an open-source vector database built to power AI applications and embedding similarity search. Milvus makes unstructured data search more accessible, and provides a consistent user experience regardless of the deployment environment.*

*Milvus was released under the open-source Apache License 2.0 in October 2019.*

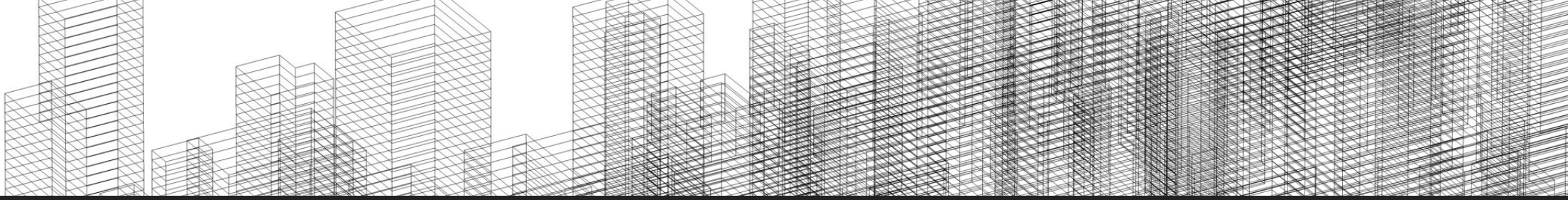
**Presenter:** Jun Gu, Milvus Project Evangelist

**Resources:**

Github: <https://github.com/milvus-io/milvus>

LF AI & Data website: <https://lfaidata.foundation/projects/milvus/>

Project Level Request to TAC: Graduation



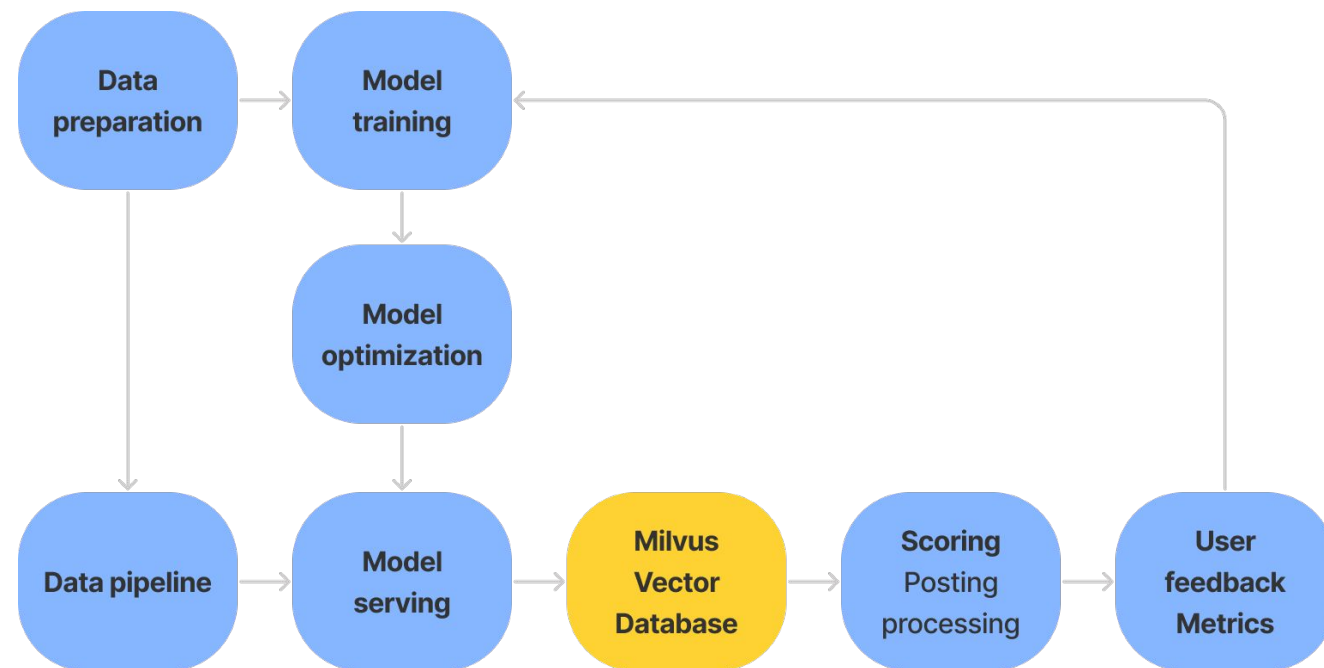
# **Milvus Project: 2021 Update and Request for Graduation**

Jun Gu, Milvus project evangelist

# Milvus Vector Database

<https://milvus.io>

- The embedding serving infrastructure for production AI applications
- Common application scenarios:
  - Computer vision: Content based image retrieval system, etc
  - NLP: Chat bot, document semantic search, etc
  - Personalized recommender system
  - Computational chemistry: molecular analysis
  - etc

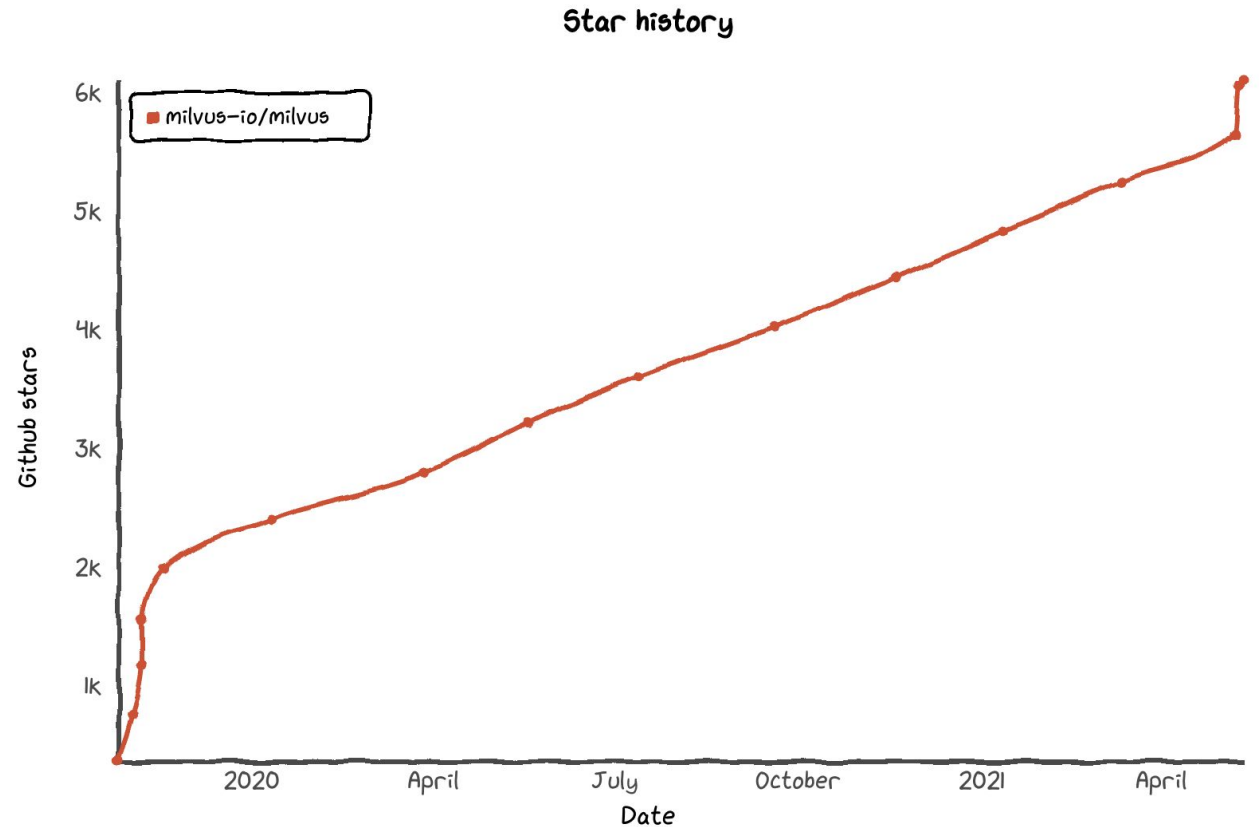


# The graduation requirements

- Have a healthy number of code contributions coming from at least five organizations.
- Have reached a minimum of 1000 stars on GitHub.
- Have achieved and maintained a Core Infrastructure Initiative Best Practices Gold Badge.
- Have demonstrated a substantial ongoing flow of commits and merged contributions for the past 12 months\*.
- Have completed at least one collaboration with another LF AI & Data hosted project.
- Have a technical lead appointed for the representation of the project on the LF AI & Data Technical Advisory Council.

# Milvus community (till May 12th, 2021)

- Contributors: 137
  - Zilliz
  - Beike
  - iQiyi
  - Yandex
  - Montage Tech
  - Arkie
  - ...
- GitHub stars: 6.1K
- GitHub forks: 887
- Docker hub downloads: 320K
- Known community users: 2.3K
  - Slack: 700+
  - Wechat group: 1,700+



# Best Practices Badge

<https://bestpractices.coreinfrastructure.org/en/projects/3563>



## milvus

[Expand panels](#) [Show all details](#) [Hide met & N/A](#)

Projects that follow the best practices below can voluntarily self-certify and show that they've achieved a Core Infrastructure Initiative (CII) badge. [Show details](#)

If this is your project, please show your badge status on your project page! The badge status looks like this: `cii best practices passing` Here is how to embed it: [Show details](#)

These are the `passing` level criteria. You can also view the `silver` or `gold` level criteria.

▼ Basics	13/13 ●
▼ Change Control	9/9 ●
▼ Reporting	8/8 ●
▼ Quality	13/13 ●
▼ Security	16/16 ●
▼ Analysis	8/8 ●

# Commit History

Mar 17, 2019 – May 6, 2021

Contributions: Commits ▾

Contributions to master, excluding merge commits and bot accounts





# Governance and Procedures

- Contributors Guide

<https://github.com/milvus-io/milvus/blob/master/CONTRIBUTING.md>

- Governance Information

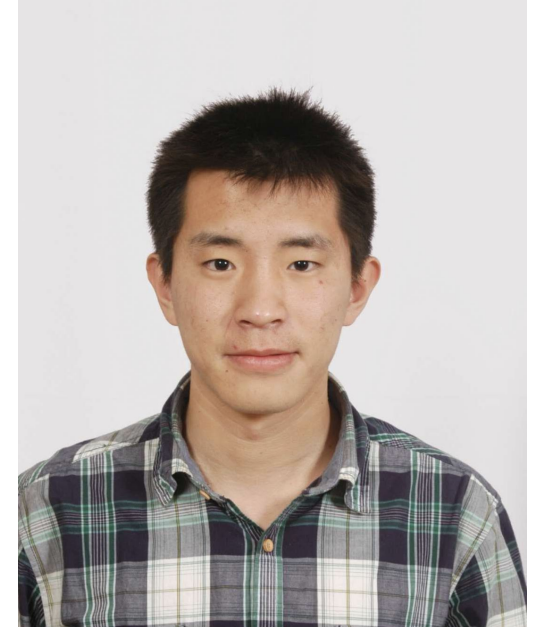
<https://github.com/milvus-io/milvus/blob/master/COMMUNITY.md>

# Milvus Project Leader

## **Xiaofan, Luan**

7 years working experience focused on open source database and big data system design. After graduation from Cornell university with a master degree, Xiaofan works for Oracle, Hedvig and Alibaba cloud. Xiaofan has built several distributed storage and database systems in public cloud.

Xiaofan is currently working at Zilliz as the head of engineering, in charge of R&D of Milvus open source project and the Zilliz cloud platform.



# Collaboration with LF AI & Data projects and members

- Adopted ONNX (graduated project) to manage deep learning models in the Milvus bootcamp applications
- Working with GSI Technology (general member) to support GSI APU processor in Milvus
- Working with Baidu (premier member) in multiple Paddlepaddle and EDL cases
- Initial discussion with Vulkan Kompute (sandbox project) about the integration

The image displays two screenshots of GitHub repositories. The top screenshot shows the 'milvus-io/bootcamp' repository, specifically the file 'milvus\_with\_different\_models.ipynb' in the 'etc/onnx\_and\_milvus' directory. The commit is by 'shiyu22' and is titled 'new-bootcamp'. The bottom screenshot shows the 'PaddlePaddle/PaddleRec' repository, specifically the file 'milvus\_tool' in the 'recserving' directory. The commit is by 'Bennu-Li' and is titled 'Update'. The repository contains several files: 'README.md', 'config.py', 'milvus\_insert.py', and 'milvus\_recall.py', all of which were updated 2 months ago.

# 2021 Roadmap - Completed

- Adopted most popular ANNS algorithm libraries
- Supported near real time data injection
- Implemented the 1st sharding proxy
- Provided SDK for the most popular programming language
- Milvus 1.0 - The first long-term support release

# 2021 Roadmap

- 6.30.2021 Milvus 2.0 - Deliver Distributed Vector database
  - Fully managed Failure Recovery and service discovery
  - Support Vector search with scalar filtering/ point query
  - Support new orm-style apis
  - Tunable consistency
  - Grafana | Prometheus support
- 10.31.2021 Milvus 2.1 - Deliver complete set of DML functions
  - Support deletes/updates
  - Support string/varbinary type
  - Segment Compaction implementation
  - Memory/Load based Balancer
  - K8s deployment support
  - Apache Spark/Flink Connector

# 2022 Roadmap

- 2.28.2022 Milvus 2.2 - Cloud Native Vector database
  - Multi tenant support
  - QueryNode Group isolation
  - BulkLoad insert
  - Traffic Control/ back pressure
  - Distributed Tracing/Log Aggregation
  - Embedded milvus for data scientist
  - Kafka Support

# Long-term Roadmap

- Federated Queries by Spark/Presto
- Time Traverse
- Traffic Control/ back pressure
- Hot/Cold Data separation
- Performance boosting
- Incremental Backup
- Cost based Query Optimization
- On Disk vector Indexing
- Change Data capture
- Data Importer/Transformer

# Executive Summary

- The Milvus project meets all graduation criteria for LF AI & Data graduating projects
- We have been in incubation for 15 months and grew the project to become a strong player in the open source AI ecosystem
- The Milvus project has no requirements for any spending at this time
- We look forward to your positive vote to graduate the project



# TAC Vote on Project Graduation Proposal: Milvus

## **Proposed Resolution:**

The TAC approves the Milvus as an Graduation project of the LFAI & Data Foundation

# Next Steps

The Governing Board will review the Milvus deck during their January call and we will have an agenda item for the board to vote on the graduation.

After a positive vote, LF AI & Data staff will work with Milvus team to announce the graduation and project promotion

# Invited Presentation - KOSA.AI

Layla Li <layla@kosa.ai>

Sonali Sanghrajka <sonali@kosa.ai>



# KOSA

The world's first automated responsible AI system

June 2021

Amsterdam, Netherlands/ Nairobi, Kenya

# OUR TEAM

LAYLA LI



**Co-founder & CEO**

7+ years delivering crucial tech solutions at Tesla, Philips and other tech innovators



SONALI SANGHRAJKA



**Co-founder & CCO**

10+ of strategy and bd experience in market entry of products each worth ~\$500M





**Artificial intelligence to deliver \$13 trillion economic activity by 2030, but 85% will generate false results due to bias in data or algorithm.**

\*McKinsey, Gartner



# This problem costs hidden millions today

## 30% Missed Revenue

from affected customers in the year following a data misuse incident.\*

## Increased Risks

Risks of costly litigations in non-compliant AI systems

\$\$\$\$

## 3x Customer Trust

put on ethics over competency when assessing companies' trustworthiness

## 1 in 6 Digital Talent Loss

because they won't help build potentially harmful products



# 2021 has been a big year in AI Governance

TE

## Uber hit with default 'robo-firing' ruling after another EU labor rights GDPR challenge

≡

The New York Times

⋮

### *Who Is Making Sure the A.I. Machines Aren't Racist?*

When Google forced out two well-known artificial intelligence experts, a long-simmering research controversy burst into the open.



AI NEWS

EU regulation sets fines of €20M or up to 4% of turnover for AI misuse

MIT  
Technology  
Review

## This has just become a big week for AI regulation

The EU has unveiled its new AI rules—but an announcement from the FTC may have more teeth.





# Obstacles for each stakeholder



Leadership



Dev Team

- **Current tool kit:** Consulting services (\$\$\$\$)
- **Problems faced:** Executives don't truly understand AI practices within their organizations

- **Current tool kit:** Open-source libraries such as AIF360 or FairLearn (Lack of guidance and resources to get started)
- **Problem faced:** 36% data science teams cite management resistance as main barrier towards delivering business value\*



# KOSA: Optimizely for fairness-conscious AI/ML teams

---

A system that drives  
organizational change

KOSA answers 4 critical questions:

- What are the biases throughout my AI/ML process?
- How do we mitigate them?
- What can I gain from it?
- Why does my model make decisions the way it does?

## Executive Teams

KOSA's ROI calculator helps quantify financial impact of AI bias.

## Operations Teams

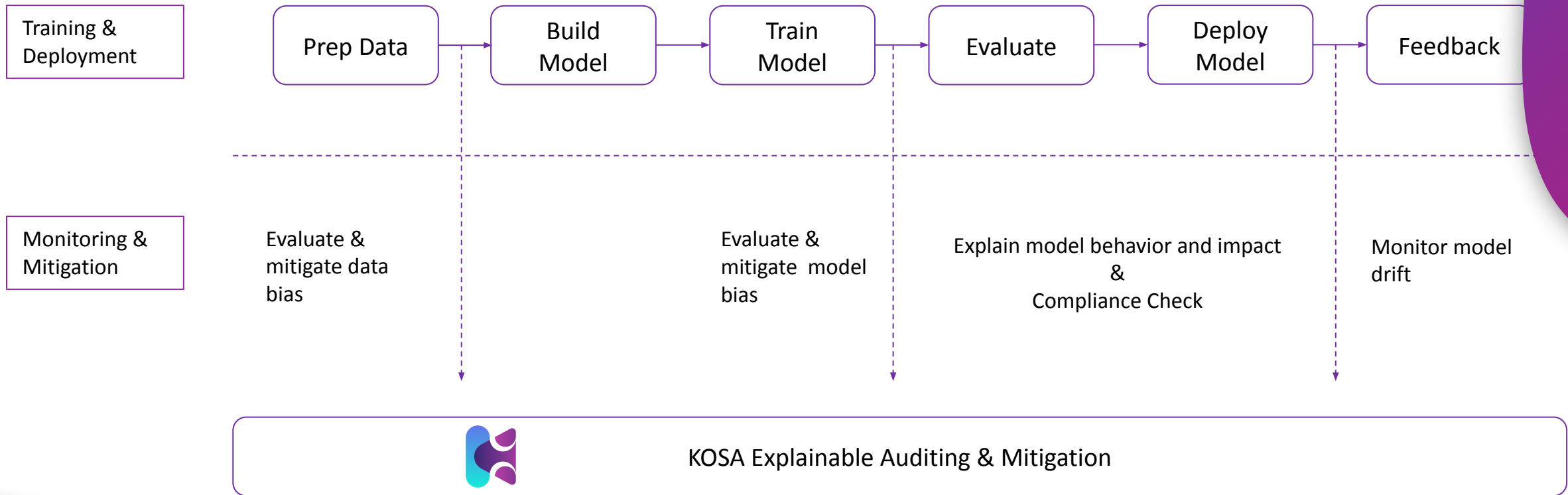
KOSA helps de-risk products from compliance and regulatory needs.

## AI/ML Teams

KOSA provides evaluation & mitigation solutions integrated with the system of choice.



# KOSA IN ML WORKFLOW



(Zero code integration and fully automated)



# KOSA's Open-Source Commitment

## Research Partners



**The International Digital Health & AI  
Research Collaborative**  
Partnership on fair AI definition research



**Know Center: Research Center for  
Data-Driven Business & Big Data Analytics**  
Partnership on ethical AI research

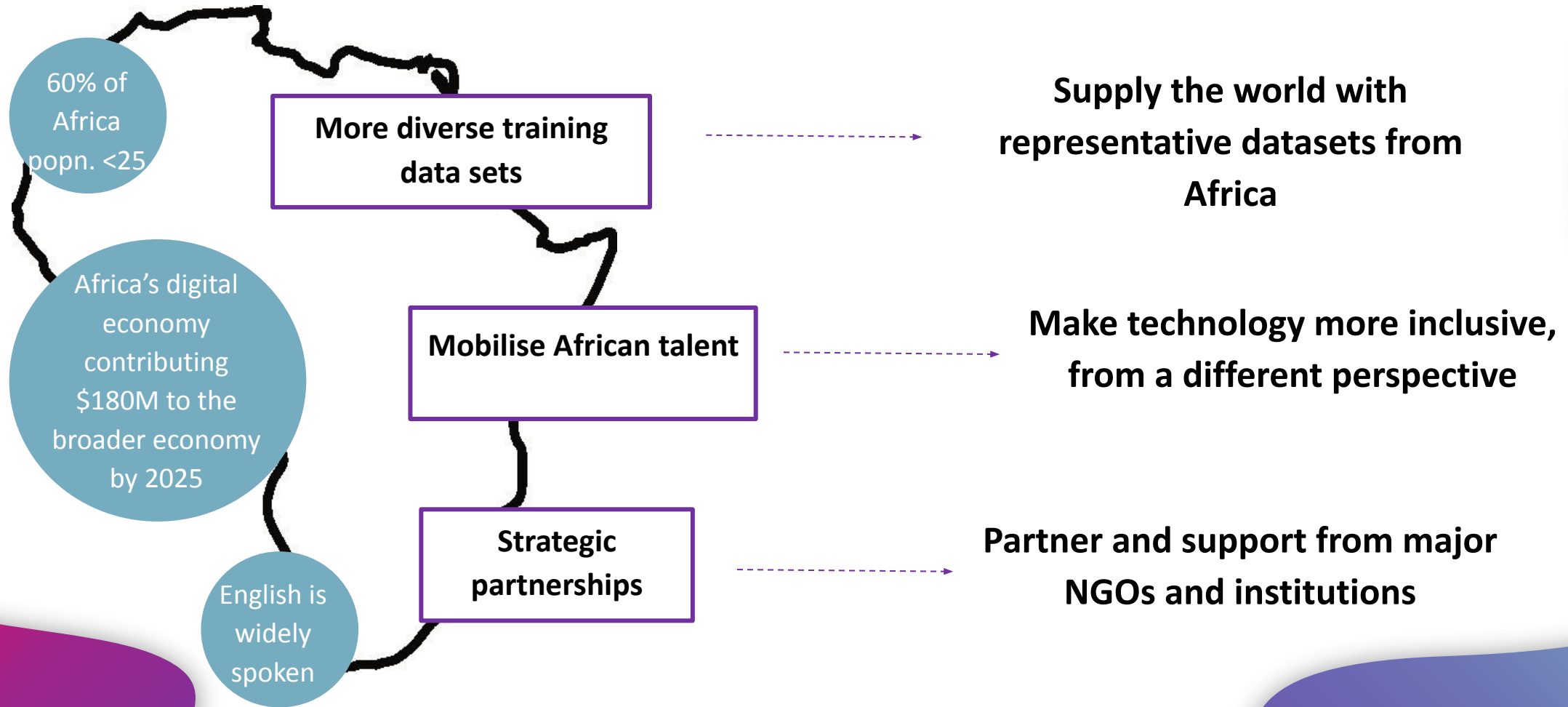


**Africa Health Business**  
Partner on African digital healthcare  
collaboration



**University of Massachusetts Amherst**  
Research collaboration on AI bias mitigation

# KOSA IN AFRICA



# Current Focus and Development

## Pricing Model

- Free: Our open-source research, limited usage and functionalities of our software available
- Pro: \$6250/m based on typical data usage & # of models
- Enterprises: Custom solutions

## Segment Focus

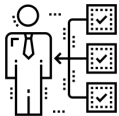
- Healthcare
  - Financial Services
  - Government
- (1000+ employees and tech mature companies in EU, mostly in regulated industries)

## Expansion

- Tech roadmap: from supervised learning to Computer vision to NLP
- Industry expansion: HR, automotives



Healthcare delivery



Credit/Risk assessment



Insurance



Hiring



# THANK YOU!

Contact info:

Layla LI: | E: [layla@kosa.ai](mailto:layla@kosa.ai)  
M: +31 0620314550

Sonali Sanghrajka: | E: [sonali@kosa.ai](mailto:sonali@kosa.ai)  
M: +254 722 518 410

Request a demo at: [KOSA.ai](https://KOSA.ai)

# LF AI & Data - TAC Election

Jacqueline Cardoso (LF)



# Upcoming TAC Chairperson Election

As announced during today's TAC meeting, there will be an opening for the **LF AI & Data Foundation TAC Chairperson**; the role will be effective through 2021. Eligible voting members are listed on the TAC wiki [here](#). **Please review the following election timeline:**

- › *Nomination period start:* May 27th, nominations should include a short bio + statement of intent
- › *Nomination period end:* June 3rd, 5pm ET
- › *Voting period start:* June 7th, LF to send slate of respective nominees and voting instructions to TAC Voting Members
- › *Voting period end:* June 14th, 5pm ET
- › *Election winner announced:* June 15th

Please begin to think about your participation and if you are interested in nominating yourself or another TAC voting member. **Information about the role:**

- › General TAC and Chairperson details can be viewed within the LF AI & Data Charter under section 7 [here](#) - A summary of the role:
- › Represent the TAC as a voting member on the Governing Board (will attend monthly meeting)
- › Lead TAC agenda and meetings with coordination among the TAC representatives and broader community
- › Attend sync meetings with LF AI & Data staff to discuss overall TAC activities and planning
- › General representation of the TAC and the LF AI & Data technical community

# LF AI & Data - Badging for Projects

Andre de Waal (OpenDS4All)  
Mandy Chessell (Egeria)

Goal: Reward contributions!



- **OpenDS4All - Technical Steering Committee badge**
- Be an active member of the TSC for at least 1 year
- **OpenDS4All - Contributor badge**
- Contributing expertise as an instructor to a workshop, bootcamp
- Contributing expertise as an advocant (such as speaker at conferences, workshops, panels)
- Create a module (Powerpoint slide deck)
- Create a module (Jupyter Notebook)
- Create a use case (Markdown file and Jupyter Notebook)
- Translate a module to another language (Spanish, French, German etc.)
- Translate a Jupyter Notebook from Python to R
- Record a module or part of a module
- Substantial edits and/or contributions to existing content

**NB!** The OpenDS4All Technical Steering Committee will evaluate the recipient's contribution and approve/disapprove whether a badge will be issued or not

# LF AI & Data - Annual Project Reviews

Jim Spohrer (TAC), John Mertic (LF)

## Annual Review schedule

Created by John Mertic, last modified by Jim Spohrer just a moment ago

Presentation Template <https://docs.google.com/presentation/d/1CcEf7DyXoobLMtt7Tdbu7f9L6xSUw6seBBokBkGvoqy0/edit?usp=sharing>

Date	Project	Presenter
April 6, 2021	Egeria	Mandy Chessell - <a href="#">recording</a> / <a href="#">deck</a>
April 6, 2021	OpenDS4all	Abdre de Waal - <a href="#">recording</a> / <a href="#">deck</a>
May 20, 2021	ONNX	Jim Spohrer - <a href="#">recording</a> / <a href="#">deck</a>
July 15, 2021	EDL	Ti Zhou (confirmed)
July 29, 2021	Angel	
July 29, 2021	Adlik	
Aug 26, 2021	Acumos	Amit Kumar (confirmed)
Aug 26, 2021	Sparklyr	
Sept 9, 2021	Marquez	
Sept 9, 2021	Milvus	
Sept 23, 2021	NNStreamer	
Sept 23, 2021	ForestFlow	
Oct 7, 2021	Ludwig	
Oct 7, 2021	Amundsen	

Oct 21, 2021	AI Fairness 360
Oct 21, 2021	AI <a href="#">Explainability 360</a>
Oct 21, 2021	Adversarial Robustness Toolbox
Nov 4, 2021	<a href="#">Horovod</a>
Nov 4, 2021	FEAST
Nov 18, 2021	<a href="#">SOAJS</a>
Nov 18, 2021	Delta
Dec 2, 2021	<a href="#">DataPractices.org</a>
Dec 2, 2021	<a href="#">JanusGraph</a>
Dec 16, 2021	<a href="#">Pyro</a>
Jan 6, 2021	<a href="#">Datashim</a>
Jan 6, 2022	<a href="#">Flyte</a>
Jan 20, 2022	<a href="#">RosaeNLG</a>
Jan 20, 2022	<a href="#">SubstraFramework</a>
	<a href="#">MLX</a>
	<a href="#">VulcanKompute</a>

[Schedule:](https://wiki.lfaidata.foundation/display/DL/Annual+Review+schedule) <https://wiki.lfaidata.foundation/display/DL/Annual+Review+schedule>

# LF AI & Data - General Updates

 LF AI & DATA

**Machine Learning**

Framework	Platform	Library
Accord.NET, FedAI, Microsoft LightGBM, MAHOUT, ML.NET, RAY	Angel (Graduated), ForestFlow (Incubating), Ax, cortex, H2O, Kubeflow, mlflow, SeldonIO, SELDON, VIZML VIZML	CartoDB, Flashlight, MediaPipe, mlpack, OpenCV, Sonnet, XGBoost, xLearn

**Deep Learning**

Framework	Platform	Library	Tool
SINGA, Chainer, CNTK, DeepDetect, Databricks AI, dy/net, Alibaba Euler, MindSpore, mxnet, PyTorch, TensorFlow	BigDL, Catalyst, DL4J, jino, Onepool, Polyaxon, TensorFlow	fast.ai, Keras, PyTorch, PyTorch	BoTorch, Intel Distiller, PloyML, PyTorch, tvn

**Reinforcement Learning**

Reinforcement Learning
CartoDB RL, Coach, DeepMind Lab, Dossena, Horizon, OpenAI, Google Puffer, Google SEED RL, Data Science

**Programming**

Programming
Pyro (Graduated), DASK, IPYNB, Julia, MARS, Numba, NumPy, NYOKA, primegrate, PyMC3, python, R, SciPy, Stan

**Data**

Education	Lineage	Relational DB	Store & Format	Versioning	Operations	Feature Engineering	Stream Processing	SQL Engine	Visualization	Pipeline Management	Labeling & Annotation	Governance
DATAPRACTICES.ORG (Incubating), OpenDS4All (Incubating)	OpenLineage (Sandbox)	CouchDB, MySQL	JanusGraph, Milvus (Incubating), ARROW, CEPH, DELTA LAKE, druid, HUDI, influxdb, pandas, Parquet, pilosa, YEARCH, vespa	ANCESTRAL STREAM, DVC, PySystem	Amundsen, datashim, MARQUEZ (Incubating), FEAST (Incubating)	NNStreamer (Incubating), kafka, logstash, PRAVEG, PREFERCT, samza, Uber uReplicator	APACHE DRILL, HAWQ, presto, SQLFlow, trino	bokeh, IBM Data Science, plotly Dash, Uber deck.gl, Ecco!, Google Facets, Grafana, Metabase, RCloud, reDash, NYU, ODSUPERT, Google Searchbar	Analytics Zoo, DAGSTER, TENDON	CVAT, Doccano, Labelbox, HITACHI, Labelimg, Microsoft WIT	EGERIA (Graduated)	

**Model**

Inference	Benchmarking	Training	Parameter	Format & Interface	Marketplace	Workflow	Tool
ADLIX (Incubating), KFServing, MNN, NVIDIA, NVIDIA, uTensor	DAWNbench, MLPerf	REBAR, LEWIG (Incubating), Microsoft, Pelanform	EFFICIENT, COIL	ONNX (Graduated), Uber Neuropod	Acumos (Graduated), Machine Learning exchange (Sandbox), IBM	Flyte (Incubating), Airflow, nifi, argo, Azkaban, Cadenza, Cooler, DataBolt, Spotify, mLeap, PREFERCT, TRAINS	Qualcomm AIMET, FACEBOOK dlm, Microsoft Wagon, AWS MMS, MDB, Amazon Neo-AI, NETRIN, GNN, PipelineAI, studio.ml, Google TensorFlow, AWS Turi

**Trusted & Responsible AI**

Explainability	Adversarial	Bias & Fairness
AI Explainability 360 (Incubating), Microsoft, WASHINGTON STATE UNIVERSITY Lime, Google Lucid, SHAP, SKATER, Treetrainer	ALIBI, ELIS, AdyBox, qcover torch	AI Fairness 360 (Incubating), Aequitas, Audit AI, Fairlearn

**Distributed Computing**

Computing & Management	Interface
EDL (Incubating), SOAJS (Incubating), Bahir, MESOS, Apache Ranger, SPARK, STORM, GNS3, NETFLIX genie, kubernetes, Nauta, OPENSHIFT, Singularity	sparklyr (Incubating), TORRE, LVVY

The LF AI & Data landscape explores open source projects in Artificial Intelligence and Data and their sub-domains.

[l.faidata.foundation](https://l.faidata.foundation)

**DLFAI & DATA Landscape**

**Security & Privacy**



Google, IBM HElib, Microsoft SEAL, SUBSTRA, Google, TRUST
-----------------------------------------------------------












**Natural Language Processing**











Natural Language Processing
DELTA (Incubating), RosaeNLG (Sandbox), Google ALBERT, AllenNLP, Google Bert, fastText, flair, RASA, spaCy, Facebook Lstm, Facebook XLNet, YouTokenToMe

**Notebook Environment**

Apache Zeppelin, BeakerX, colab, Elyra, IP[y]: IPython, jupyter, IBM, Polynote, Streamlit
-------------------------------------------------------------------------------------------

Machine Learning	Framework	Platform	Library	Framework	Platform	Library	Tool	Reinforcement Learning	Programming
		 Graduated  Incubating		Deep Learning				Reinforcement Learning	 Graduated

Data	Education	Lineage	Relational DB	Store & Format	Versioning	Operations	Feature Engineering	Stream Processing	SQL Engine	Visualization	Pipeline Management	Labeling & Annotation	Governance
	 Incubating  Incubating	 Sandbox	 Incubating  Incubating		 Incubating  Incubating  Incubating		 Incubating	 Incubating					 Graduated

Model	Inference	Benchmarking	Training	Parameter	Format & Interface	Marketplace	Workflow	Tool	Explainability	Adversarial	Bias & Fairness
	 Incubating		 Graduated  Incubating		 Graduated	 Graduated  Sandbox	 Incubating		 Incubating	 Incubating	 Incubating

Distributed Computing	Computing & Management	Interface	Security & Privacy	Natural Language Processing	Notebook Environment
	 Incubating  Incubating	 Incubating	 <p>The LF AI &amp; Data landscape explores open source projects in Artificial Intelligence and Data and their respective sub-domains.</p> 	Security & Privacy	 Incubating  Sandbox



## 2020 TAC Meetings Summary

Jan Feb Mar	16: Milvus (Zilliz)*	13: <i>MLOps Work (LF CD)</i>  27: <i>Collective Knowledge (Coral Reef)</i>	12: NNStreamer (Samsung)*  26: ForestFlow (?)*
Apr May Jun	9: <i>Trusted AI &amp; ML Workflow (LF)</i>  23: <i>Open Data Hub (Red Hat)</i>	7: Ludwig (Uber)*  21: <i>SnapML (IBM)</i>	4: <i>Trusted AI (AI for Good, Ambianic.ai, MAIEI)</i>  18: Fairness, Explainability, Robustness (IBM)*
Jul Aug Sep	16: <i>Mindspore (Huawei)</i>  30: Amundsen (Lyft)*	16: <i>Delta (Didi)</i> <b>16: Horovod (Uber/LF)**</b>  30: <i>ModelDB (?)</i> 30: <i>Egeria, OpenDS4All, BI&amp;AI (LF ODPI)</i>	10: SOAJS (HeronTech)* 10: Delta (Didi)* 24: FEAST (Gojek)* <b>24: Egeria, (LF ODPI)**</b> 24: OpenDS4All (ODPI)* 24: BI&AI Committee (ODPI)
Oct Nov Dec	8: <i>Fairness, Explainability, Robustness (LF)</i>  22: <i>OpenLineage (DataKins)</i> 22: <i>IDA (IBM/Salesforce)</i>	5: DataPractices.Org (WorldData/LF)* 5: <i>Kubeflow-On-Prem (Google, Arrikto/Intel)</i>  19: <i>OpenDS4All, DataPractices.Org, edX Ethical AI (LF)</i>	3: TBD - JanusGraph (LF)* 3: <i>TBD - RosaeGL (?)</i>  17: TBD – Seldon Core (Seldon)* <b>17: TBD – Pyro (Uber/LF)**</b>

(Entity)\* = incubating vote

**\*\* bold = graduate vote**

*Italics = invited project presentation*

## 2021 TAC Meetings Pipeline Summary

Jan Feb Mar	14: Datashim(IBM)* 28: Project Lifecycle Stages vote Invited talks Sedna & CIM	11: Invited talks Egeria CI & Mentorships 25: Flyte (Lyft) *	11: RosaeNLG ( ) Sandbox Proposal Invited talk Elyra-AI (IBM) 25: Substra Framework (Substra)*
Apr May Jun	8: Invited talk JINAAI 22: Egeria & OpenDS4All - project update	6: ML eXchange (MLX) (IBM) Sandbox Vulcan Kompute ( ) Sandbox 20: OpenLineage (Datakin) Sandbox	3: KOSA.ai 17: GSI
Jul Aug Sep	1: Canceled for holiday 15: TonY (Linkedin)	5: TBD - Project updates 19: TBD - Project updates	?: Open Data Hub (Red Hat) ? Ray (Anyscale.io) ?: Pachyderm (Pachyderm) ?: DataHub (LinkedIn) ?: Kubeflow-On-Prem (Google, Arrikto, Intel)
Oct Nov Dec	?: Vespa (Verizon Media) ?: KubeflowServing (Google, Arrikto, Seldon) ?: Kubeflow Pipeline (Google, Bloomberg) ?: Common Knowledge (Code Reef) ?: Couler (Ant Financial)	?: Snorkle (Snorkle) ?: Plotly (DASH) ?: Melody (Substra) ?: mloperator (Polyaxen) ?: SnapML (IBM)	?: PMML/PFA (DMG.org) ?: Mindspore, Volcano (Huawei) ?: TransmorgrifAI (Salesforce) ?: AIMET (Qualcomm) ?: Elyra-AI (IBM)

(Entity)\* = incubating vote

\*\* **bold** = graduate vote

*Italics* = invited project presentation

Getting to know the projects more

# Active and growing developer community

Cumm. Jan 1- Dec 31, 2020 vs. Jan 1, 2020 to Apr 2, 2021

**8.92K**

Contributors

+7.25%

**34.65K**

PRs/Changesets

+7.21%

**100.03K**

Commits

+11.03%

**22.54K**

Total issues

+5.19%

**358**

Repositories

+3.07%

**2.08K**

Slack messages

+68.75%

**9.61K**

Contributors

**37.15K**

PRs/Changesets

**111.07K**

Commits

**23.71K**

Total issues

**369**

Repositories







**3.51K**

Slack messages





















# Projects

<https://landscape.lfai.foundation/card-mode?project=company>




Graduated LF AI & Data Projects (6)

 Acumos LF AI & Data Foundation ★ 16	 Angel-ML LF AI & Data Foundation ★ 6,247	 Egeria LF AI & Data Foundation ★ 509	 Horovod LF AI & Data Foundation ★ 11,246	 ONNX LF AI & Data Foundation ★ 10,681	 Pyro LF AI & Data Foundation ★ 6,939
-----------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------

Incubating LF AI & Data Projects (20)

 ADLIK LF AI & Data Foundation ★ 256	 Adversarial Robustness Toolbox LF AI & Data Foundation ★ 2,250	 AI Explainability 360 LF AI & Data Foundation ★ 840	 AI Fairness 360 LF AI & Data Foundation ★ 1,397	 Amundsen LF AI & Data Foundation ★ 2,164	 DataPractices.ORG LF AI & Data Foundation ★ 10	 datashim LF AI & Data Foundation ★ 101	 DELTA LF AI & Data Foundation ★ 1,413	 EDL Elastic Deep Learning LF AI & Data Foundation ★ 132	 Feast LF AI & Data Foundation ★ 1,881
 Flyte LF AI & Data Foundation ★ 1,421	 ForestFlow LF AI & Data Foundation ★ 46	 JanusGraph LF AI & Data Foundation ★ 3,972	 Ludwig LF AI & Data Foundation ★ 7,678	 Marquez LF AI & Data Foundation ★ 630	 Milvus LF AI & Data Foundation ★ 6,260	 NNStreamer LF AI & Data Foundation ★ 362	 OpenDS4All LF AI & Data Foundation ★ 310	 SOAJS LF AI & Data Foundation ★ 46	 sparklyr LF AI & Data Foundation ★ 785

Sandbox LF AI & Data Projects (3)

 Machine Learning eXchange LF AI & Data Foundation ★ 11	 OpenLineage LF AI & Data Foundation ★ 272	 RosaeNLG LF AI & Data Foundation ★ 20
--------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------

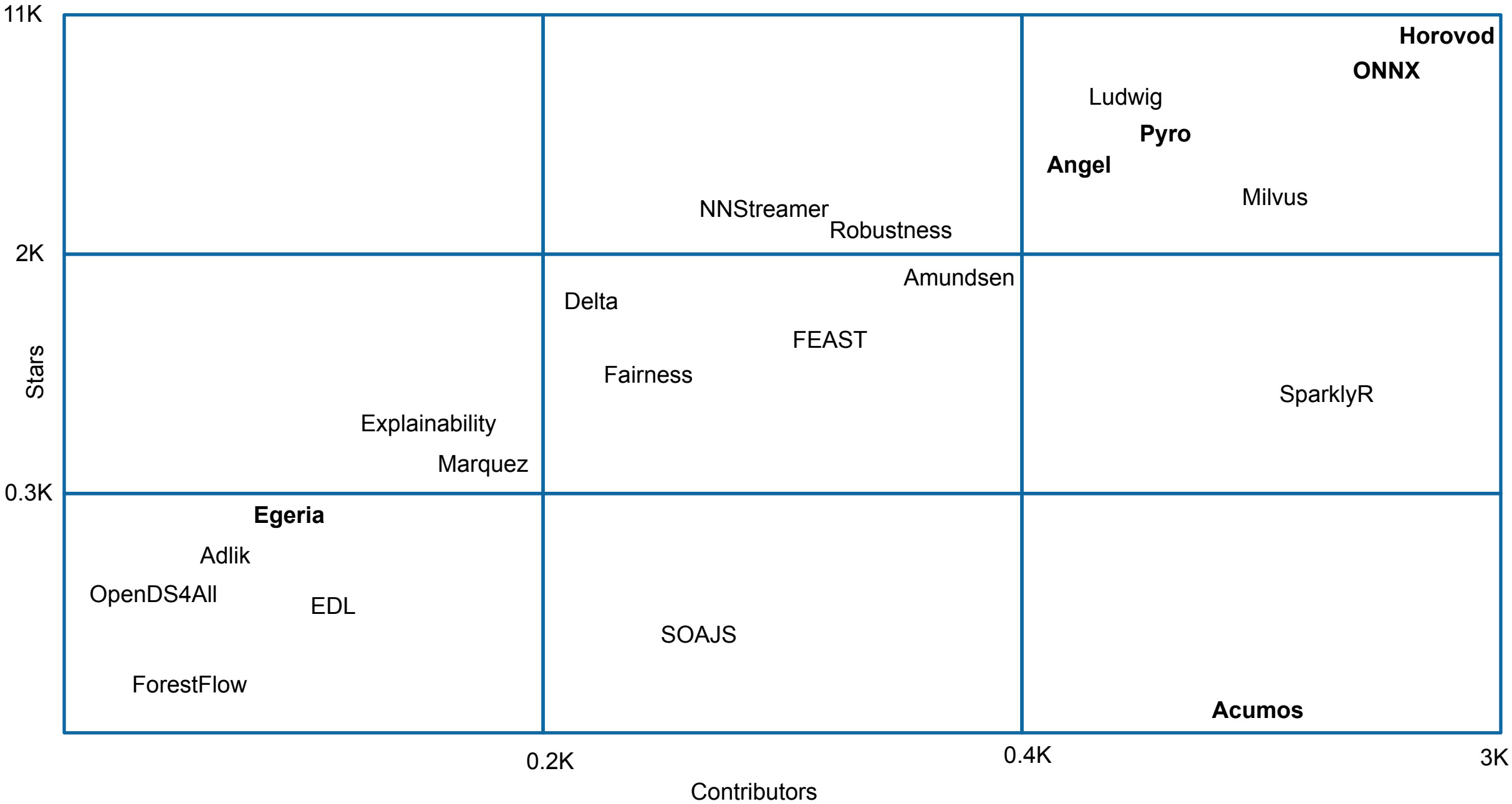
\* Missing Substra Framework (pending logo redesign)

# New projects in 2021

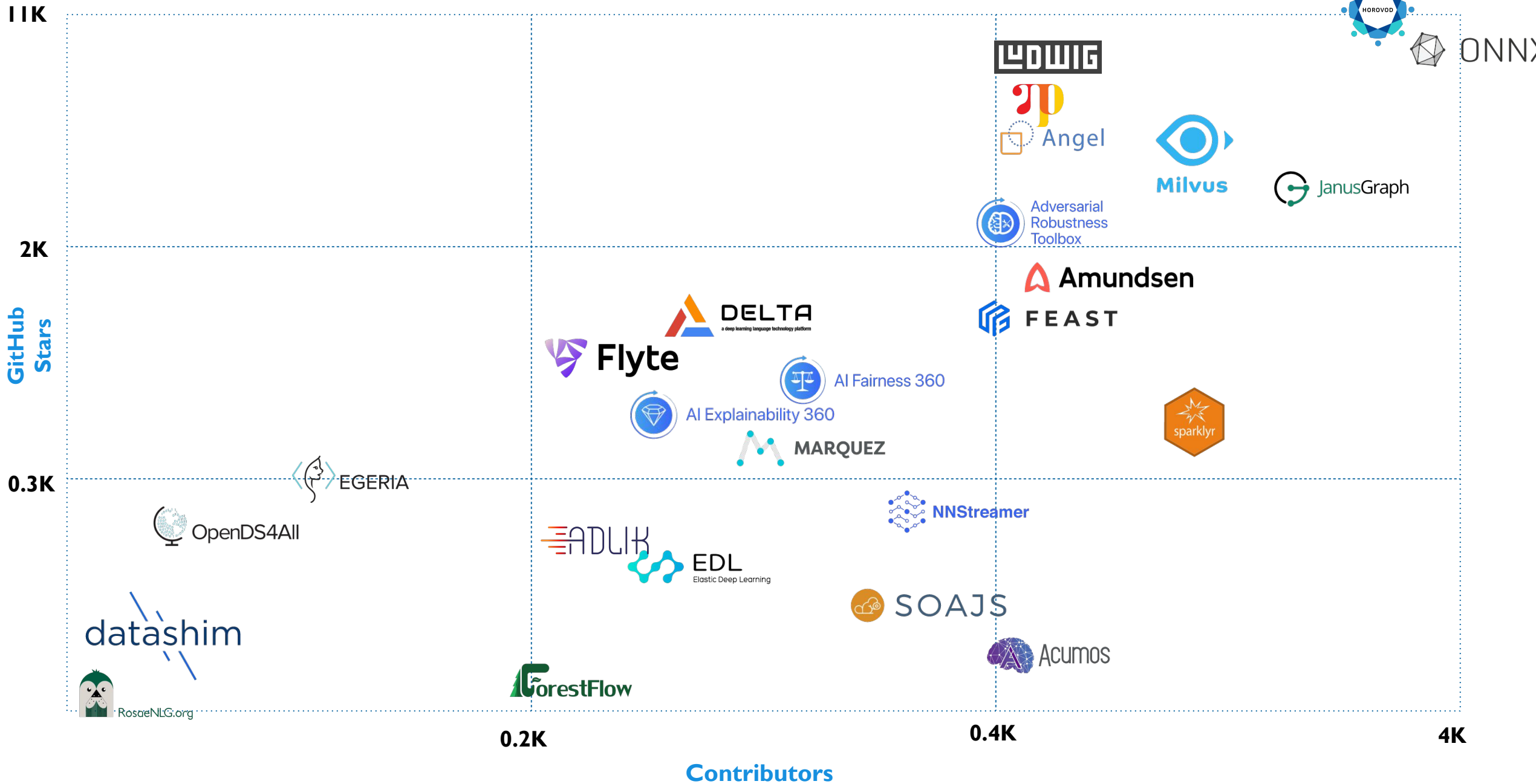
→ Track incoming proposals via: <https://github.com/lfai/proposing-projects>

1. **Datashim:** Open source enablement and acceleration of data access for Kubernetes/OpenShift workloads in a transparent and declarative way
2. **Flyte:** Production-grade, declarative, structured and highly scalable cloud-native workflow orchestration platform
3. **RosaeNLG:** Open source project, template-based Natural Language Generation (NLG) automating the production of relatively repetitive texts based on structured input data and textual templates, run by a NLG engine
4. **Substra Framework:** Low-layer framework, offering secure, traceable, distributed orchestration of machine learning tasks among partners.
5. **ML eXchange:** Data and AI Assets Catalog and Execution Engine
6. **Kompute: Blazing fast, mobile-enabled, asynchronous, and optimized for advanced GPU processing usecases.**
7. **Open Lineage:** Open standard for metadata and lineage collection designed to instrument jobs as they are running

Data from November 23, 2020 – Stars and Contributors

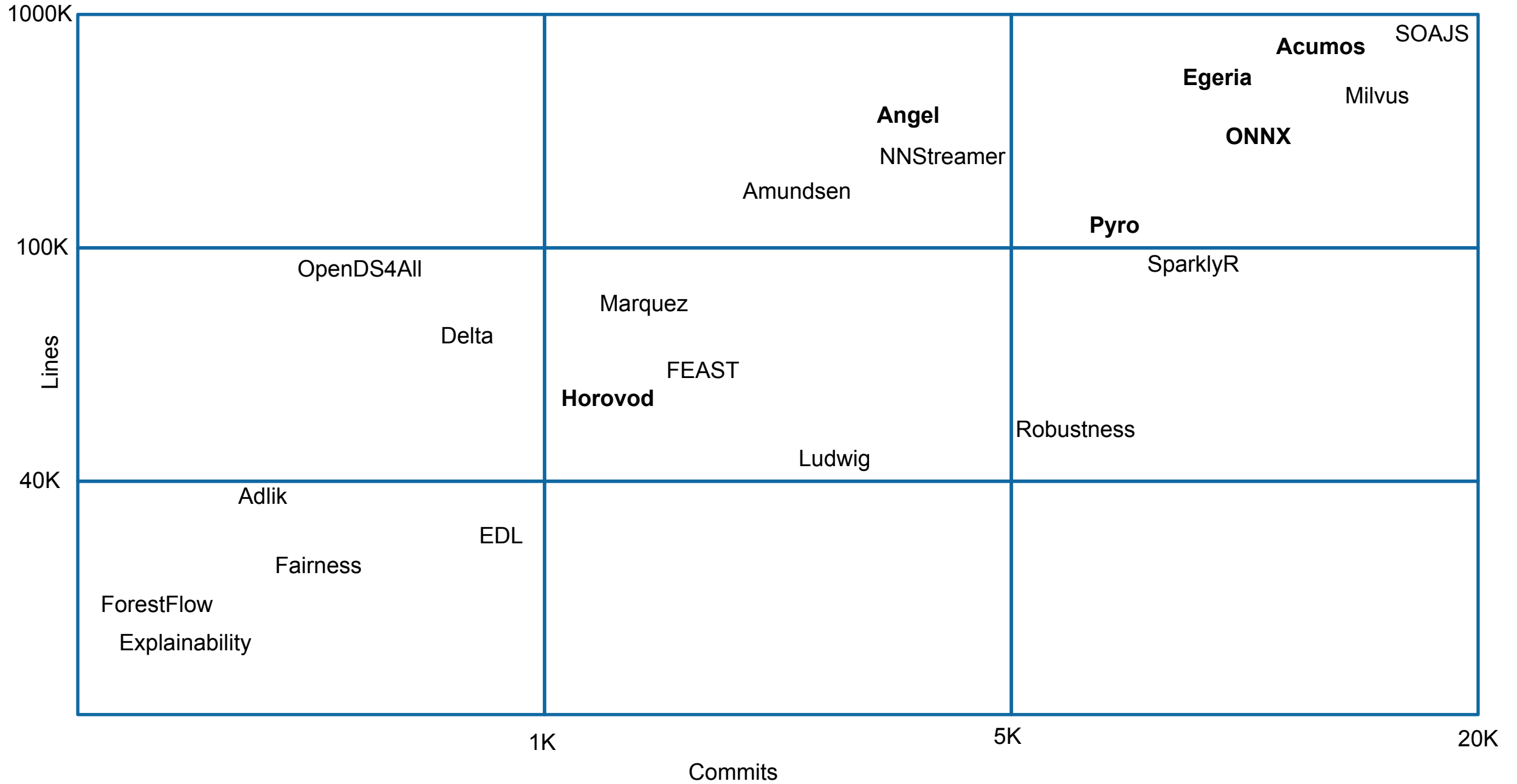


# Data Coverage (Mar 19, 2021) - Stars and Contributors





Data from November 23, 2020 – Lines of Code and Commits



# Data Coverage (Mar 19, 2021) - Line of Code and Commits



# Looking to host a project with LF AI & Data

- › Hosted project stages and life cycle:

<https://lfaidata.foundation/project-stages-and-lifecycle/>

- › Offered services for hosted projects:

<https://lfaidata.foundation/services-for-projects/>

- › Contact:

Jim Spohrer (TAC Chair) and Ibrahim Haddad (ED, LF AI & Data)

# Promoting Upcoming Project Releases

We promote project releases via a blog post and on LF AI & Data [Twitter](#) and/or [LinkedIn](#) social channels

For links to details on upcoming releases for LF AI & Data hosted projects visit the [Technical Project Releases wiki](#)

If you are an LF AI & Data hosted project and would like LF AI & Data to promote your release, reach out to [pr@lfai.foundation](mailto:pr@lfai.foundation) to coordinate in advance (min 2 wks) of your expected release date.

# Note on quorum

As LF AI & Data is growing, we now have 18 voting members on the TAC.

TAC representative - please ensure you attend the bi-weekly calls or email Jacqueline/Ibrahim to designate an alternate representative when you can not make it.

We need to ensure quorum on the calls especially when we have items to vote on.

# Updates from Outreach Committee

# Upcoming Events

- › Upcoming Events
  - › Visit the [LF AI & Data Events Calendar](#) or the [LF AI & Data 2021 Events wiki](#) for a list of all events
  - › To participate visit the [LF AI & Data 2021 Events wiki page](#) or email [info@lfaidata.foundation](mailto:info@lfaidata.foundation)
  
- › Please consider holding virtual events

To discuss participation, please email [events@lfaidata.foundation](mailto:events@lfaidata.foundation)

# Upcoming Events

<https://lfaidata.foundation/events/>

- **OSS Global (hybrid) - Seattle, WA, USA - Sept 27-30 - [Event Website](#)**
  - a. **Mini-Summit, Booth, Track**



# LF AI PR/Comms

- › Please follow LF AI & Data on [Twitter](#) & [LinkedIn](#) and help amplify news via your social networks - Please retweet and share!
  - › Also watch for news updates via the tac-general mail list
  - › View recent announcement on the [LF AI & Data Blog](#)
- › Open call to publish project/committee updates or other relevant content on the [LF AI & Data Blog](#)
- › To discuss more details on participation or upcoming announcements, please email [pr@lfaidata.foundation](mailto:pr@lfaidata.foundation)

# Call to Participate in Ongoing Efforts

 OLF AI & DATA

# Trusted AI

- › **Leadership:**  
Animesh Singh (IBM), Souad Ouali (Orange), and Jeff Cao (Tencent)
- › **Goal:** Create policies, guidelines, tooling and use cases by industry
- › **Slack conversation channel:**  
#trusted-ai-committee  
<https://lfaifoundation.slack.com/archives/CPS6Q1E8G>
- › **Github:**  
<https://github.com/lfai/trusted-ai>
- › **Wiki:**  
<https://wiki.lfai.foundation/display/DL/Trusted+AI+Committee>
- › **Email lists:**  
<https://lists.lfaidata.foundation/g/trustedai-committee/>
- › **Next call:** Monthly alternating times  
<https://wiki.lfai.foundation/pages/viewpage.action?pageId=12091895>

# ML Workflow & Interop

- › **Leadership:**  
Huang “Howard” Zhipeng (Huawei)
- › **Goal:**  
Define an ML Workflow and promote cross project integration
- › **Slack conversation channel:**  
#ml-workflow  
<https://lfaifoundation.slack.com/archives/C011V9VSMQR>
- › **Wiki:**  
<https://wiki.lfaidata.foundation/pages/viewpage.action?pageId=10518537>
- › **Email lists:**  
<https://lists.lfaidata.foundation/g/mlworkflow-committee>
- › **Next call:** Monthly check calendar/slack  
<https://wiki.lfai.foundation/pages/viewpage.action?pageId=18481242>

# BI & AI

- › **Leadership:**  
Cupid Chan (Index Analytics)
- › **Goal:** Identify and share industry best practices that combine the speed of machine learning with human insights to create a new business intelligence and better strategic direction for your organization.
  
- › **Slack conversations channel:**  
**#bi-ai-committee**  
<https://lfaifoundation.slack.com/archives/C01EK5ND073>
- › **Github:**  
<https://github.com/odpi/bi-ai>
- Wiki:**  
<https://wiki.lfaidata.foundation/pages/viewpage.action?pageId=35160417>
- Email lists:**  
<https://lists.lfaidata.foundation/g/biai-discussion>
- Next call:** Monthly community call TBD

# Ongoing effort to create AI Ethics Training

Initial developed course by the LF: Ethics in AI and Big Data - published on edX platform:

<https://www.edx.org/course/ethics-in-ai-and-big-data>

The goal is to build 2 more modules and package all 3 as a professional certificate - a requirement for edX

- › **To participate:**  
<https://lists.lfaidata.foundation/g/aiethics-training>

# Upcoming TAC Meetings

# Upcoming TAC Meetings (Tentative)

- › Jun 3: Milvus Graduation Proposal (TBD), KOSA.ai
- › June 17: GSI invited presentation, Saishruthi Swaminathan (AIOps Committee proposal) (TBD)
- › July 1: Canceled Holiday
- › July 15: TonY (LinkedIn)
- › Aug 5: TBD - Annual project review
- › Aug 19: TBD - Annual project review
- ›

Please send agenda topic requests to [tac-general@lists.lfaidata.foundation](mailto:tac-general@lists.lfaidata.foundation)

# TAC Meeting Details

- › To subscribe to the TAC Group Calendar, visit the wiki: <https://wiki.lfaidata.foundation/x/cQB2>
- › Join from PC, Mac, Linux, iOS or Android: <https://zoom.us/j/430697670>
- › Or iPhone one-tap:
  - › US: +16465588656,,430697670# or +16699006833,,430697670#
- › Or Telephone:
  - › Dial(for higher quality, dial a number based on your current location):
  - › US: +1 646 558 8656 or +1 669 900 6833 or +1 855 880 1246 (Toll Free) or +1 877 369 0926 (Toll Free)
- › Meeting ID: 430 697 670
- › International numbers available: <https://zoom.us/u/achYtcw7uN>



# Open Discussion

# Mission

To build and support an open community and a growing ecosystem of open source AI, data and analytics projects, by accelerating innovation, enabling collaboration and the creation of new opportunities for all the members of the community

# Legal Notice

- › The Linux Foundation, The Linux Foundation logos, and other marks that may be used herein are owned by The Linux Foundation or its affiliated entities, and are subject to The Linux Foundation's Trademark Usage Policy at <https://www.linuxfoundation.org/trademark-usage>, as may be modified from time to time.
- › Linux is a registered trademark of Linus Torvalds. Please see the Linux Mark Institute's trademark usage page at <https://lmi.linuxfoundation.org> for details regarding use of this trademark.
- › Some marks that may be used herein are owned by projects operating as separately incorporated entities managed by The Linux Foundation, and have their own trademarks, policies and usage guidelines.
- › TWITTER, TWEET, RETWEET and the Twitter logo are trademarks of Twitter, Inc. or its affiliates.
- › Facebook and the “f” logo are trademarks of Facebook or its affiliates.
- › LinkedIn, the LinkedIn logo, the IN logo and InMail are registered trademarks or trademarks of LinkedIn Corporation and its affiliates in the United States and/or other countries.
- › YouTube and the YouTube icon are trademarks of YouTube or its affiliates.
- › All other trademarks are the property of their respective owners. Use of such marks herein does not represent affiliation with or authorization, sponsorship or approval by such owners unless otherwise expressly specified.
- › The Linux Foundation is subject to other policies, including without limitation its Privacy Policy at <https://www.linuxfoundation.org/privacy> and its Antitrust Policy at <https://www.linuxfoundation.org/antitrust-policy>, each as may be modified from time to time. More information about The Linux Foundation's policies is available at <https://www.linuxfoundation.org>.
- › Please email [legal@linuxfoundation.org](mailto:legal@linuxfoundation.org) with any questions about The Linux Foundation's policies or the notices set forth on this slide.