Meeting of the Technical Advisory Council (TAC)

April 8, 2021



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: Ifaidata.foundation

Wiki: <u>wiki.lfaidata.foundation</u>

> GitHub: github.com/lfaidata

> Landscape: https://landscape.lfaidata.foundation or https://landscape.lfaidata.foundation or https://landscape.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/leiDNJvXCqSZHT4Zk -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 March 25 (5 mins)
- Invited Presentation (25 minutes)
 - > JINA AI (Han Xiao)
- > LF AI General Updates (10 minutes)
- > Open Discussion (15 minutes)



TAC Voting Members

* = still need backup specified on wiki

| | <u>'</u> | | |
|------------------|----------------------------|--|--|
| 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 March 15th, 2021 Minutes

Draft minutes from the March 25^h TAC call were previously distributed to the TAC members via the mailing list

Proposed Resolution:

That the minutes of the March 25th meeting of the Technical Advisory Council of the LF AI & Data Foundation are hereby approved.



Invited Presentation - JINA AI

Han Xiao <han.xiao@jina.ai>

TLFAI & DATA

Invited Presentation: JINA Al

An easier way to build neural search on the cloud.

Jina is geared towards building search systems for any kind of data, including <u>text</u>, <u>images</u>, <u>audio</u>, <u>video</u> and <u>many</u> <u>more</u>. With the modular design & multi-layer abstraction, you can leverage the efficient patterns to build the system by parts, or chaining them into a <u>Flow</u> for an end-to-end experience.

Presenter: Han Xiao <han.xiao@jina.ai>

Resources:

Github: https://github.com/jina-ai/jina

Stars: 2686

Website: https://jina.ai/





What is Neural Search?

Searching unstructured data using unstructured query

- Search image via image
- Search video via video
- Search audios via audio
- Search items via sketch
- Search PDF via image
- Search audio via image
- Search map via image





What is Neural Search?

Searching unstructured data using unstructured query

- Search image via image
- Search video via video
- Search audios via audio
- Search items via sketch
- Search PDF via image
- Search audio via image
- Search map via image

Multi-modality Search

Cross-modality Search





What is Neural Search?

Searching unstructured data using unstructured query

- Search image via image
- Search video via video
- Search audios via audio
- Search items via sketch
- Search PDF via image
- Search audio via image
- Search map via image

Multi-modality Search

Cross-modality Search

Deep Learning

Cloud-Native

OSS ecosystem





Jina Al is



Tensorflow for search



A design pattern



The next Elastic









Our mission

We provide an **open-source** neural search ecosystem for **businesses and developers**, allowing everyone to search for information in **all kinds of data** with high availability and scalability.



Search Al Opensource Cross-border



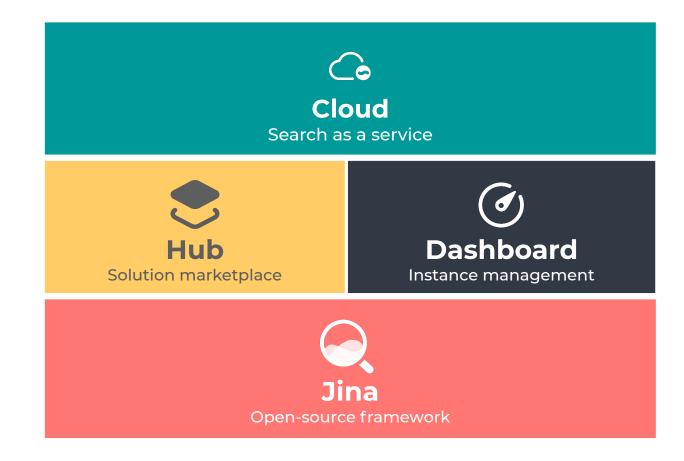
- 1 -

Products



Product Overview

Jina Al Product Landscape







Jina is "Tensorflow for Search"





Universal deep learning framework





Universal neural search framework

The **design pattern** of neural search

Opensource as the core, ecosystem around it





Multi/Cross-Modality Search on **Unstructured Data**



Cloud-Native



Search Al in Production

Universal



Full-stack Ownership



Time-saver



Plug&Play





Shop-the-look



Fashion director Shelly Vella's pick of the season's most exuberant prints





Current Solutions





How to implement "Shop-the-look" in 5 min

```
from jina import Flow

f = Flow().add(uses='cnn-encoding.yml').add(uses='simple-index.yml')

with f:
    f.index(image_data)
```



How to implement "Shop-the-look" in 5 min

```
Use Docker image
from jina import Flow
f = (Flow().add(uses='docker://preproc', host='123.45.678.9:8000')-
           .add(uses='cnn-encoding.yml', parallel=3)
           .add(uses='simple-index.yml', shards=3))`
                                                                  Decentralized
with f:
 f.index(image_data)
                                                 Parallelization & Sharding
```



Saving human time



What just happened?

This is Jina's hello-world, end-to-endly. It downloads Fashion-MNIST dataset and indexes 60.00 stored into multiple shards. We then randomly sample unseen images as *Queries*, ask Jina to retriwhere the left-most column is query image.

Intrigued? Learn more about Jina and checkout our Github!

margaret zeum more acour ma and enconcer car cran

Precision@50: 63.16% Recall@50: 0.53%

Query Top-K Results



Shop the look, Ecommerce

LOC: ~150 50L Python + 93L YAML

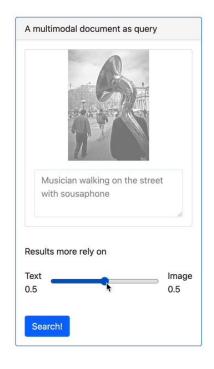
Save >1000 hours



Chatbot, QA, Customer service

LOC: ~50 21L Python + 27L YAML

Save >500 hours



Multimodal, Rich doc search

LOC: ~250

9L Python + 237L YAML

Save >1500 hours



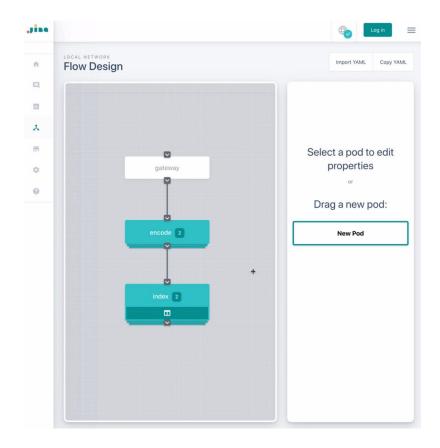
Full stack developer experience on Neural Search

Full-stack env.

Low-code env.

No-code Env.

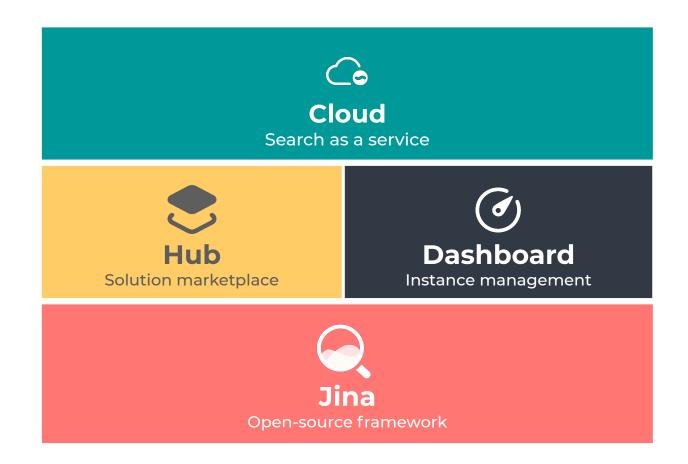
```
!Flow
version: '1.0'
pods:
    - uses: docker://preproc
    host: 123.45.6.789
    - uses: cnn-encoding.yml
    parallel: 3
    - uses: simple-index.yml
    parallel: 3
```





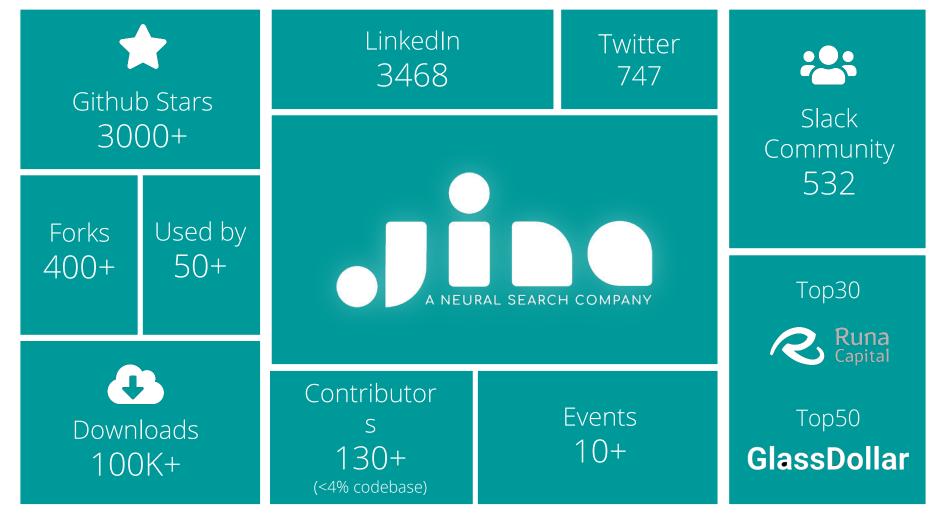
Product Overview

Jina Al Product Landscape



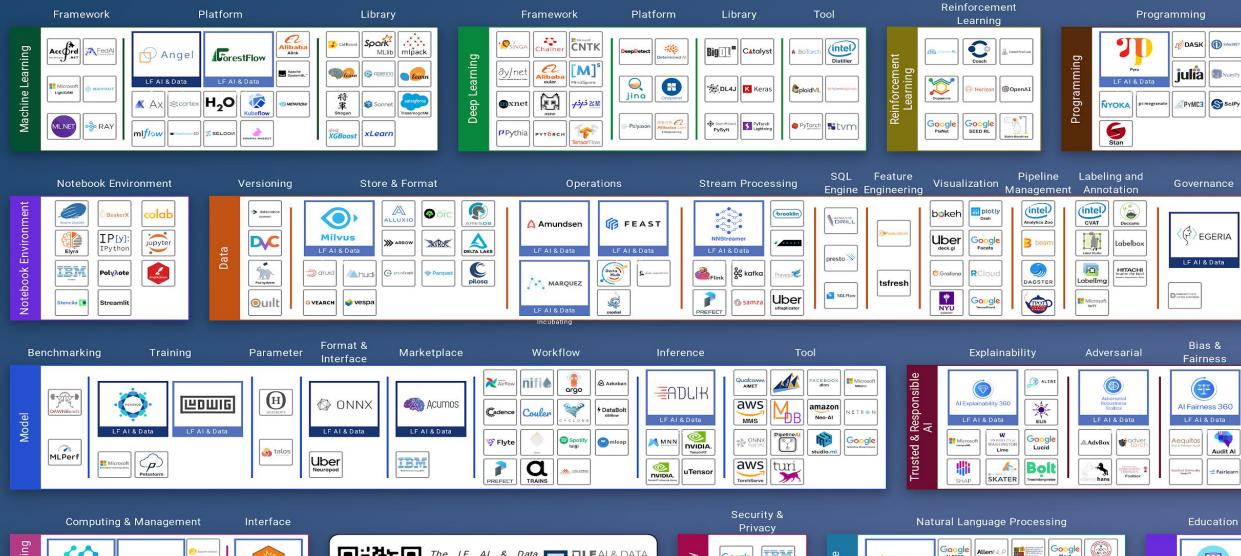


Since 2020 May,



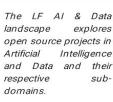
LF Al & Data - General Updates

ILFAI & DATA



















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 <u>Parquet</u>

Ceph

Stream Processing

Flink Kafka

Logstash (*) FluentD (*)

Relational DB

Postgres MySOL

<u>CouchDB</u>

SOL Engine Presto (*)

Visualization Bokeh <u>D3</u> Plotly **Facets**

Grafana Seaborn Superset (*) **TensorBoard Prometheus**

Data

CLDA

Governance Egeria

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 H20 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 Katih

Format & Interface **ONNX**

Marketplace **MAX** (*)

Workflow Kubeflow Pipelines **Tekton**

Airflow (*) Nifi (*) Argp (*) Mleap (*) Volcano (*)

Tool

KFServing

Clipper (*)

MMS (*)

ONNX Runtime

TorchServe (*)

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 PvText Spacy Transformers (*)

Education OpenDS4All



2020 TAC Meetings Summary

| 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) | 7: Ludwig (Uber)* | 4: Trusted AI (AI for Good, Ambianic.ai, MAIEI) |
| | 23: Open Data Hub (Red Hat) | 21: SnapML (IBM) | 18: Fairness, Explainability, Robustness (IBM)* |
| Months Seb In gep | <i>16: Mindspore (Huawei)</i> 30: Amundsen (Lyft)* | 16: Delta (Didi) 16: Horovod (Uber/LF)** 30: ModelDB (?) 30: Egeria, OpenDS4All, BI&Al (LF ODPi) | 10: SOAJS (HeronTech)* 10: Delta (Didi)* 24: FEAST (Gojek)* 24: Egeria, (LF ODPi)** 24: OpenDS4All (ODPi)* 24: BI&Al 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

| 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-Al (IBM) 25: Substra Framework (Substra)* |
|--|--|---|--|
| Apr May Jun Jul Aug Sep | 8: Invited talk JINA AI 22: TBD | 6: Egeria & OpenDS4All - project update 20: TBD - Project updates | 3: TBD - Project updates 17: TBD - Project updates |
| | 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) ?: Mellody (Substra) ?: mloperator (Polyaxen) ?: SnapML (IBM) | ?: PMML/PFA (DMG.org) ?: Mindspore, Volcano (Huawei) ?: TransmorgrifAl (Salesforce) ?: AIMET (Qualcomm) ?: Elyra-Al (IBM) |

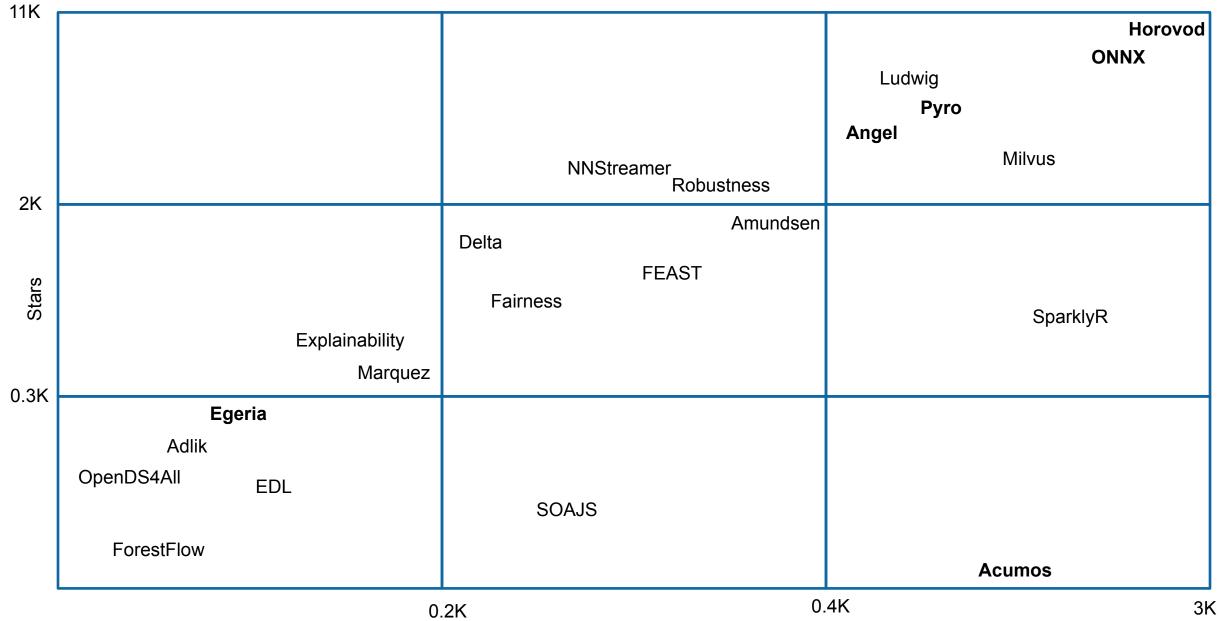
(Entity)* = incubating vote

** bold = graduate vote

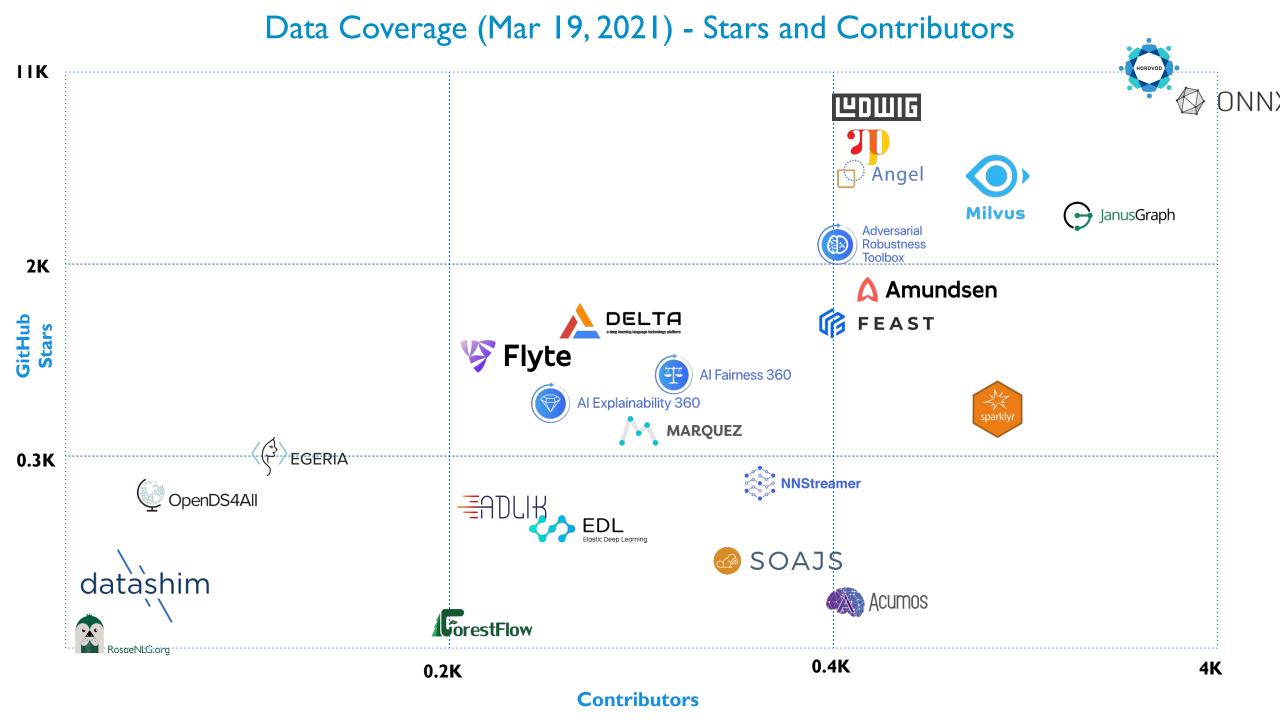
Italics = invited project presentation

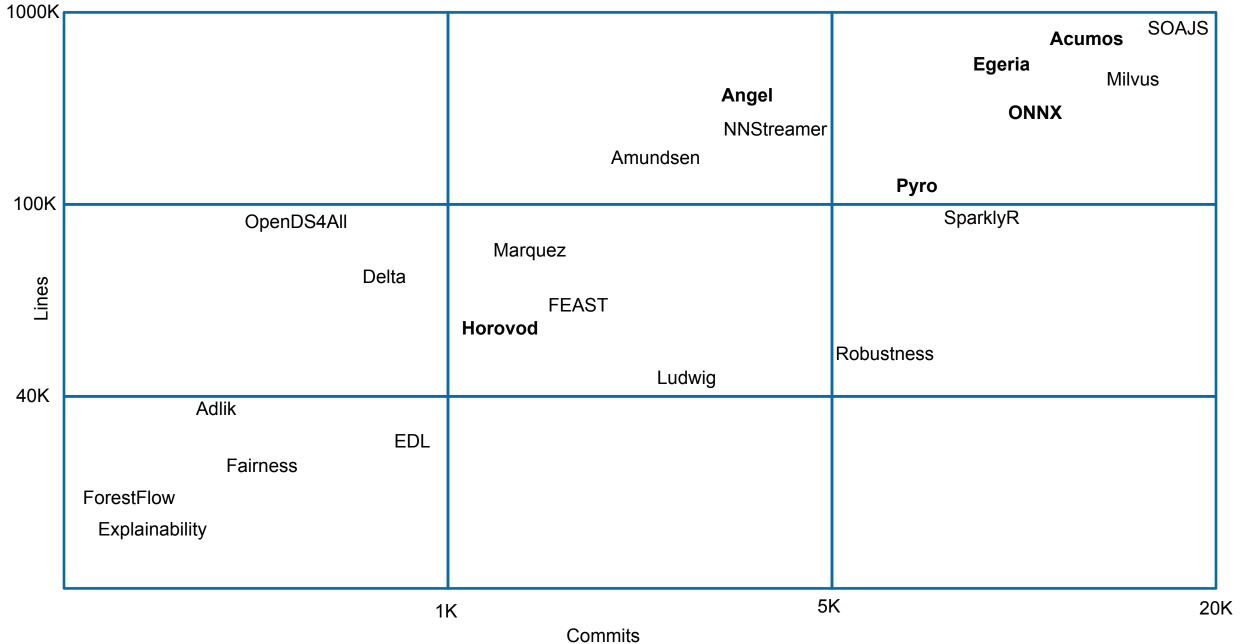
Getting to know the projects more





Contributors





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



Commits

Looking to host a project with LFAI & 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 LFAI & Data Twitter and/or LinkedIn social channels

For links to details on upcoming releases for LFAI & Data hosted projects visit the Technical Project Releases wiki

If you are an LFAI & Data hosted project and would like LFAI & 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

ILFAI & DATA

Upcoming Events

- Upcoming Events
 - > Visit the LFAI & Data Events Calendar or the LFAI & Data 2021 Events wiki for a list of all events
 - To participate visit the LFAI & 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/

- Sept 29 Oct 1, 2021 OSS Global
 - a. Mini-Summit, Booth, Track

LF AI PR/Comms

- Please follow LF AI & Data on <u>Twitter</u> & <u>LinkedIn</u> 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 <u>LFAI & Data Blog</u>

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

TLFAI & DATA

Trusted Al

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+Al+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 Al 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-a nd-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

ILFAI & DATA

Upcoming TAC Meetings (Tentative)

```
April 22: TBD
May 6: Egeria and ODPi - Project updates
May 20: TBD - Project updates
Jun 3: TBD - Project updates
June 17: TBD - Project updates
July 8: TBD - Project updates
July 22: TBD - Project updates
Aug 5: TBD - Project updates
Aug 19: TBD - 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

TLFAI & DATA

Mission

To build and support an open community and a growing ecosystem of open source Al, 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.





