

# Meeting of the Technical Advisory Council (TAC)

April 22, 2021

 **DLF** AI & DATA

# Anti-Trust 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>
- ›
- › 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 (5 mins)
- › Approval of Minutes from April 8th (5 mins)
- › Welcome new Associate Member “Banque de France”
  
- › Annual Project Reviews (40 minutes)
  - › Egeria (Mandy Chessell)
  - › OpenDS4All (Andre de Waal)
  
- › CI Working Group - project CI survey insight (5 minutes)
- › LF AI General Updates (2 minutes)
- › Open Discussion (3 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 Soinenen	<a href="mailto:jonne.soininen@nokia.com">jonne.soininen@nokia.com</a>
SAS	Nancy Rausch	<a href="mailto:nancy.rausch@sas.com">nancy.rausch@sas.com</a>
Tech Mahindra	Nikunj Nirmal	<a href="mailto:nn006444@techmahindra.com">nn006444@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	Nat Subramanian	<a href="mailto:natarajan.subramanian@techmahindra.com">natarajan.subramanian@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>

# Approval of April 8th, 2021 Minutes

Draft minutes from the April 8<sup>th</sup> TAC call were previously distributed to the TAC members via the mailing list

## **Proposed Resolution:**

- › That the minutes of the April 8<sup>th</sup> meeting of the Technical Advisory Council of the LF AI & Data Foundation are hereby approved.

# Welcome to new associate member - Banque de France

Associate





# Annual Project Review - Egeria

Mandy Chessell <mandy\_chessell@uk.ibm.com>

# Annual Review for Egeria

April/22/2021

 **DLF** AI & DATA



# Egeria

## Brief Description:

Egeria is the world's first open source metadata standard. It provides open APIs, event formats, types and integration logic so organizations can share data management and governance across the entire enterprise without reformatting or restricting the data to a single format, platform, or vendor product.

TSC Chairperson: [Mandy Chessell](#)

## Contributed by:

Initially contributed to ODPI by IBM and ING Bank in August 2018, and added as a Graduate Project in November 2020



## Key Links:

Github: <https://github.com/odpi/egeria>

Website: <https://egeria-project.org>

Artwork:

<https://artwork.lfaidata.foundation/projects/egeria/>

Mailing lists:

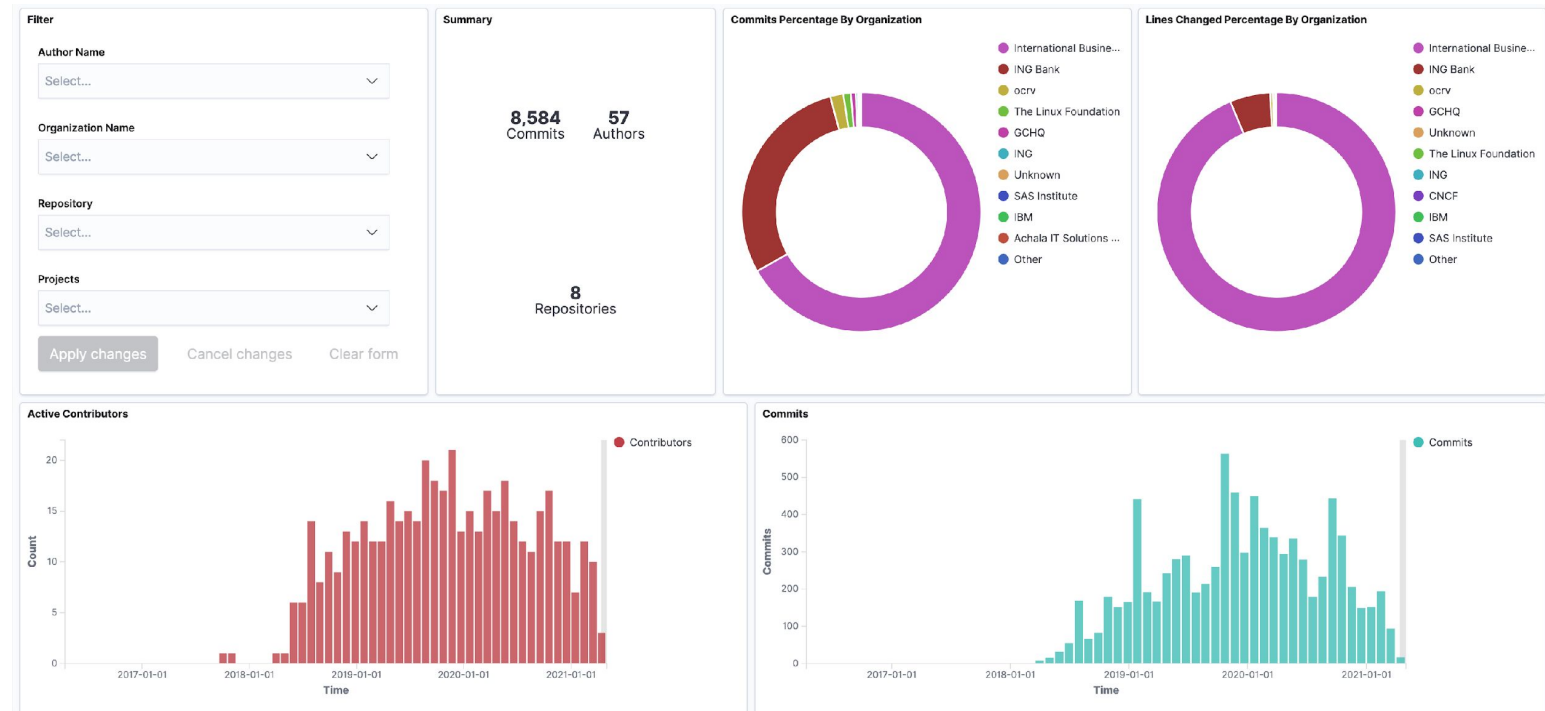
- > [egeria-announce](#)
- > [egeria-technical-discuss](#)
- > [egeria-tsc](#)





# Organizations contributing

- > IBM
- > ING
- > OCRV (Russian Railways)
- > SAS
- > GCHQ



# Contributions



**16.87M**  
Lines Of Code Changed

**8.56K**  
Commits

**57**  
Contributors

**1**  
No Of Sub Projects

**8**  
Repositories

## Top 10 Contributors By Commits

[View All](#)

NAME	LINES OF CODE	COMMITTS	%
Mandy Chessell	1.70M	2.04K	10.08%
Christopher Grote	4.59M	1.31K	27.23%
Nigel Jones	230.12K	1.13K	1.36%
Graham Wallis	281.97K	584	1.67%
David Radley	8.95M	500	53.02%
Guy Ackermans	189.83K	463	1.13%
Daniela Otelea	76.76K	335	0.45%
Raluca Popa	60.62K	263	0.36%
Ilya Ksenofontov	63.32K	151	0.38%

## Top 10 Organizations By Commits

[View All](#)

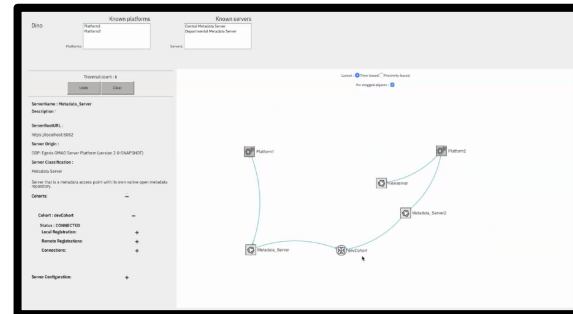
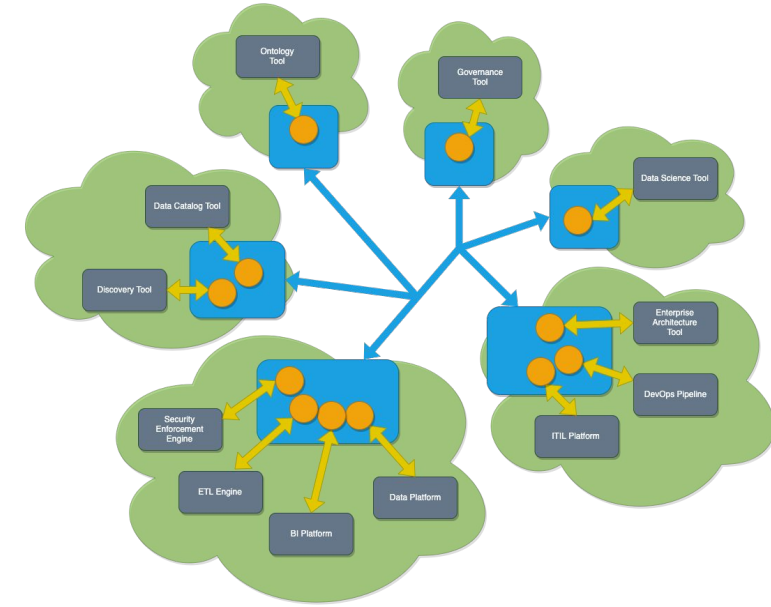
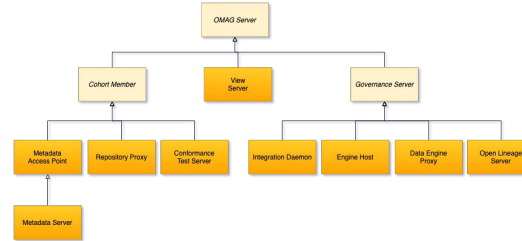


- International Business Machines Corporation (IBM)
- ING Bank
- ocrv
- The Linux Foundation
- GCHQ
- ING
- Unknown
- SAS Institute
- IBM
- Others

Projects	Commits	Authors	Repositories	Added Lines	Removed Lines	Lines Changed	Avg. Time To Commit (Hrs)
EGERIA	8,584	57	8	9,048,818	7,823,500	16,872,318	0.974

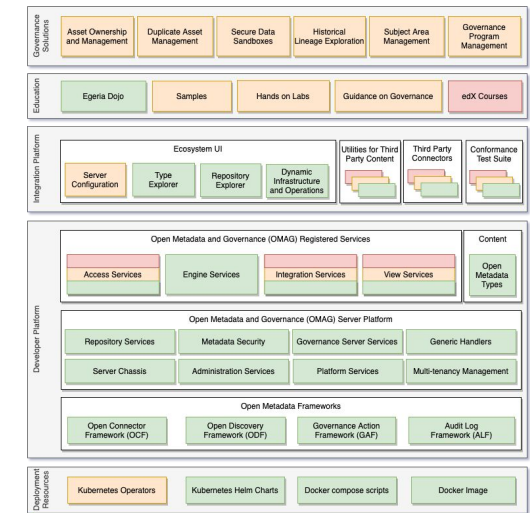
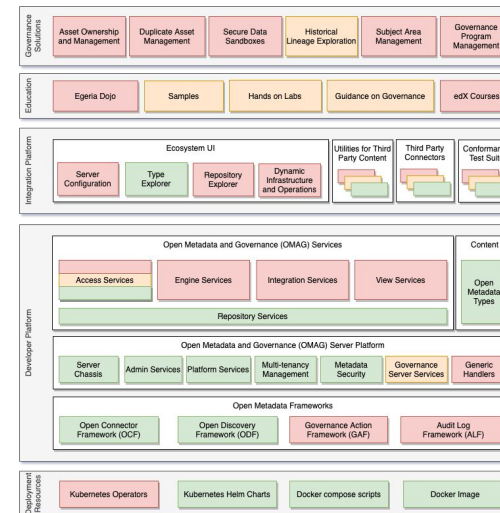
# Key Achievements in the past year

- Monthly releases: 1.5-2.8
- Migration of CI/CD pipeline to GitHub Actions
- Use of TLS by platform
- Consolidation of Server Types
  - Integration Daemon
  - Engine Host
- ING in production with open lineage
- Graph based ecosystem UI
- Metadata Governance
- Historic metadata repository
- 3 day Dojo education



April 2020

April 2021



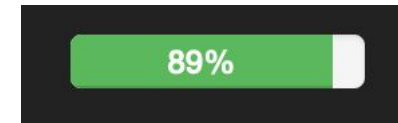
# Areas the project could use help on

- › CII Badge
- › Hosting of Javadoc and Swagger (output of build)
- › Hosting of complex test environments

# CII Badge Progress

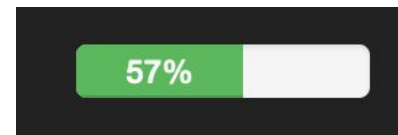
## > Silver

- > Hyperlink to all achievements within 48 hours of public recognition
- > Use of accessibility best practices
- > TLS verification by default
- > TLS certification before sending private information
- > Use of hardening mechanism for software defects
- > Security assurance case



## > Gold

- > Small tasks for beginners
- > 2FA enabled (with cryptography)
- > Code review requirements
- > Code reviews for 50% of changes
- > Project websites secured against MITM attacks





# Feedback on working with LF AI & Data

- › Provision of confluence WIKI is very helpful
- › Value opportunities to work with related projects
  - › Marquez / Egeria  Open Lineage

# TAC Open Discussion

# Annual Project Review - OpenDS4All

Andre de Waal <[andre.dewaal@ibm.com](mailto:andre.dewaal@ibm.com)>

# Annual Review for OpenDS4All

04/22/2021

 **OLFAI & DATA**



# OpenDS4All

## Brief Description:

OpenDS4All is an open source project built to accelerate the creation of data science curricula at academic institutions.

TSC Chairperson: [André de Waal](#)

## Contributed by:

Initially contributed to ODPi by IBM and the University of Pennsylvania in August 2019, and added as an Incubation Project in November 2020



# OpenDS4All

## Key Links:

Github: <https://github.com/odpi/opends4all>

Artwork:

<https://artwork.lfaidata.foundation/projects/opends4all/>

## Mailing lists:

- > [opends4all-announce](#)
- > [opends4all-technical-discuss](#)
- > [opends4all-tsc](#)



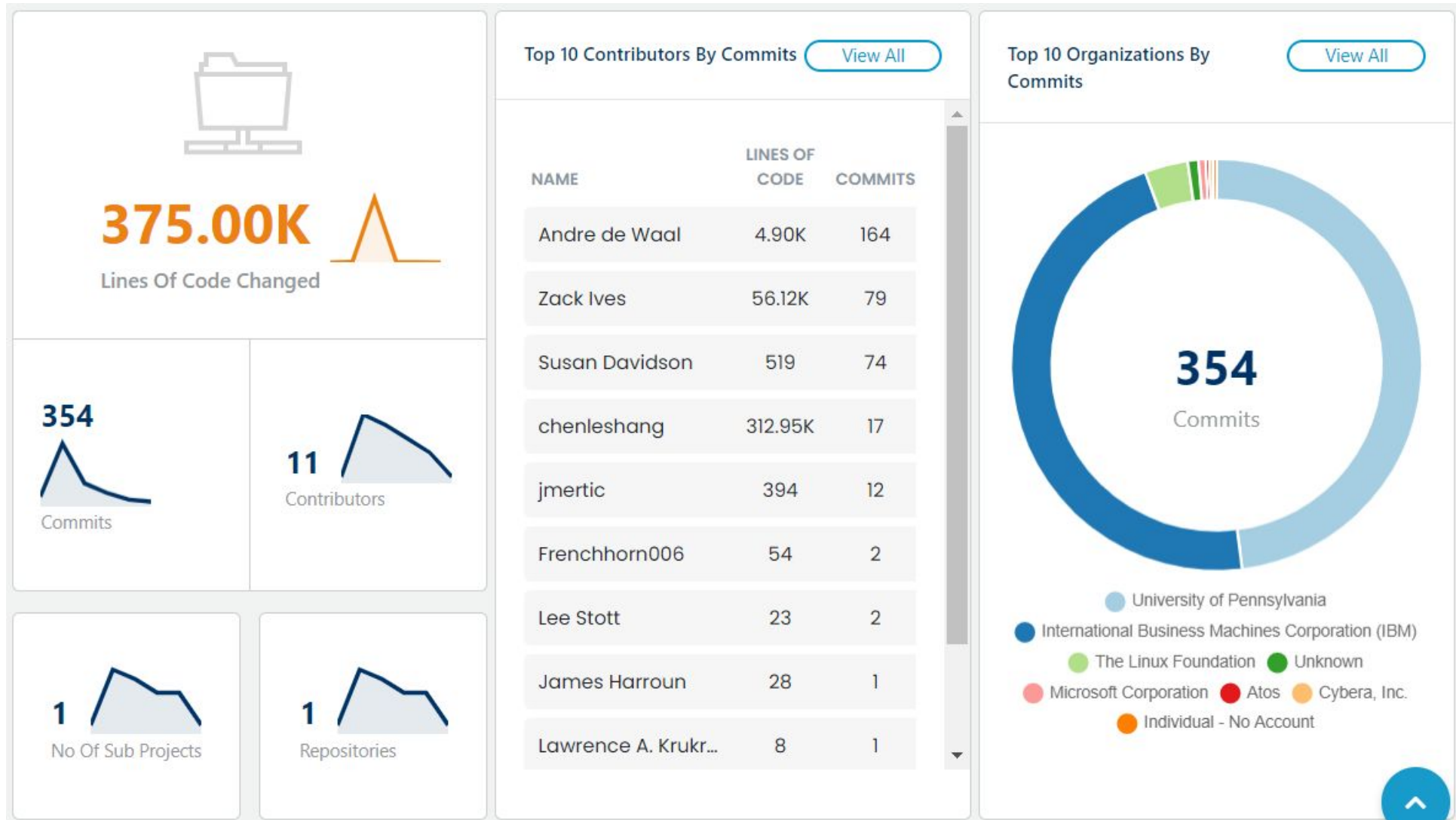


# Organizations contributing

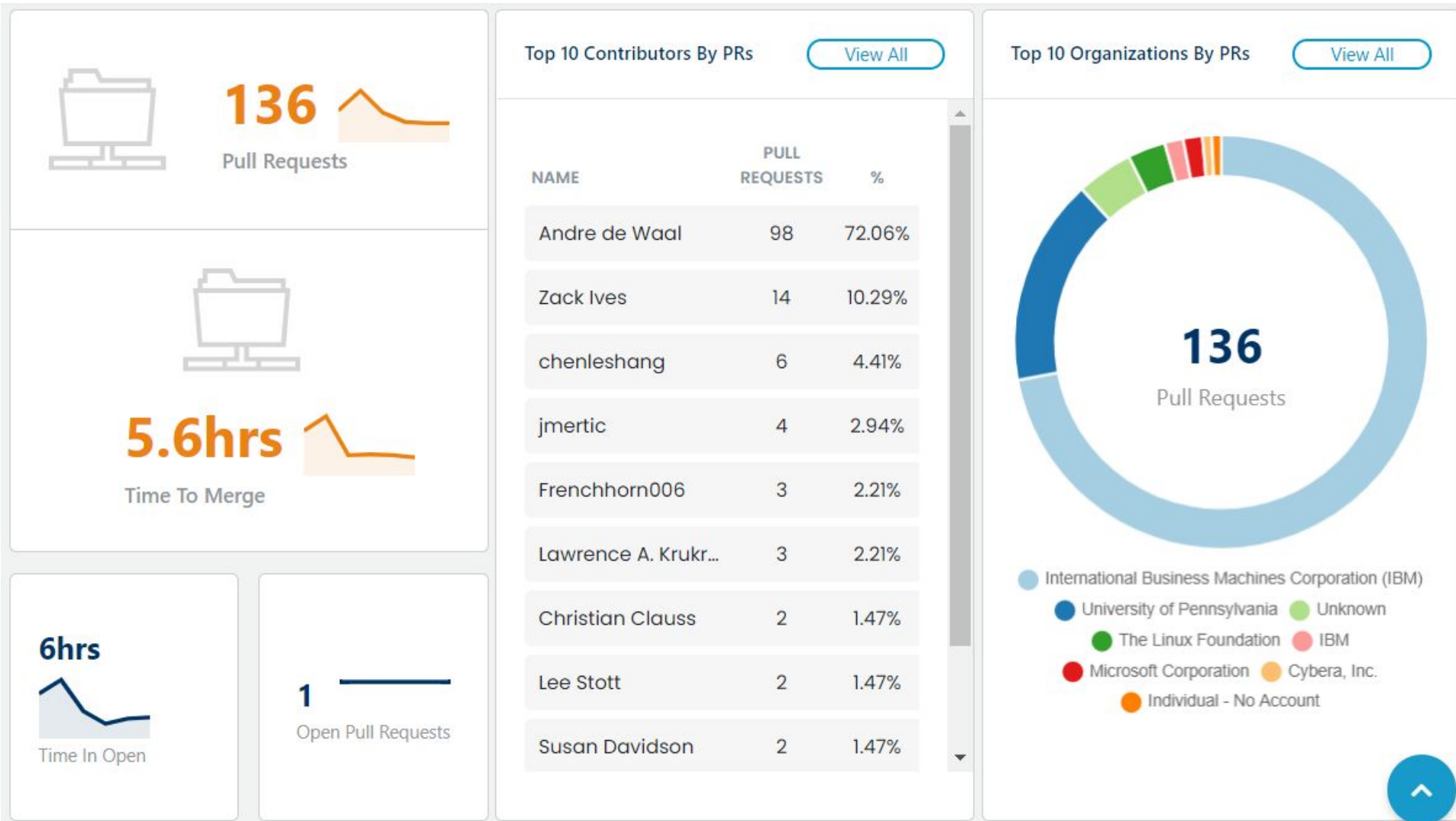
- Linux Foundation
- University of Pennsylvania
- IBM
- SAS Institute
- Kennesaw State University
- Oklahoma State University
- University of Washington at Tacoma
- University of Berkeley
- Duke University

- North Carolina State University

# Contributions



# Contributions





# Key Achievements in the past year

- Binder computing environment integration
- Presentations
  - UIDP 30, ADSA, Data Con LA 2020,
- Adoption of content by:
  - Kennesaw State University
  - Oklahoma State University
  - University of North Carolina at Charlotte
  - Imperial College London
  - University of Liverpool

**OLFAI & DATA** NSDC (Northeast Student Data Corps)

## Do you want to create a world class **Data Science Curriculum?**



A **Data Science Curriculum Kit** built on Open Source (Python)

- Built by professors at the University of Pennsylvania
- Launched under the Linux Foundation Governance to ensure updates and growth
- Initiated by IBM
- Goal: Accelerate the development of Data Science programs at academic institutions with limited resources and build a supporting community

### 1 AN OPEN EDUCATION PROJECT

- Covers the core of Data Science
- Approved content
- Available from everywhere
- No cost

### 2 15 DATA SCIENCE MODULES

- Modules mainly independent of each other
- Covers 50-minute lectures
- Jupyter Notebooks (Python)
- Mix and match

### 3 LICENSES

- CC BY 4.0
- Apache 2.0
- Free to share and adapt with attribution

## Modules (lessons)

			OVERVIEW	FOUNDATION	DATA WRANGLING AND INTEGRATION
			Introduction: What is Data Science	How your Computer Works and Graph Theory	Getting Started in Data Science: Data Acquisition and Wrangling
SCALABLE DATA PROCESSING	SCALABLE DATA PROCESSING	MACHINE LEARNING	MACHINE LEARNING	EXPLORATORY DATA ANALYSIS	DATA AND KNOWLEDGE MODELING
Cluster-Based Processing of Graph Data	Scaling to Big Data and the Cloud	Unsupervised Machine Learning	Supervised Machine Learning: Linear and Logistic Regression	Information Visualization, aka Visual Analytics	Data Representation and Modeling
SCALABLE DATA PROCESSING	SCALABLE DATA PROCESSING	MACHINE LEARNING	MACHINE LEARNING	MODEL ASSESSMENT	ETHICS
Efficient Data Processing	Network Centrality, PageRank and Graph Analysis Using Linear Algebra	Supervised Machine Learning: Overview, Decision Trees, Random Forests	Supervised Machine Learning: Artificial Neural Networks	Training Robust Models	Data Science Ethics

### Collaborate by becoming a

- Adopter (use part or all the content)
- Contributor (contribute lessons, Jupyter Notebooks, Use Cases, etc.)
- Committer (become involved in the day-to-day running of the project and validate new contributions)
- Technical Steering Committee Member (provide overall direction and leadership)

<https://github.com/odpi/OpenDS4All>

For more information, sample curricula, contact [andre.dewaal@ibm.com](mailto:andre.dewaal@ibm.com)

- Incorporation of DATAPRACTICES.org content

# Areas the project could use help on

- Building the OpenDS4All community
  - Adopters
  - Contributors
    - Getting contributions on new areas/developments in Data Science
  - Committers
    - Getting people enthusiastic about Data Science involved
- Collaboration with other projects

# Feedback on working with LF AI & Data

- LF AI & Data have been very supportive!!!
- Put OpenDS4All in contact with universities that are members of LF AI & Data
  - AIVANCITY (France)
  - Pennsylvania State University
- Continued support from Linux Foundation & LF AI & Data TAC Chairperson

# TAC Open Discussion

# Incubation/Graduation Project review criteria

To be accepted into the Graduation stage, a project must meet the Incubation stage requirements plus:

- 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\*.
- Receive the affirmative vote of two-thirds of the TAC and the affirmative vote of the Governing Board.
- Have completed at least one collaboration with another LF AI & Data hosted project
- Have a technical lead appointed for representation of the project on the LF AI & Data Technical Advisory Council.

# CI Working Group - Project CI usage analysis

# Background

- › John surveyed all projects to get a sense of what CI systems and methodologies were in place. Questions asked:
  - › Do you have a CI/CD environment?
  - › Describe the environment and technologies used
  - › What works well with this environment?
  - › What things would you like to improve on or add to your environment?
- › Responses came from 11 projects:
  - › Angel, DataPractices.org, Datashim, Flyte, Ludwig, Marquez, Milvus, NNStreamer, OpenDS4All, Pyro, RosaeNLG

# Key findings

- All code projects have a CI/CD environment
  - Datapractices and OpenDS4All do not
- GitHub Workflows and TravisCI most popular solutions used
  - Some usage of CircleCI, homegrown solutions.
  - Some on TravisCI moving to GitHub
  - Docker, SonarCloud, Kubernetes, readthedocs also being used
- Challenges
  - Limitations of free tier resources; need for cluster environments and different types of deployment architectures
  - Common asset management tooling ( Dockerhub, etc )

Raw data at

[https://docs.google.com/spreadsheets/d/e/2PACX-1vSgNRr4WhH5rTIQxfcd3mhTlqaxdn6a1nqfg3rjch-cLk8rzDq1\\_bAVAsiaB8i\\_pkxZbAay-5fPszb/pubhtml](https://docs.google.com/spreadsheets/d/e/2PACX-1vSgNRr4WhH5rTIQxfcd3mhTlqaxdn6a1nqfg3rjch-cLk8rzDq1_bAVAsiaB8i_pkxZbAay-5fPszb/pubhtml)



## Next steps ( suggested )

- › Collaborate on “CI Best Practices” document outlining recommendations and resources for projects.
  - › API Documentation hosting
  - › Test case result hosting
- › Look at funding specialized infrastructure ( clusters, GPUs, different architectures ).
- › Other ideas?

# LF AI & Data - General Updates

 LF AI & DATA

### Machine Learning

<b>Framework</b>	<b>Platform</b>	<b>Library</b>
Accord AI, FedAI, Microsoft LightGBM, MAHOUT, ML.NET, RAY	Angel (Graduated), ForestFlow (Incubating), H2O, Kubeflow, mflow, Project30, SELDOV, VIZUAL MARKET	Carboast, mlpack, openml, Ecarr, 將軍 (Ji Jiangun), Sonnet, xgboost, xgboost-translator, xgboost

### Deep Learning

<b>Framework</b>	<b>Platform</b>	<b>Library</b>	<b>Tool</b>
SINGA, Chainer, CNTK, dyNet, Alibaba Euler, MindSpore, mxnet, nnon, 深度学习 (Deep Learning), Pythia, PYTORCH, TensorFlow	DeepDetect, Determined AI, jina, Chopsal, Polyaxon, TensorFlow Extended (TF Extended)	BigDL, Catalyst, DL4J, Keras, PyTorch, PyTorch Lightning	BoTorch, Intel Distiller, SpleadML, Intel OpenVINO, PyTorch, TVM

### Reinforcement Learning

<b>Reinforcement Learning</b>
OpenAI Gym, Coach, DeepMind Lab, Dopamine, Horizon, OpenAI, Google PaaS, Google SEED RL, Stable Baselines

### Programming

<b>Programming</b>
Pyro (Graduated), Julia, MARS, NumPy, NYOKA, pomegranate, PyMC3, SciPy, Stan

### Notebook Environment

<b>Notebook Environment</b>
Apache Databricks, BeakerX, Colab, Elyra, IPython, Jupyter, IBM, PolyNote, Streamlit

### Data

<b>Data</b>
JanusGraph (Incubating), Milvus (Incubating), ALLUXIO, ICEBERG, ARROW, DELTA LAKE, druid, hudi, Parquet, pilosa, VEARCH, vespa

### Operations

<b>Operations</b>
Amundsen (Incubating), dataShim (Incubating), MARQUEZ (Incubating), FEAST (Incubating), Hive, cloudf, DataHub, DVC, Oulit, WH FLABS whylogs, tsfresh

### Feature Engineering

<b>Feature Engineering</b>
FEAST (Incubating), NNStreamer (Incubating), Flink, kafka, PULSAR, samza, Uber uReplicator

### Stream Processing

<b>Stream Processing</b>
Apache Drill, HAWQ, presto, SQLFlow, bokeh, plotly, Uber deck.gl, Ecco, Google Facets, Grafana, RCloud, NYU, OO SuperSet, Google

### Pipeline Management & Annotation

<b>Pipeline Management &amp; Annotation</b>
Labeling, Labelbox, Labeling, HITACHI, Microsoft VATT, EGERIA (Graduated)

### Model

<b>Model</b>
ADLIX (Incubating), KFerving, MNN, NVIDIA, NVIDIA, uTensor, MLPerf

### Parameter

<b>Parameter</b>
ONNX (Graduated), Uber NeuroPod, Acumos (Graduated), IBM

### Workflow

<b>Workflow</b>
Flyte (Incubating), Airflow, nifi, argo, Arkaban, Cadence, Couler, Cylone, Databolt, Spotify, mlmap, PREDICT, TRAINS

### Tool

<b>Tool</b>
Qualcomm AIMEE, Facebook dlm, Microsoft Azure, Amazon MMS, Amazon Neo-AI, Amazon Netron, PipelineAI, studio.ml, Google, AWS TorchServe, Turi

### Explainability

<b>Explainability</b>
AI Explainability 360 (Incubating), Microsoft, University of Washington Line, Google Lucid, SHAP, SKATER, Bolt, TrainInterpreter

### Adversarial

<b>Adversarial</b>
Adversarial Robustness Toolbox (Incubating), AdvBox, AdvTorch, hars, Foolbox, AI Fairness 360 (Incubating), Aquitas, Audit AI, Fairlearn

### Distributed Computing

<b>Distributed Computing</b>
EDL (Incubating), SOAJS (Incubating), Bahir (Incubating), sparkly (Incubating), MESOS, Apache Spark, SPARK, GNS3, APACHE TORRE, LOVY, NETFLIX genie, kubernetes, Nauta, Singularity



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

[lfaidata.foundation](https://lfaidata.foundation)



### Security & Privacy




<b>Security &amp; Privacy</b>
Google, HELIB, Microsoft SEAL, SUBSTR, Google, TensorFlow Privacy, TensorFlow


### Natural Language Processing










<b>Natural Language Processing</b>
DELTA (Incubating), RosaeNLG (Sandbox), Google ALBERT, Allen NLP, Google Bert, fastText, flair, LUON, Kashgari, Facebook LAGER, fastText, intel NLP Architect, INLP, ParAI, Facebook Pytorch, RASA, spaCy, Transformers, Facebook XLM, YouTokenToMe

### Education

<b>Education</b>
DATAPRACTICES.ORG (Incubating), OpenDS4All (Incubating)

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

Notebook Environment	Store & Format	Versioning	Operations	Feature Engineering	Stream Processing	SQL Engine	Visualization	Pipeline Management	Labeling & Annotation	Governance
										 Graduated

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

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



# Suggested Additions

## Project Key

**Yellow** = not in [Landscape](#), maybe should be added

## Programming

[Numpy](#)  
[Numba](#)  
[SciPy](#)  
[Dask](#)  
[Julia](#) (\*)  
[Python](#)  
[Rstudio](#)

## Notebooks

[Flyra](#)  
[I-python](#)  
[Jupyter Notebooks](#)  
[PixieDust](#)  
[Rmarkdown](#)

## Security & Privacy

[HE-Lib](#) (\*)  
[TensorFlow Privacy](#)  
[TF-Encrypted](#)

## Distributed Computing

*Management*  
[OpenShift](#)  
[Kubernetes](#)  
[Mesos](#)  
[Ranger](#)  
[Storm](#)

*Interface*  
[Sparklyr](#)  
[Toree](#)  
[Livy](#)  
[Spark-NLP](#)

## Data

*Versioning*  
[Pachyderm](#) (\*)

*Store & Format*  
[Alluxio](#)  
[Arrow](#)  
[Avro](#)  
[Delta Lake](#) (\*)

[Druid](#)  
[JanusGraph](#)  
[Parquet](#)  
[Ceph](#)

*Stream Processing*

[Flink](#)  
[Kafka](#)  
[Logstash](#) (\*)  
[FluentD](#) (\*)

*Relational DB*

[Postgres](#)  
[MySQL](#)  
[CouchDB](#)

*SQL Engine*  
[Presto](#) (\*)

*Visualization*

[Bokeh](#)  
[D3](#)  
[Plotly](#)  
[Facets](#)  
[Grafana](#)  
[Seaborn](#)  
[Superset](#) (\*)  
[TensorBoard](#)  
[Prometheus](#)

## Data

*Governance*  
[Egeria](#)  
[CLDA](#)

*Feature Engineering*  
[Tsfresh](#)

*Operations*  
[FEAST](#) (\*)  
[Amundsen](#) (\*)  
[Hive](#) (\*)  
[Snorkel](#) (\*)

*Pipeline Management*  
[Beam](#)

*Labeling & Annotation*  
[Vott](#) (\*)

*Exploration*  
[Hue](#)  
[Kibana](#)

## Machine Learning

*Framework*  
[LightGBM](#)  
[Mahout](#)  
[Ray](#) (\*)

*Platform*  
[Kubeflow](#)  
[H2O](#)  
[SystemML](#)  
[Mlflow](#) (\*)  
[Seldon](#) (\*)  
[Marvin-AI](#) (\*)

*Library*  
[Scikit-learn](#)  
[XGBoost](#)  
[cat-boost](#)  
[SparkML](#)

## Deep Learning

*Framework*  
[TensorFlow](#)  
[PyTorch](#)  
[MX-Net](#)

*Library*  
[Keras](#)

## Reinforcement Learning

[DeepMind Lab](#) (\*)  
[OpenAI Gym](#) (\*)

## Model

*Inference*  
[TensorRT](#)  
[TensorRT Inference](#)

*Benchmarking*  
[MLPerf](#)

*Training*  
[Horovod](#) (\*)

*Parameter*  
[HyperOpt](#)  
[Katib](#)

*Format & Interface*  
[ONNX](#)

*Marketplace*  
[MAX](#) (\*)

*Workflow*  
[Kubeflow Pipelines](#)  
[Tekton](#)

[Airflow](#) (\*)  
[Nifi](#) (\*)  
[Argp](#) (\*)  
[Mleap](#) (\*)  
[Volcano](#) (\*)

*Tool*  
[KFServing](#)  
[ONNX Runtime](#)  
[TorchServe](#) (\*)  
[Clipper](#) (\*)  
[MMS](#) (\*)

## Trusted AI

*Explainability*  
[AI Explainability 360](#)  
[Alibi](#) (\*)  
[LIME](#)  
[SHAP](#)

*Bias & Fairness*  
[AI Fairness 360](#)

*Adversarial Attacks*  
[Adversarial Robustness Toolbox](#)

## Natural Language Processing

[UIMA](#)  
[BERT](#)  
[Core NLP](#)  
[Lucene](#)  
[PyText](#)  
[Spacy](#)  
[Transformers](#) (\*)

*Education*  
[OpenDS4All](#)

## 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: TBD	6: Egeria & OpenDS4All - project update 20: TBD - Project updates	3: TBD - Project updates 17: TBD - Project updates
Jul Aug Sep	8: TBD - Project updates 22: TBD - Project updates	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























# Projects (28)

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


## Graduated LF AI & Data Projects (6)

 <b>Acumos</b> ★ 15 LF AI & Data Foundation	 <b>Angel</b> ★ 6,149 Angel-ML LF AI & Data Foundation	 <b>Egeria</b> ★ 326 LF AI & Data Foundation	 <b>Horovod</b> ★ 10,913 LF AI & Data Foundation	 <b>ONNX</b> ★ 9,871 LF AI & Data Foundation	 <b>Pyro</b> ★ 6,784 LF AI & Data Foundation
---	---	--	---	--	--

## Incubating LF AI & Data Projects (20)

 <b>ADLIK</b> ★ 237 Adlik LF AI & Data Foundation	 <b>Adversarial Robustness Toolkit</b> ★ 2,054 LF AI & Data Foundation	 <b>AI Explainability 360 Toolkit</b> ★ 782 LF AI & Data Foundation	 <b>AI Fairness 360 Toolkit</b> ★ 1,246 LF AI & Data Foundation	 <b>Amundsen</b> ★ 1,864 LF AI & Data Foundation	 <b>DataPractices</b> ★ 10 LF AI & Data Foundation	 <b>Datashim</b> ★ 79 LF AI & Data Foundation	 <b>DELTA</b> ★ 1,394 LF AI & Data Foundation	 <b>Elastic Deep Learning (EDL)</b> ★ 125 LF AI & Data Foundation	 <b>Feast</b> ★ 1,539 LF AI & Data Foundation
 <b>Flyte</b> ★ 1,217 LF AI & Data Foundation	 <b>ForestFlow</b> ★ 43 LF AI & Data Foundation	 <b>JanusGraph</b> ★ 3,869 LF AI & Data Foundation	 <b>Ludwig</b> ★ 7,587 LF AI & Data Foundation	 <b>Marquez</b> ★ 518 LF AI & Data Foundation	 <b>Milvus</b> ★ 5,259 LF AI & Data Foundation	 <b>NNStreamer</b> ★ 325 LF AI & Data Foundation	 <b>OpenDS4All</b> ★ 238 LF AI & Data Foundation	 <b>SOAJS</b> ★ 43 LF AI & Data Foundation	 <b>sparklyr</b> ★ 775 LF AI & Data Foundation

## Sandbox LF AI & Data Projects (1)

 <b>RosaeNLG</b> ★ 12 LF AI & Data Foundation
---

\* Missing Substra Framework (pending logo redesign)

# New projects in 2021

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.

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

# 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

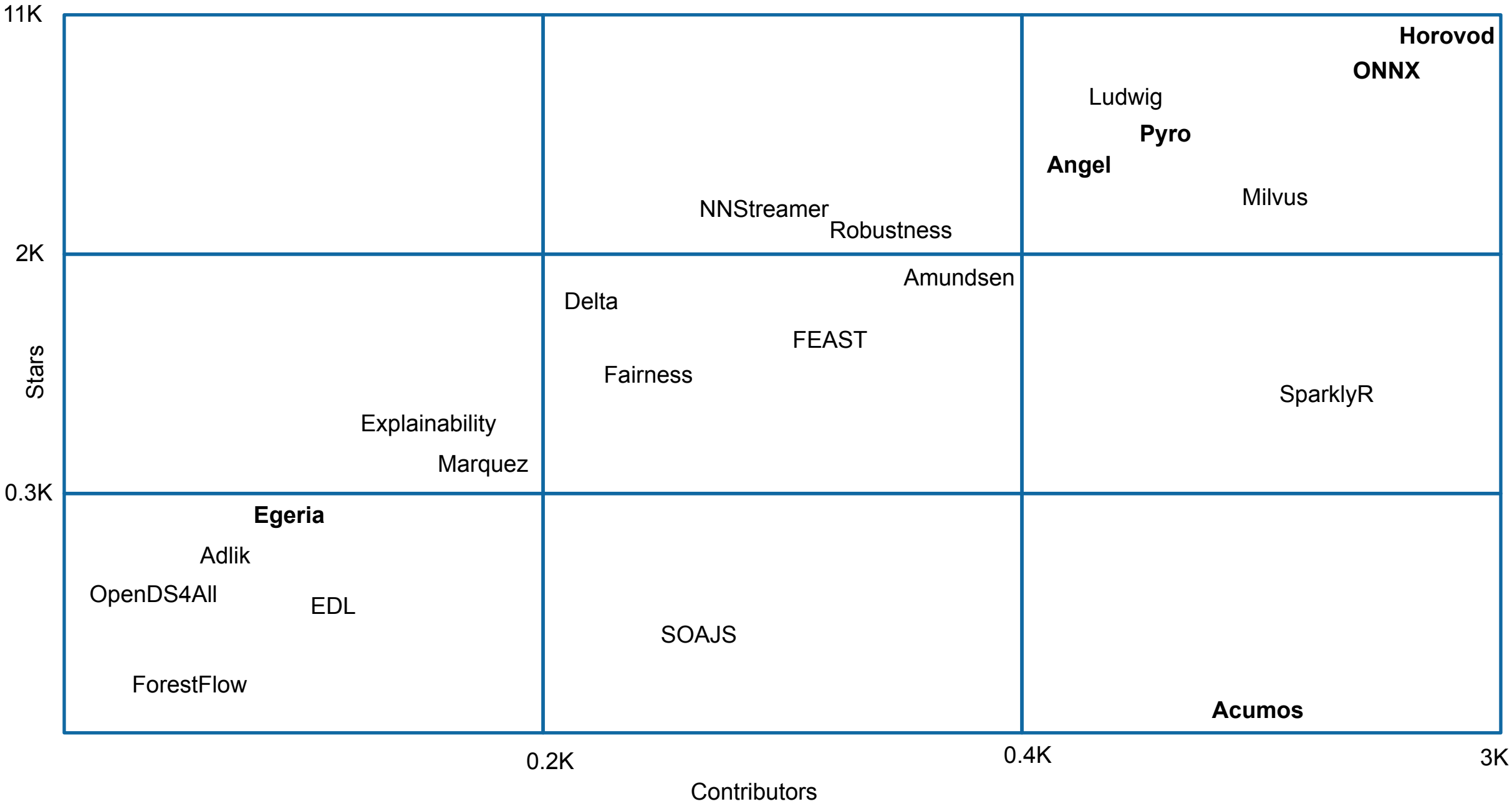
**369**

Repositories

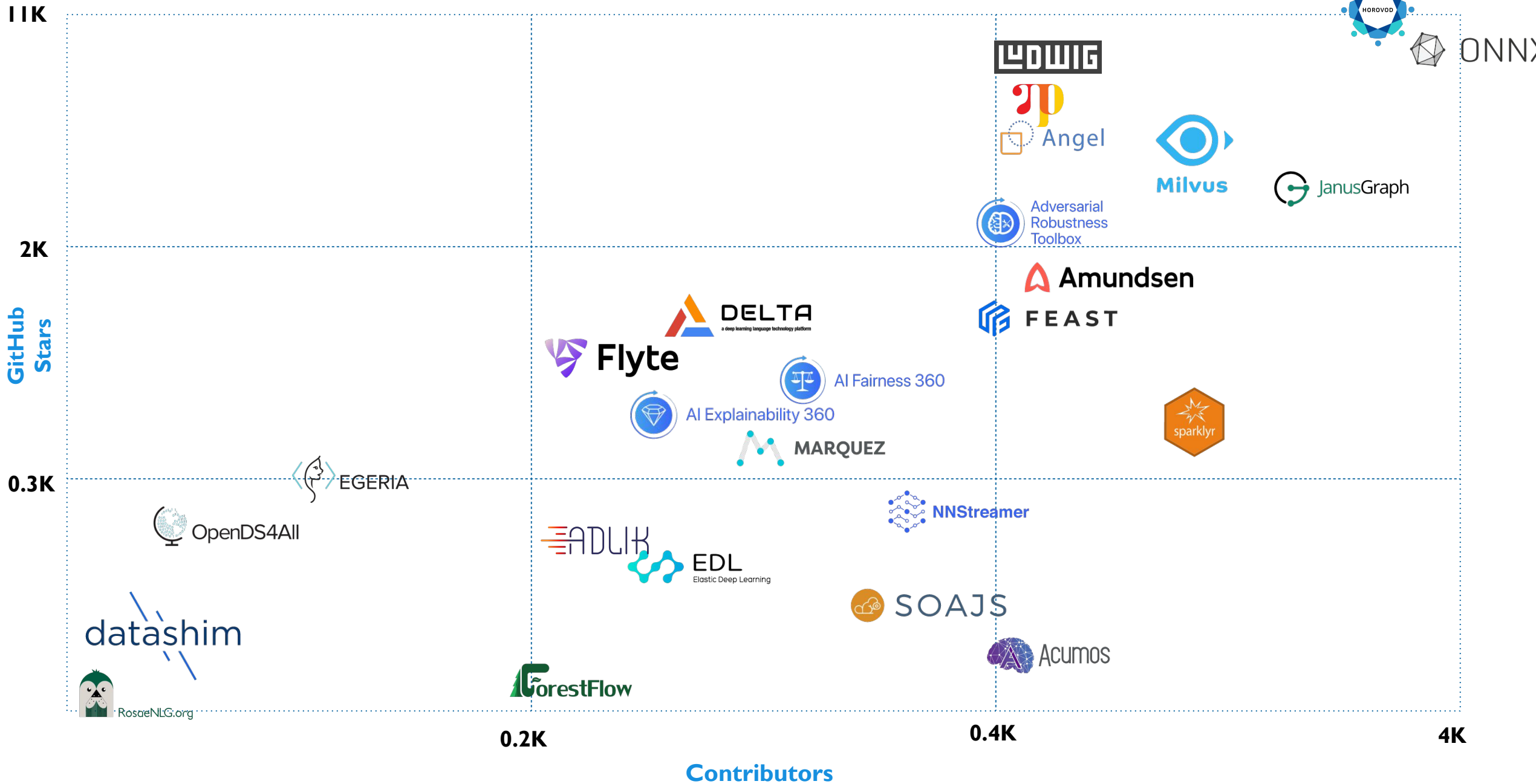
**3.51K**

Slack messages

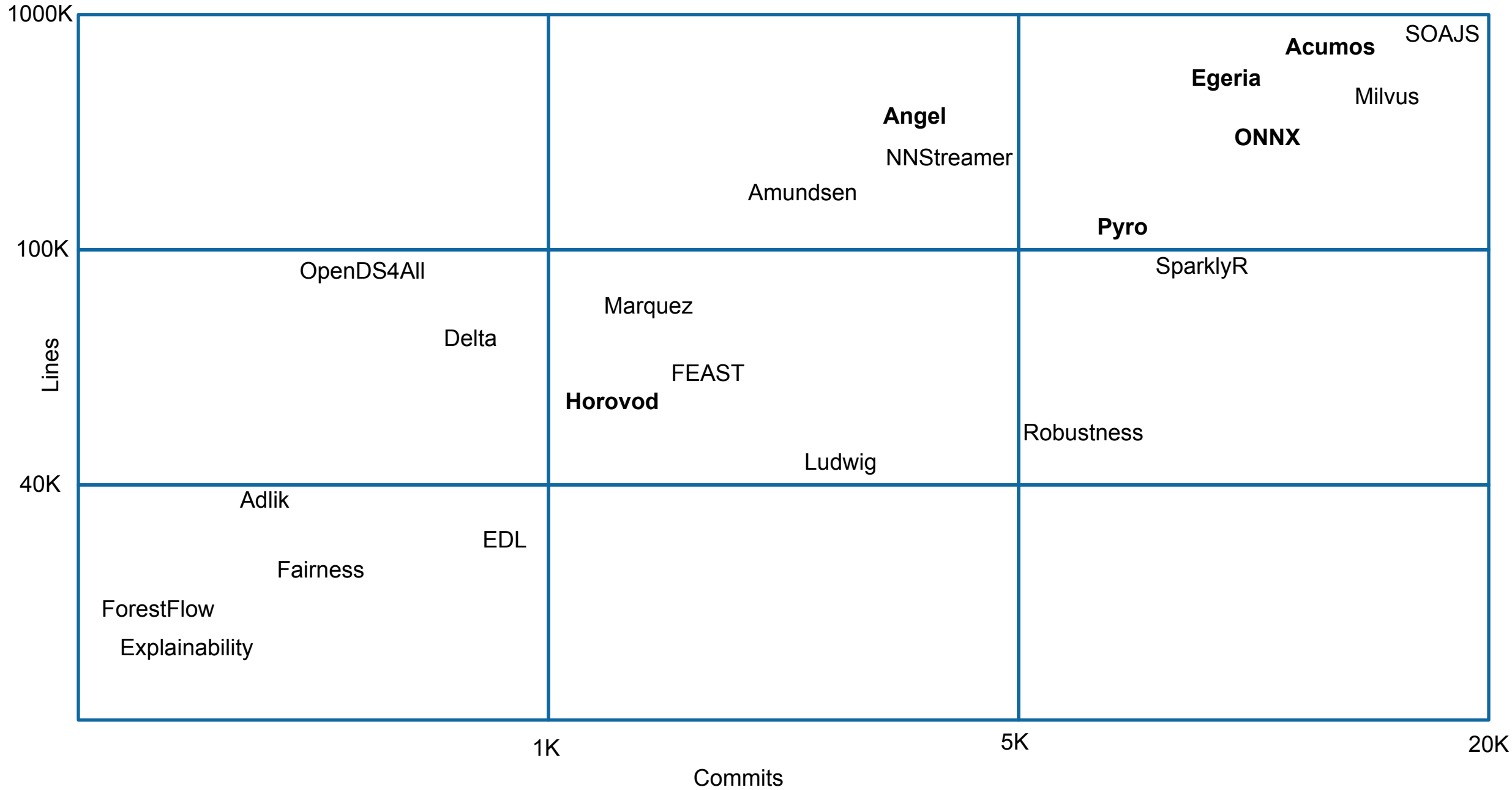
Data from November 23, 2020 – Stars and Contributors



# 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 16 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/>

- **LF Energy Spring Summit (virtual) - April 14th - [Event Website](#)**
  - a. Booth, Keynote by Ibrahim Haddad, LF AI & Data Executive Director
- **Kubernetes AI Day (virtual) - May 4th - [Event Website](#)**
  - a. Booth
- **OSS Global (hybrid) Sept 29 - Oct 1 - [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)

- › May 6: Alejandro Saucedo (Kompute) ,Animesh Singh (ML Exchange)
- › May 20: Julian Le Dem & Mandy Chessell “Open Lineage” sandbox proposal
- › Jun 3: KOSA.ai
- › June 17: TBD - Adlik Graduation
- › July 8: TBD - Annual project review
- › July 22: TBD - Annual project review
- › 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.





