Meeting of the Technical Advisory Council (TAC)

February 11, 2021

LF 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 Updegrove 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
- Wiki: wiki.lfaidata.foundation
- GitHub: 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
-
- 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 (2 mins) - welcome Pyro!
- Approval of Minutes (2 mins)

- Invited Presentation (40 minutes)
  - Egeria CI - Build and Release (Nigel Jones (Egeria, IBM))
- Invited Presentation (15 minutes)
  - Mentorships (John Mertic (LF))

- LF AI General Updates (2 minutes)
- Open Discussion (2 minutes)

TAC Voting Members

* = still need backup specified on wiki

Board Member	Contact Person	Email
AT&T	Anwar Atfab	anwar@research.att.com
Baidu	Ti Zhou	zhouti@baidu.com
Ericsson	Rani Yadav-Ranjan*	rani.yadav-ranjan@ericsson.com
Huawei	Huang Zhipeng*	huangzhipeng@huawei.com
IBM	Susan Malaika	malaika@us.ibm.com
Nokia	Jonne Soininen*	jonne.soininen@nokia.com
SAS	Nancy Rausch	nancy.rausch@sas.com
Tech Mahindra	Nikunj Nirmal	nn006444@techmahindra.com
Tencent	Bruce Tao	brucetao@tencent.com
Zilliz	Jun Gu	jun.gu@zilliz.com
ZTE	Wei Meng	meng.wei2@zte.com.cn
Graduate Project	**Contact Person**	**Email**
Acumos	Nat Subramanian	natarajan.subramanian@techmahindra.com
Angel	Bruce Tao	brucetao@tencent.com
Egeria	Mandy Chessell	mandy_chessell@uk.ibm.com
Horovod	Travis Addair*	taddair@uber.com
ONNX	Jim Spohrer (Chair of TAC)	spohrer@us.ibm.com
Pyro	Fritz Obermeyer*	fritz.obermeyer@gmail.com

Approval of January 28th, 2021 Minutes

Draft minutes from the January 28th TAC call were previously distributed to the TAC members via the mailing list

Proposed Resolution:

- That the minutes of the January 28th meeting of the Technical Advisory Council of the LF AI & Data Foundation are hereby approved.

Invited Presentation - Egeria CI - Build and Release

Nigel Jones (Egeria Project, IBM)

LF AI & DATA

Egeria – Build & Release

Nigel Jones – Egeria Maintainer
2021-02-09

LF AI & DATA

About Egeria

- Commits: 3.3k (1yr), 8.8k (total) in base repo (15k inc merges)
- Contributors: 32 (1yr), 57 (total)
- Started with single Java core project
- Primarily maven based (400+ artifacts)
- More repos & technologies – Go, Python/notebooks, javascript, node, docker, helm
- Release every month

What do we do currently –dev (all branches)

- GitHub actions
- PR
  - Build/test with maven & gradle on Java 8 & Java 11 on Linux
  - Unit tests & FVTs
  - CodeQL code scanning – security issue
  - DCO
- Merge
  - Docker images
  - Windows (slow….)
- Monthly Dependabot scan for dependency updates (security/currency)
- Additional tools for security scans (static/dependency) – NexusIQ, Sonar

Repo Management

- egeria – Main repo, original location for all code
- egeria-react-ui
- egeria-ui
- egeria-api-mocks
- egeria-samples
- egeria-k8s-operator
- egeria-dev-projects
- egeria-connector-crux
- egeria-connector-Hadoop-ecosystem
- egeria-connector-ibm-information-server
- egeria-database-connectors
- egeria-connector-sas-viya
- egeria-palisade

Recent improvements

- Migrated PR/merge builds to GitHub actions from Azure pipelines
  - More reliable & faster
  - Active ecosystem
  - Easy to add extra capabilities
- FVT tests against active server
- Code refactoring to streamline dependencies alongside corresponding repackaging & repo moves rather than all in base with maven
- Started on gradle (2-3x clean improvement, > 100x for incremental), avoids some maven limitations

Release (as of Feb 2021)

- Branch
- Additional tests
- Update versions
- Release notes
- Azure Pipelines : Build->Artifactory->BinTray->JCenter->Maven Central
- Tag, create GitHub release
- Socialize

Release: Impacted by JFrog Announcement

- https://jfrog.com/blog/into-the-sunset-bintray-jcenter-gocenter-and-chartcenter/
- Move release pipeline to GitHub actions
- Publish to GitHub maven repo a) snapshots b) release
- Publish docker images to github maven repo (dockerhub limits)
- Additionally publish/sync to Maven Central)
- Drop use of JFrog Artifactory/JCenter/BinTray
- Faster, more maintainable, can integrate with release process, version changes etc
- Concern: Limits of free account (#executors)

Other Work in progress

- Dealing with more repos- process, dependencies, refactoring of samples, certificates, testing, Clearer contract of what a repo delivers
- Improved testing & coverage
  - K8s for test - within a CI pipeline (KinD) – ie for postgres connector (in progress)
  - Conformance Test Suite
  - Notebooks
  - Cross-version testing
  - Capture more coverage data
- Packaging BOMs?
- Reports, Javadoc, licensing, release notes publishing automation

Invited Presentation - Mentorships

John Mertic (Linux Foundation)

LF AI & DATA

Invited Presentation - TBD - will do at a future time

Steven Winslow (Linux Foundation)

LF AI & DATA

Why mentorships?

- For students, it helps them learn real-world software development skills and time management
- For projects, it brings in new contributors that become maintainers with codebase experience.
- For members, it highlights up and coming talent to hire.

We generally see LF projects leverage these platforms

LFX | Mentorship

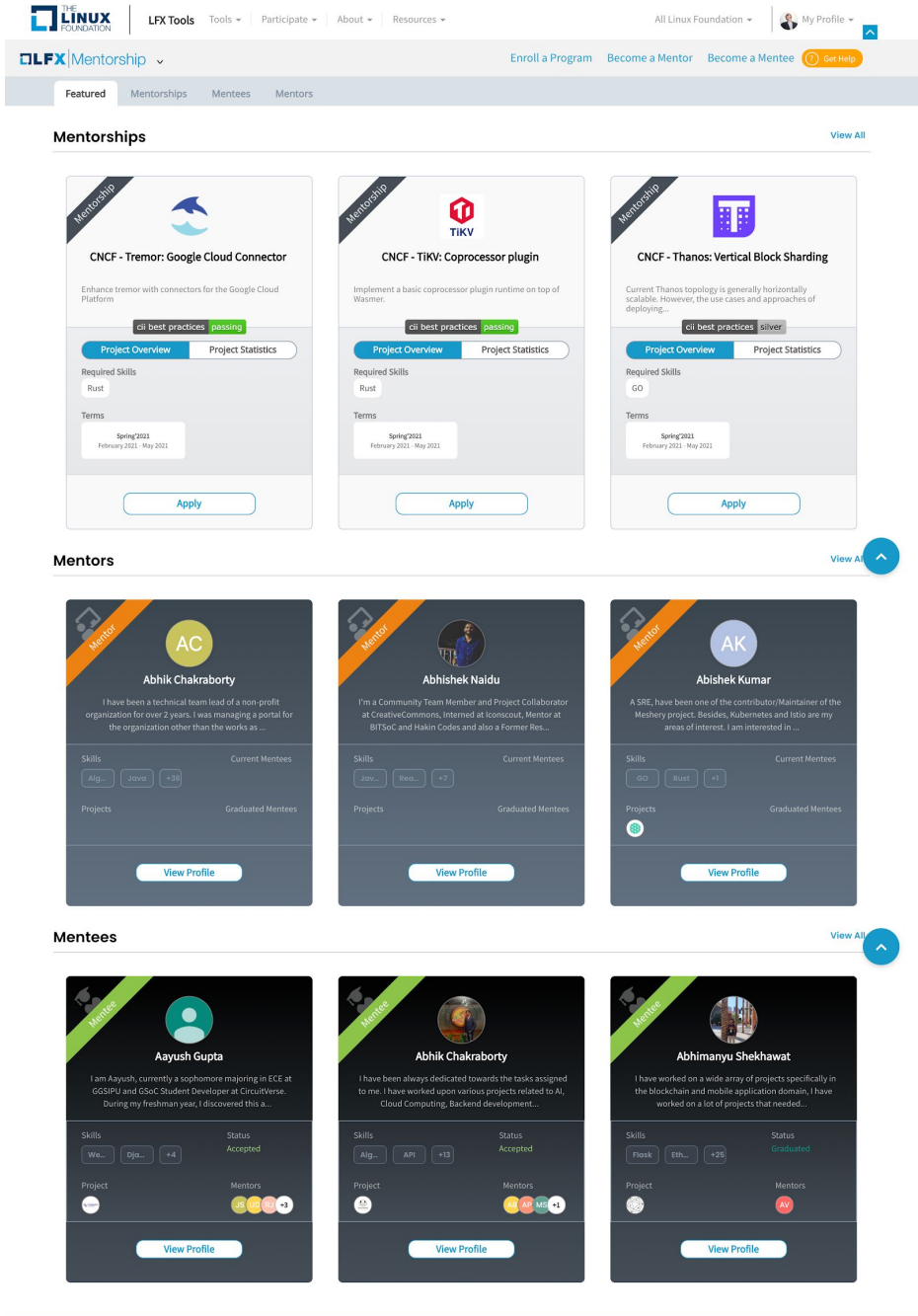
Google Summer of Code

What is LFX Mentorships?

Part of the LFX platform, it provides tooling for attracting mentees and supporting mentorships.

Project providing the mentorship provides funding, but there is the opportunity to crowdfund to support mentorships.

Three mentorship terms, spring, summer, and fall, gives the project flexibility on timelines



LF AI & DATA

Google Summer of Code

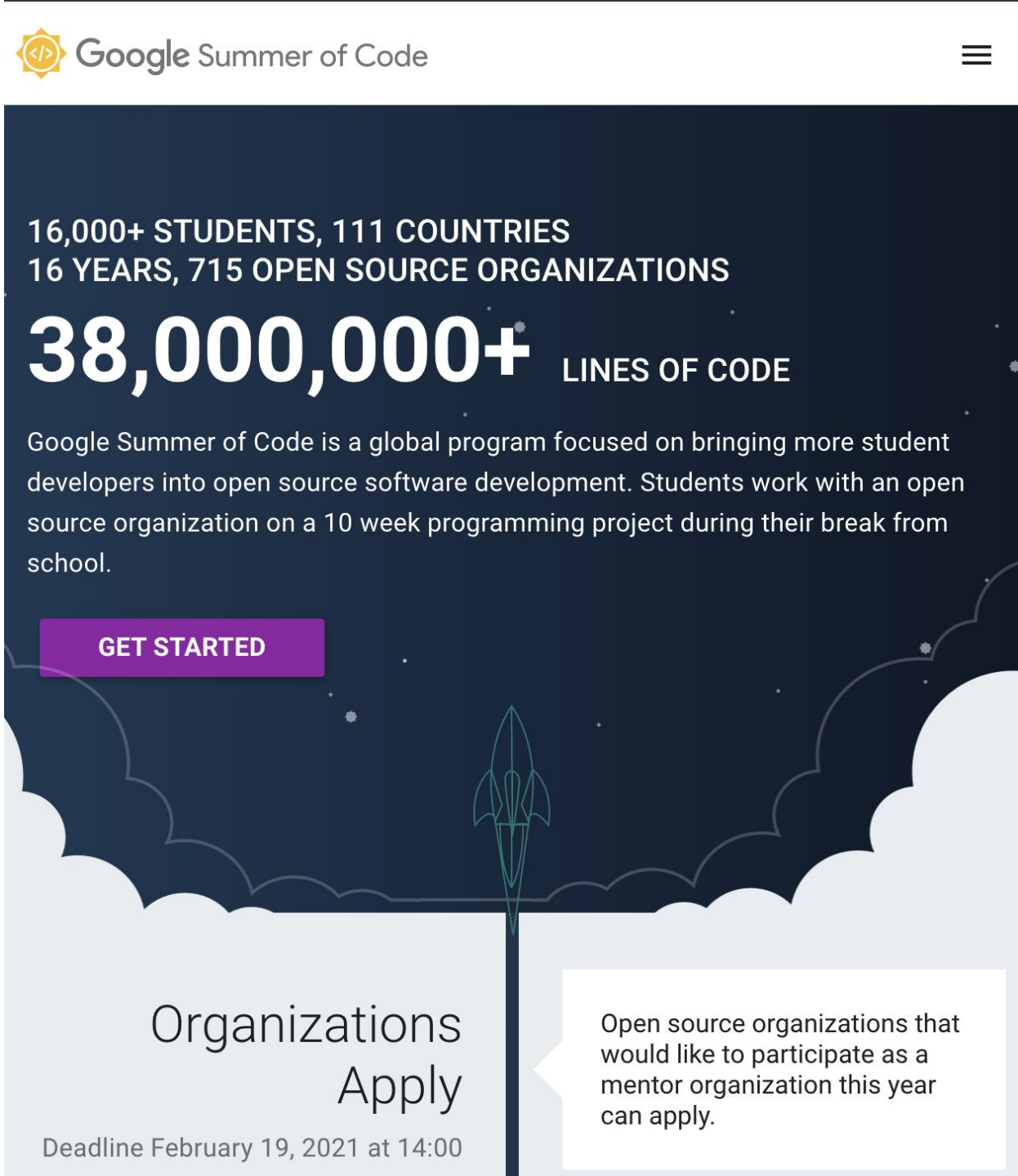
Summer program sponsored by Google for the past 16 years.

Open source organizations can apply to be a mentoring organization, and Google approves and sets a number of mentorships for each organization.

No cost to project, and provides a small stipend ($400 USD) per mentorship.

Deadline to apply is February 19, 2021

LF AI & DATA



What to do if LF AI and Data wants to do this

1. Form a working group to manage mentorships
2. Get going on the platfomr
   - Apply to GSoC as an organization (if this path is of interest)
   - Add projects to LFX Mentorship that want to do mentorships (if this path is of interest).
3. Start a project ideas page where projects can list mentorship project ideas or link to issue list of mentorship ideas
   - Example: https://github.com/cncf/mentoring/blob/master/summerofcode/2020.md
4. Be ready for inbound interest!
5. Select final mentorships and mentees

LF AI & Data - General Updates

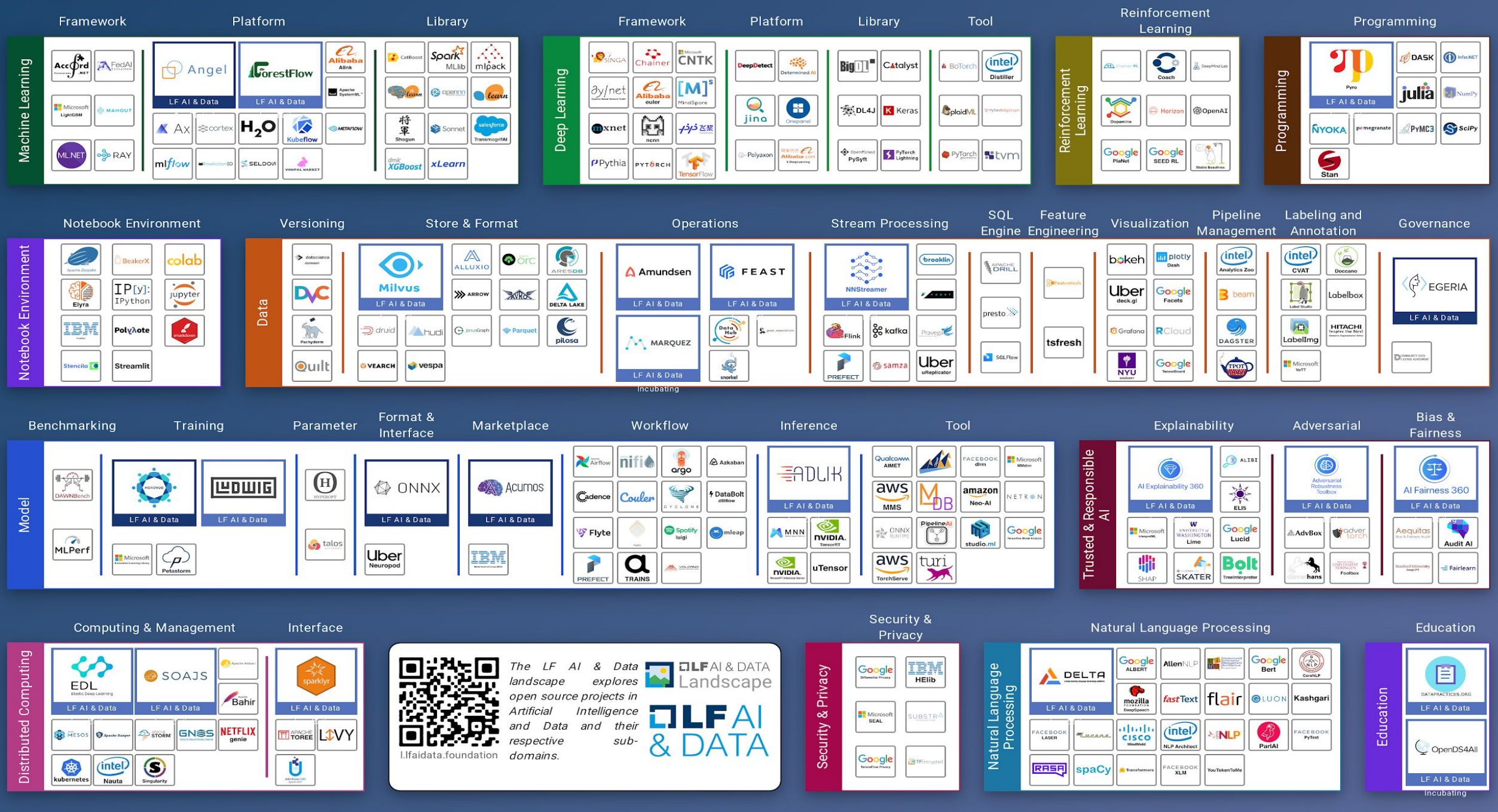
LF AI & DATA

Linux Foundation AI Landscape

2020-11-22T22:28:27Z b08f33f

See the interactive landscape at l.lfai.foundation

Greyed logos are not open source



Framework
Platform
Library
Machine Learning
Angel
LF AI & Data
ForestFlow
LF AI & Data
Framework
Platform
Library
Tool
Deep Learning
Reinforcement Learning
Reinforcement Learning
Programming
Programming
Pyro
LF AI & Data
Notebook Environment
Notebook Environment
Versioning
Store & Format
Operations
Stream Processing
SQL Engine
Feature Engineering
Visualization
Pipeline Management
Labeling and Annotation
Governance
Data
Milvus
LF AI & Data
Amundsen
LF AI & Data
FEAST
LF AI & Data
MARQUEZ
LF AI & Data
Incubating
NNStreamer
LF AI & Data
EGERIA
LF AI & Data
Benchmarking
Training
Parameter
Format & Interface
Marketplace
Workflow
Inference
Tool
Model
LF AI & Data
LF AI & Data
ONNX
LF AI & Data
Acumos
LF AI & Data
ADLIK
LF AI & Data
Explainability
Adversarial
Bias & Fairness
Trusted & Responsible AI
AI Explainability 360
LF AI & Data
Adversarial Robustness Toolbox
LF AI & Data
AI Fairness 360
LF AI & Data
Computing & Management
Interface
Distributed Computing
EDL
LF AI & Data
SOAJS
LF AI & Data
sparklyr
LF AI & Data
The LF AI & Data landscape explores open source projects in Artificial Intelligence and Data and their respective sub-domains.
l.lfaidata.foundation
LF AI & DATA Landscape
LF AI & DATA
Security & Privacy
Security & Privacy
Natural Language Processing
Natural Language Processing
DELTA
LF AI & Data
Education
Education
DATAPRACTICES.ORG
LF AI & Data
OpenDS4All
LF AI & Data
Incubating

Suggested Additions

Project Key

Yellow = not in Landscape, maybe should be added

Programming
Numpy
Numba
SciPy
Dask
Julia (*)
Python
Rstudio

Notebooks
Elyra
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

Libarary
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

LF AI & DATA

2020 TAC Meetings Summary

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

(Entity)* = incubating vote

** **bold = graduate vote**

Italics = invited project presentation

2021 TAC Meetings Pipeline Summary

Months

Jan Feb Mar	14: Data Lifecyle Framework (IBM)* 28: Tentative: Verse (Seldon)	11: MARS (Aliabab) 25: Flyte (Lyft)	11: Streams (IBM) 25: Tentative: Substra Framework
Apr May Jun	8: Adlik (ZTE)** 22: Kubeflow-On-Prem (Google, Arrikto, Intel)	?: Ray (Anyscale.io) ?: Pachyderm (Pachyderm) ?: DataHub (LinkedIn)	?: Common Knowledge (Code Reef) ?: Couler (Ant Financial)
Jul Aug Sep	?: KubeflowServing (Google, Arrikto, Seldon)	?: Kubeflow Pipeline (Google, Bloomberg)	?: Open Data Hub (Red Hat)
Oct Nov Dec	?: Vespa (Verizon Media)	?: Snorkle (Snorkle) ?: Plotly (DASH) ?: Mellody (Substra) ?: mloperator (Polyaxen) ?: SnapML (IBM)	?: PMML/PFA (DMG.org) ?: Mindspore, Volcano (Huawei) ?: TransmorgrifAI (Salesforce) ?: AIMET (Qualcomm) ?: Elyra-AI (IBM)

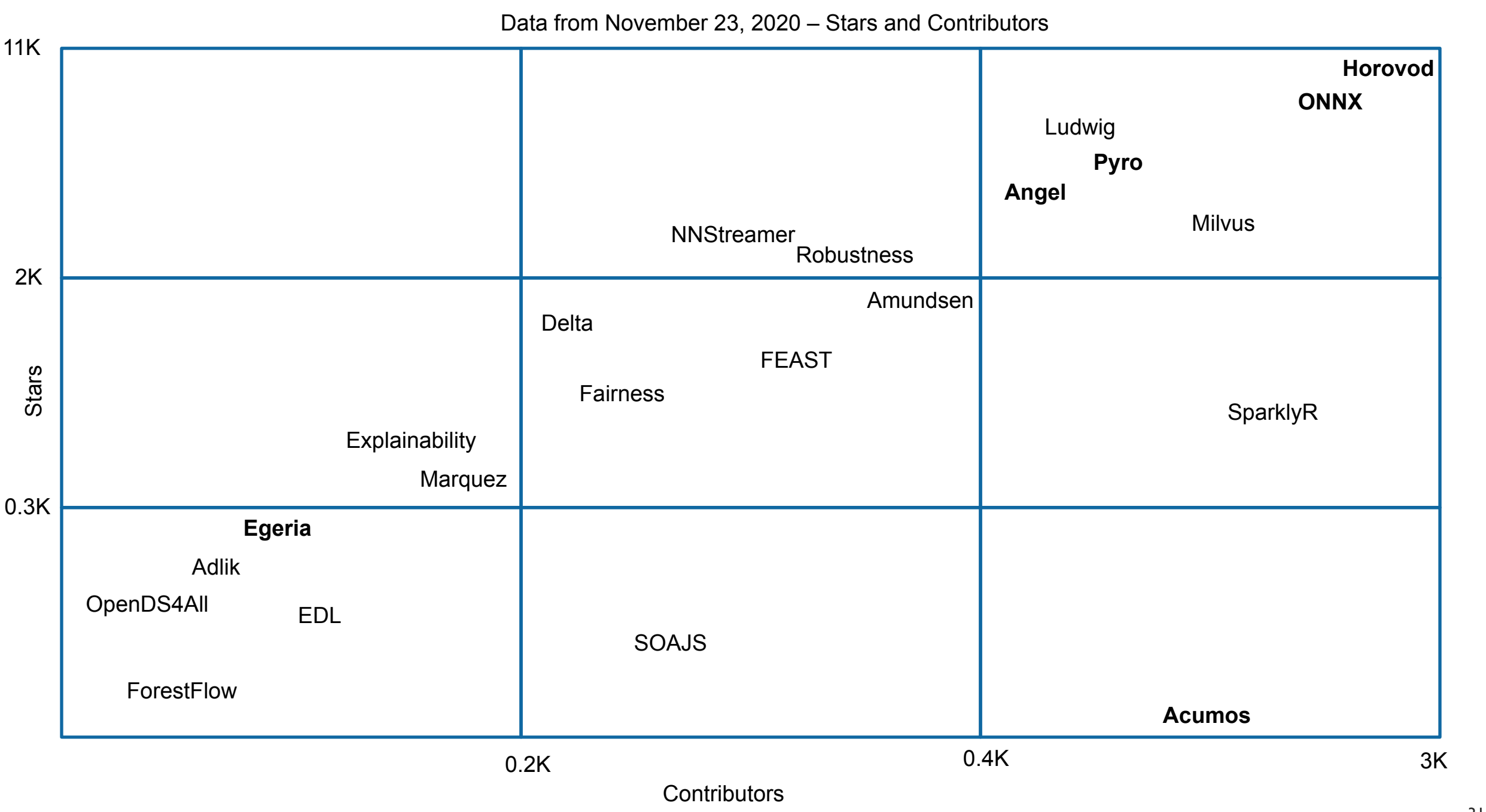
(Entity)* = incubating vote

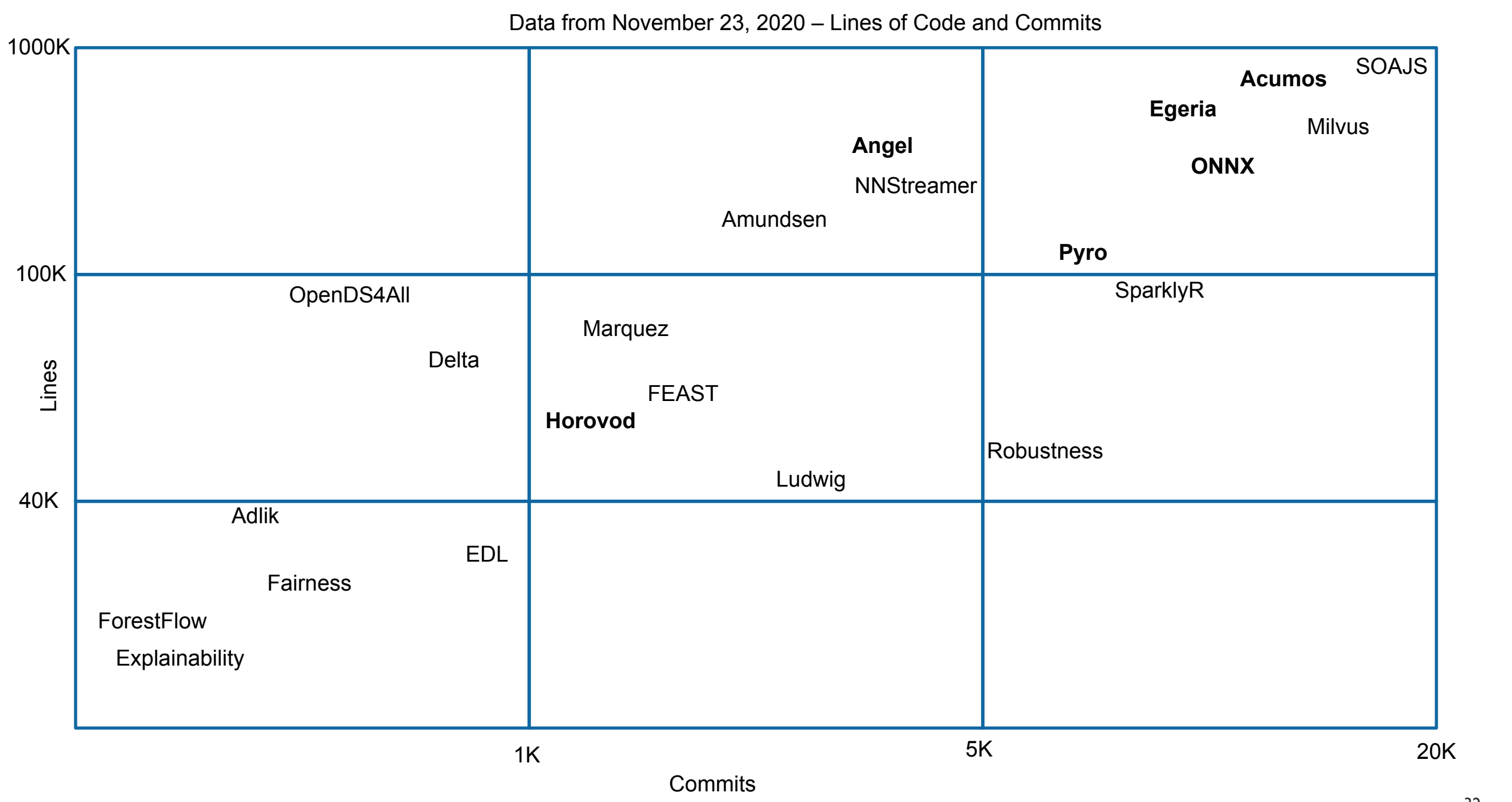
** **bold = graduate vote**

Italics = invited project presentation

Getting to know the projects more

LF AI & DATA


Data from November 23, 2020 – Stars and Contributors
11K
2K
0.3K
Stars
0.2K
0.4K
3K
Contributors
Horovod
ONNX
Ludwig
Pyro
Angel
Milvus
NNStreamer
Robustness
Amundsen
Delta
FEAST
Fairness
SparklyR
Explainability
Marquez
Egeria
Adlik
OpenDS4All
EDL
SOAJS
ForestFlow
Acumos


Data from November 23, 2020 – Lines of Code and Commits
1000K
100K
40K
Lines
SOAJS
Acumos
Egeria
Milvus
Angel
ONNX
NNStreamer
Amundsen
Pyro
SparklyR
OpenDS4All
Marquez
Delta
FEAST
Horovod
Robustness
Ludwig
Adlik
EDL
Fairness
ForestFlow
Explainability
1K
5K
20K
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 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

LF AI & DATA

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

- Please consider holding virtual events

To discuss participation, please email events@lfaidata.foundation

Upcoming Events

https://lfaidata.foundation/events/

- **March 24, 2021 - ONNX Community Virtual Meetup**
  a. **Wednesday @ 5:00 pm - 8:00 pm PT USA**
     Thursday @ 8:00am - 11am China Time
     LF AI Day: ONNX Community Virtual Meetup – March 2021
     (Virtual - Free - Asia-friendly time – Host Ti Zhou - Baidu)
- **August 4-6, 2021 - OSS North America, Vancouver**
  a. **Mini-Summit, Booth, Track**
- **Sept 29 - Oct 1, 2021 - OSS Europe, Dublin**
  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

Call to Participate in Ongoing Efforts

LF 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

LF AI & DATA

Upcoming TAC Meetings (Tentative)

- Feb 25: FLYTE (Lyft)
- Mar 11: TBD
- Mar 25: Substra Foundation
- April 8: Adlik (ZTE)
- April 22: TBD
- May 6: All project updates

Please send agenda topic requests to
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

LF AI & DATA

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

- he 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 with any questions about The Linux Foundation's policies or the notices set forth on this slide.

LF AI & DATA

LF AI & DATA

LF AI & DATA