LF AI & Data

June 2021

Ibrahim Haddad, Ph.D.
Executive Director, LF AI & Data
Ibrahim@LinuxFoundation.org

LF AI & DATA

Ecosystem Overview

LF AI & DATA

LF AI & DATA Landscape

305 projects
2.2M+ GitHub Stars
450M+ LoC
40K+ developers

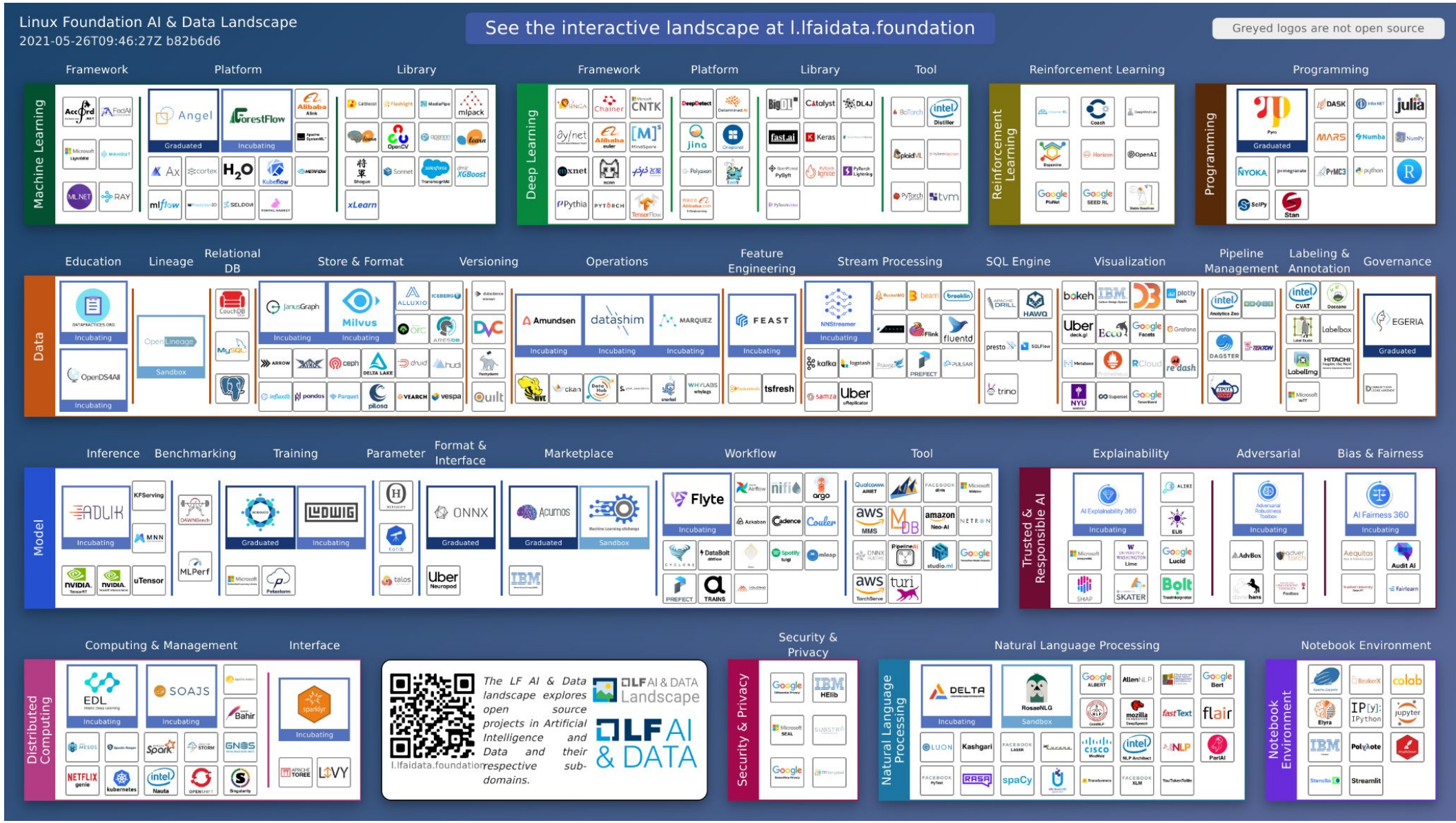
130+ founding orgs

10+ universities

13 licenses

>450M LoC

>1M new LoC weekly



Help us improve the landscape:
https://github.com/lfai/lfai-landscape/

The barrier to entry in AI is lower than ever before, thanks to open source software.

LF AI & DATA

Open source libraries, frameworks, platforms, tools, are a two-way street: they make AI accessible to everyone, and companies benefit from a community of other contributors helping accelerate open AI applied research.

LF AI & DATA

Open source software is a market accelerant

Standardization on open technologies	**Neutral environment for collaboration & cross-pollination**	**Innovation multiplier - community driven**	**Better interoperability**	**Minimizes fragmentation**
Qualifies reference architectures	**Lowers barriers to enter a new domain**	**Enables business opportunities supported by a flexible licensing model**	**Leads to better products, improved quality and security**	**Allows fast TTM and shared cost of development**

Open Source offers *unique* benefits to AI & Data

Fairness

Methods to detect and mitigate bias in datasets and models, including bias against known protected populations

Robustness

Methods to detect alterations/tampering with datasets and models, including alterations from known adversarial attacks

Explainability

Methods to enhance understandability and interpretability by persona/roles in process of AI model outcomes/decision recommendations, incl ranking and debating results and decision options

Lineage

Methods to ensure provenance of datasets and AI models, including reproducibility of generated datasets and AI models

Open Data

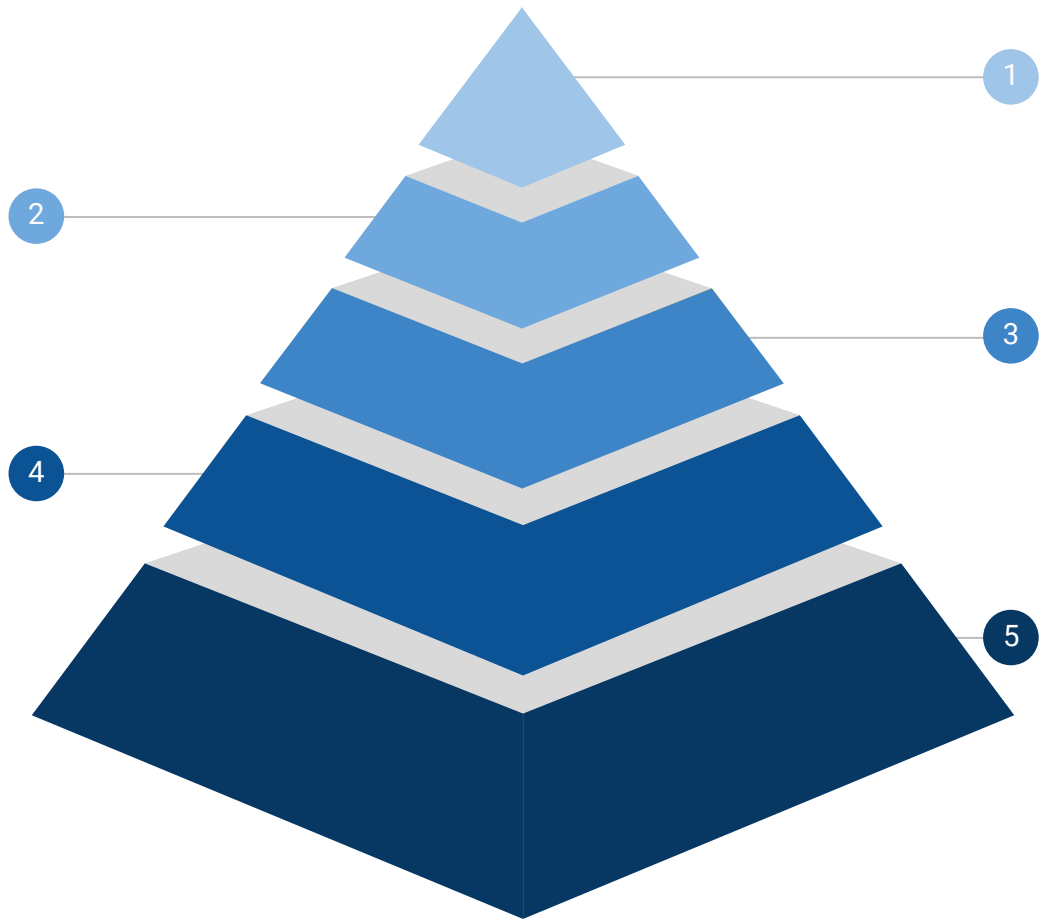
Data Manipulation Tools

Tools to clean, sort, tagg, track provenance, etc., and a governance structure for doing these thing

Ecosystem Challenges

Managing projects assets
As project grows, it is unclear who should own legal and administrative tasks that are essential to the health of the project

Governance challenges
No formal governance, governance favors the creators of the project, project backer dominating dev and governance



Glass ceiling for project investment and adoption
in the absence of a fair governance and safe haven for the project's assets

Developed internally for specific product requirements, highly specialized
The open source spinoff is a result of wanting to build an ecosystem and collaborate on building a platform

Fragmentation
Lack of integration
Lack of cross pollination

Enter LF AI & Data

LF AI & DATA

What is LF AI & Data?

- Non-profit
- Hosted at the Linux Foundation
- Collaborative effort focused on open source AI & Data
- Aims to accelerate open source innovation and development and enable collaboration and the creation of new opportunities for all the members of the community

History & Growth

- Launched in 03/2018 with 9 members and 1 hosted project
- +3 projects in 2018
- +5 projects in 2019
- +15 projects in in 2020
- +8 projects (Jan - May) 2021
- Supported by 49 member companies (all geos)
- Build an ecosystem around critical and high value projects and accelerate the adoption and sustainability of hosted projects
- >2000 subscribers to our main mailing list

Structure and Governance

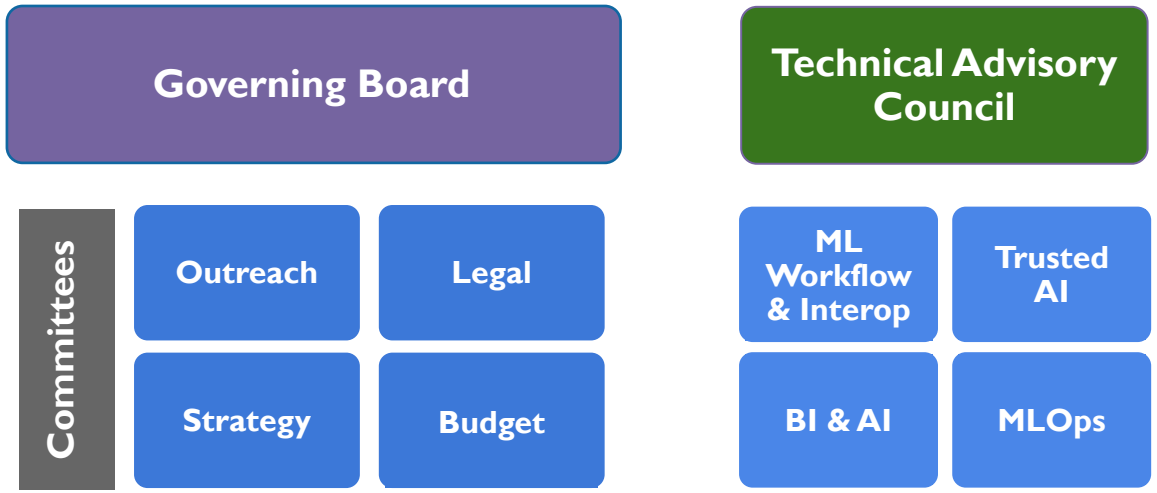
Foundation Governance

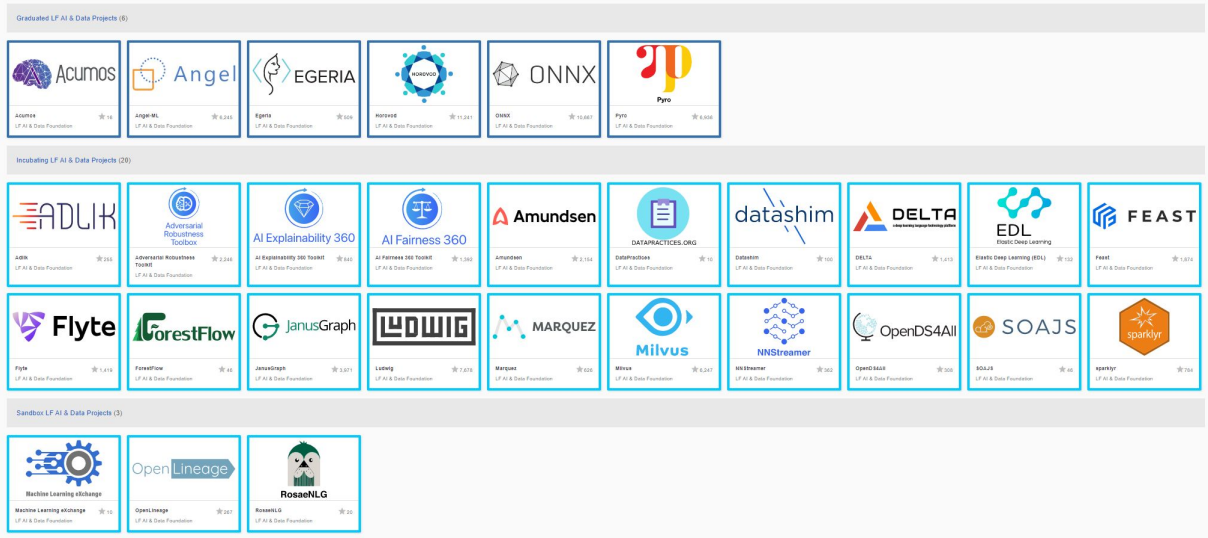
Funding effort to support hosted technical projects

Technical Coordination

Hosted Projects

Projects have their own independent technical governance





Members

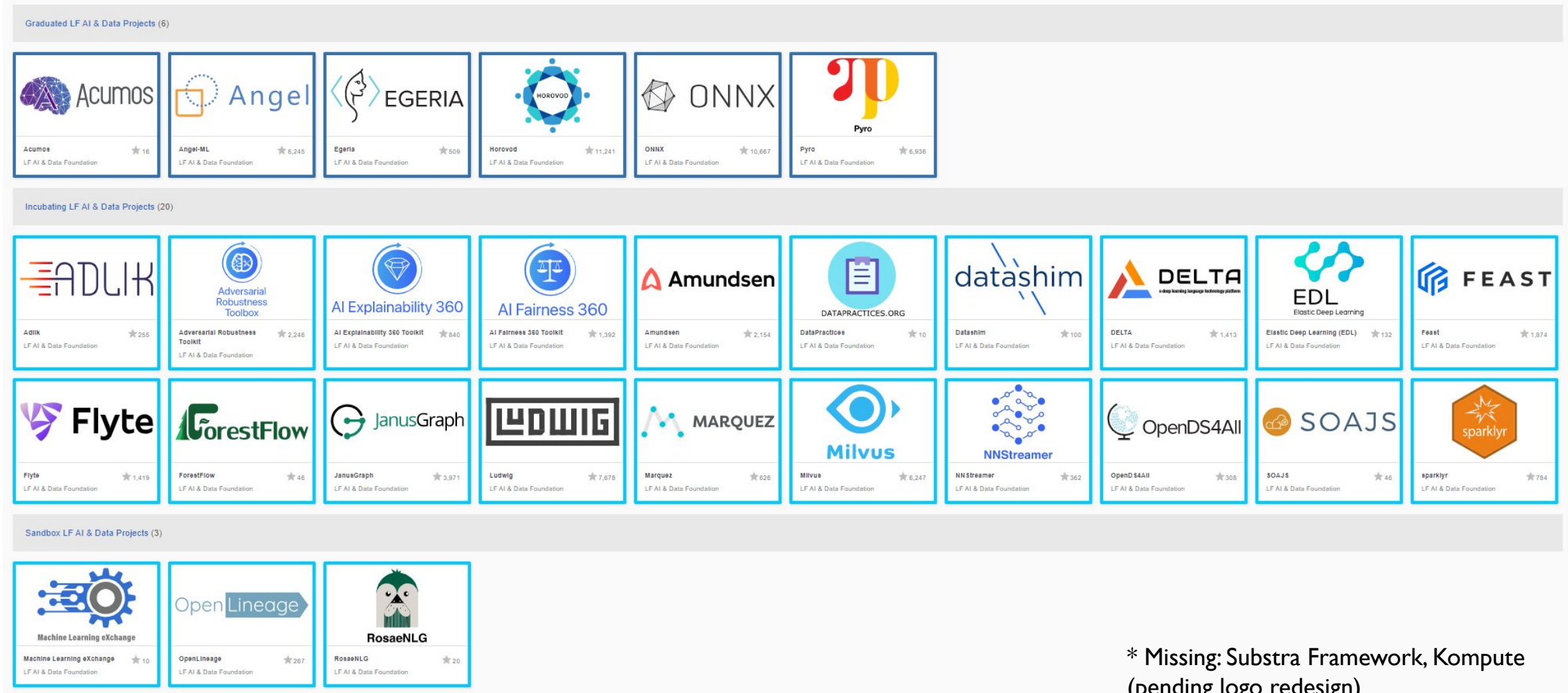
https://landscape.lfai.foundation/card-mode?project=company

High Participation by Academia and Other Member Foundations

- Visibility
- Practical experience working with open source foundations
- Internships opportunities
- Collaboration opportunities
- Publication opportunities
- Open source working experience
- Access to events

Projects (31)

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



* Missing: Substra Framework, Kompute (pending logo redesign)

Major Ongoing Efforts

Develop an ecosystem of developers, users, expertise adopting and relying on the same projects

Enable closer collaboration with the support of a single Technical Advisory Council across all technical efforts

Provide unified guidance to end users on open source software solutions, tools, interoperability, integration, and standards

Increase efforts to harmonize AI/Data open source projects, reduce fragmentation;
Rally the industry to adopt Foundation hosted projects

Provide a funding model for key projects; Increase awareness via a number of marketing and outreach activities;

Trusted AI
Become the home of open source trusted and responsible AI: Policies, guidelines, tools and use cases (by industry) to ensure the development of trustworthy AI systems and processes to develop and improve them over time

MLOps
Get community exposed to how MLOps tools work together and where to use in the pipeline;
Create template architecture for managing ML project lifecycle;
Provide R&D collaboration opps;
Fill in the gaps in use case implementation

ML Workflow & Interop
Improve interoperability and integration capabilities across hosted and external projects

Very active and growing developer community

Jan 1- Dec 31, 2020

8.92K Contributors	217K Contributions	100.03K Commits	358 Repositories
6.93K Emails	34.65K PRs/Changesets	2.08K Slack messages	22.54K Total issues

LF AI & DATA

Source: https://insights.lfx.linuxfoundation.org/projects/lfai

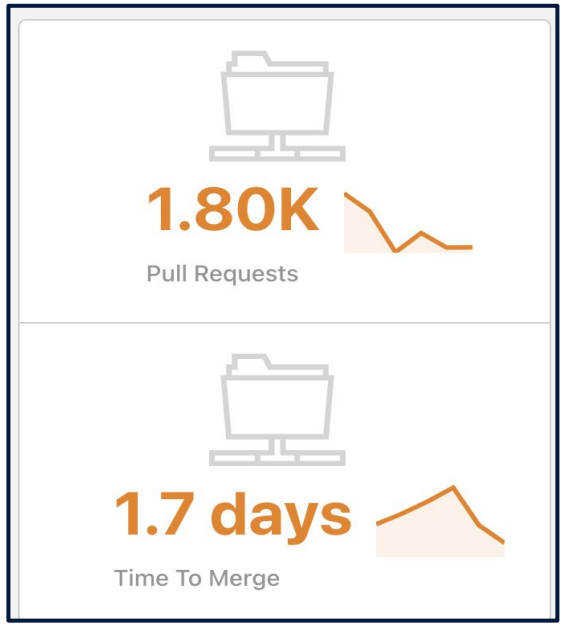
Active and growing developer community

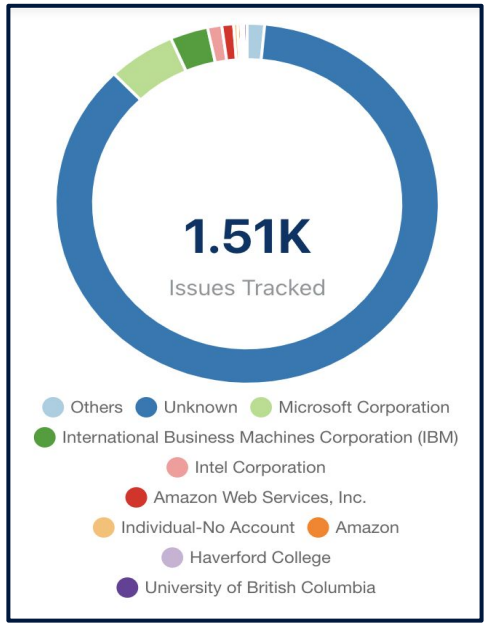
Cumm. Jan 1- Dec 31, 2020 vs. Jan 1, 2020 to Apr 2, 2021

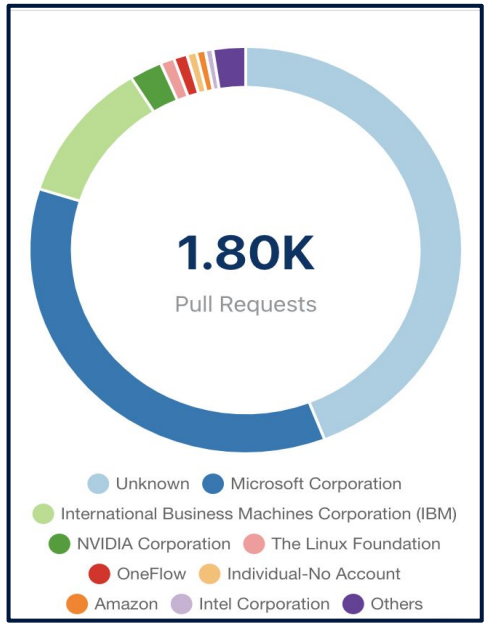
8.92K Contributors	**34.65K** PRs/Changesets	**100.03K** Commits	**22.54K** Total issues	**358** Repositories	**2.08K** Slack messages
+7.25%	+7.21%	+11.03%	+5.19%	+3.07%	+68.75%
9.61K Contributors	**37.15K** PRs/Changesets	**111.07K** Commits	**23.71K** Total issues	**369** Repositories	**3.51K** Slack messages

ONNX Dev Stats - Jan to Dec 2020

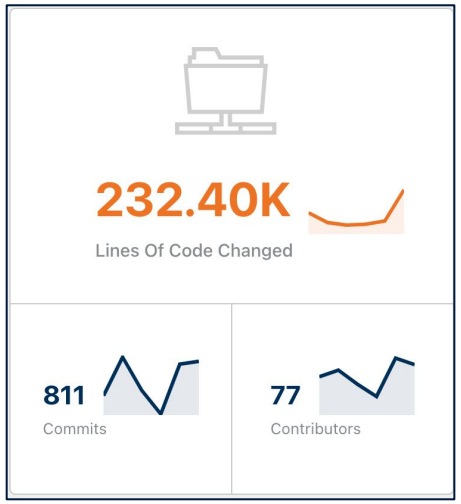
Open standard for machine learning interoperability

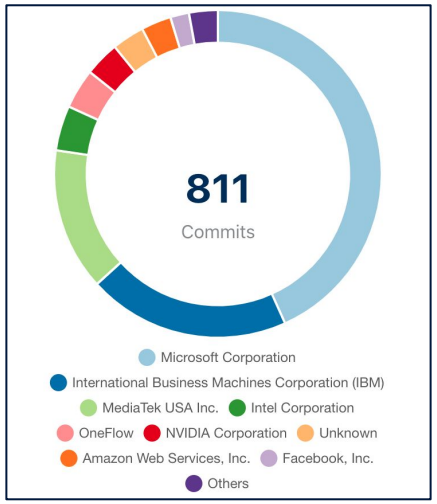






Jan 1, May 10

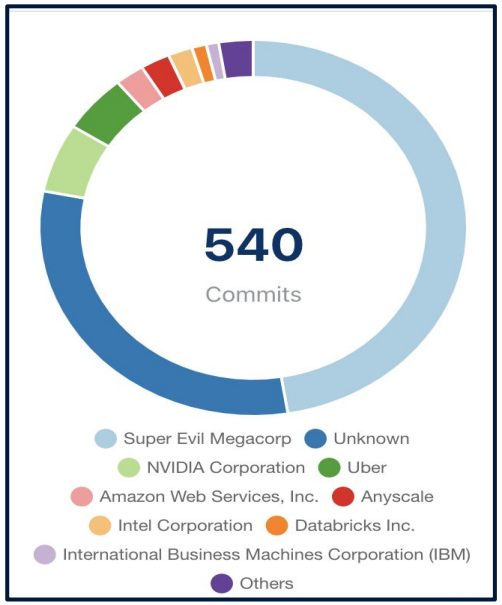


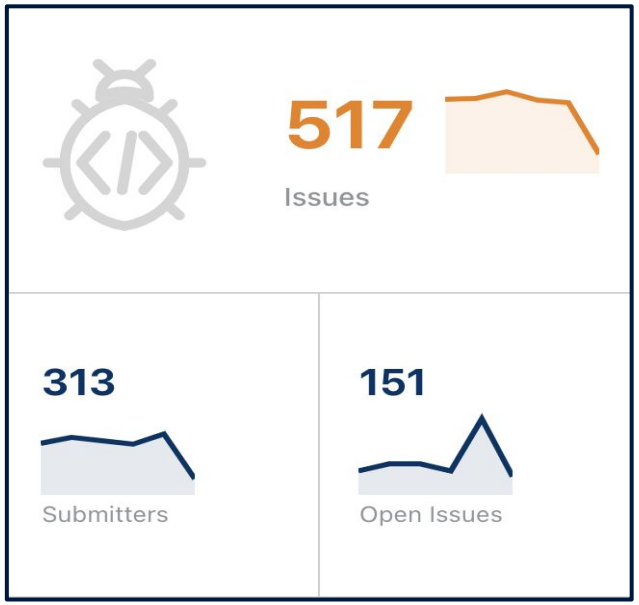


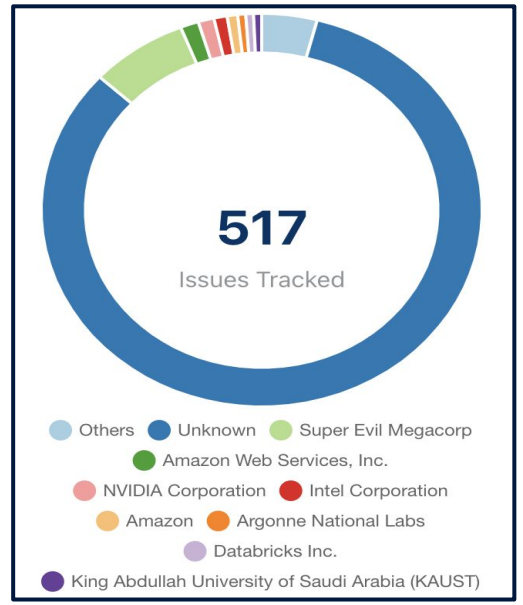
LF AI & DATA

Horovod - Dev Stats - Jan to Dec 2020

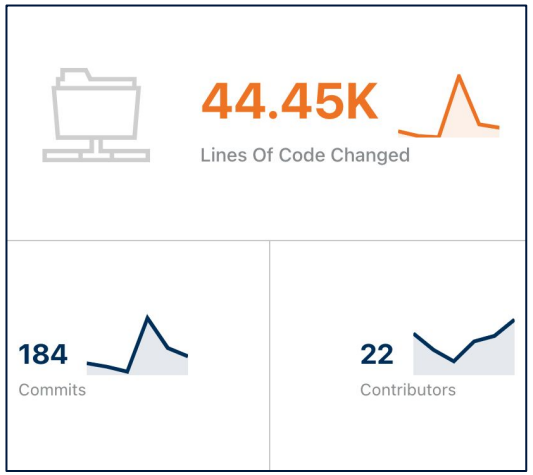
Distributed training framework for TensorFlow, Keras and PyTorch

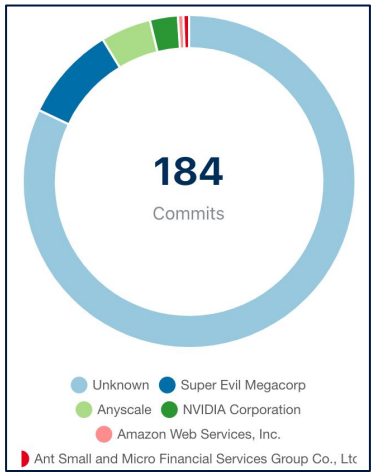




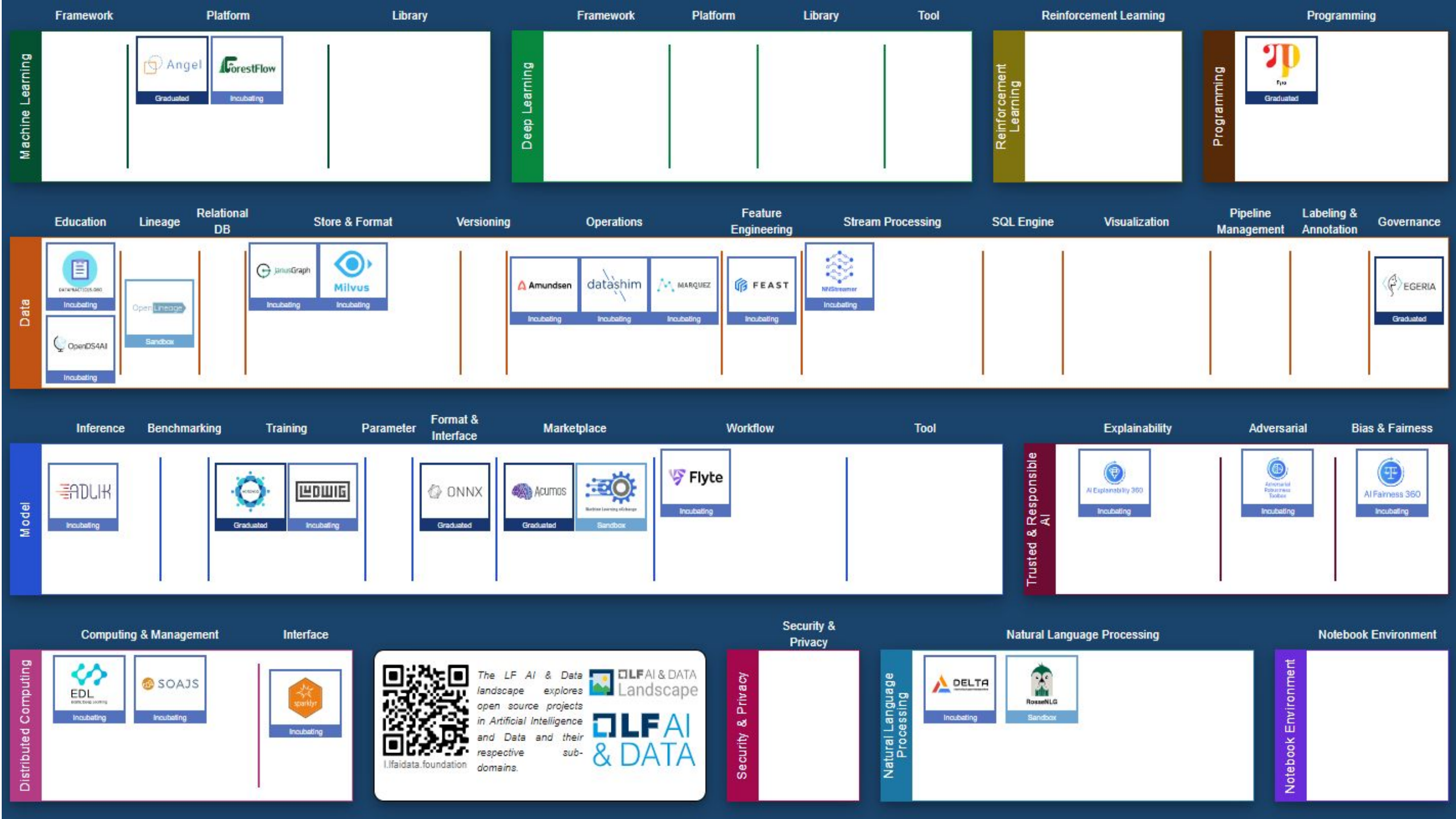


Jan 1, May 10

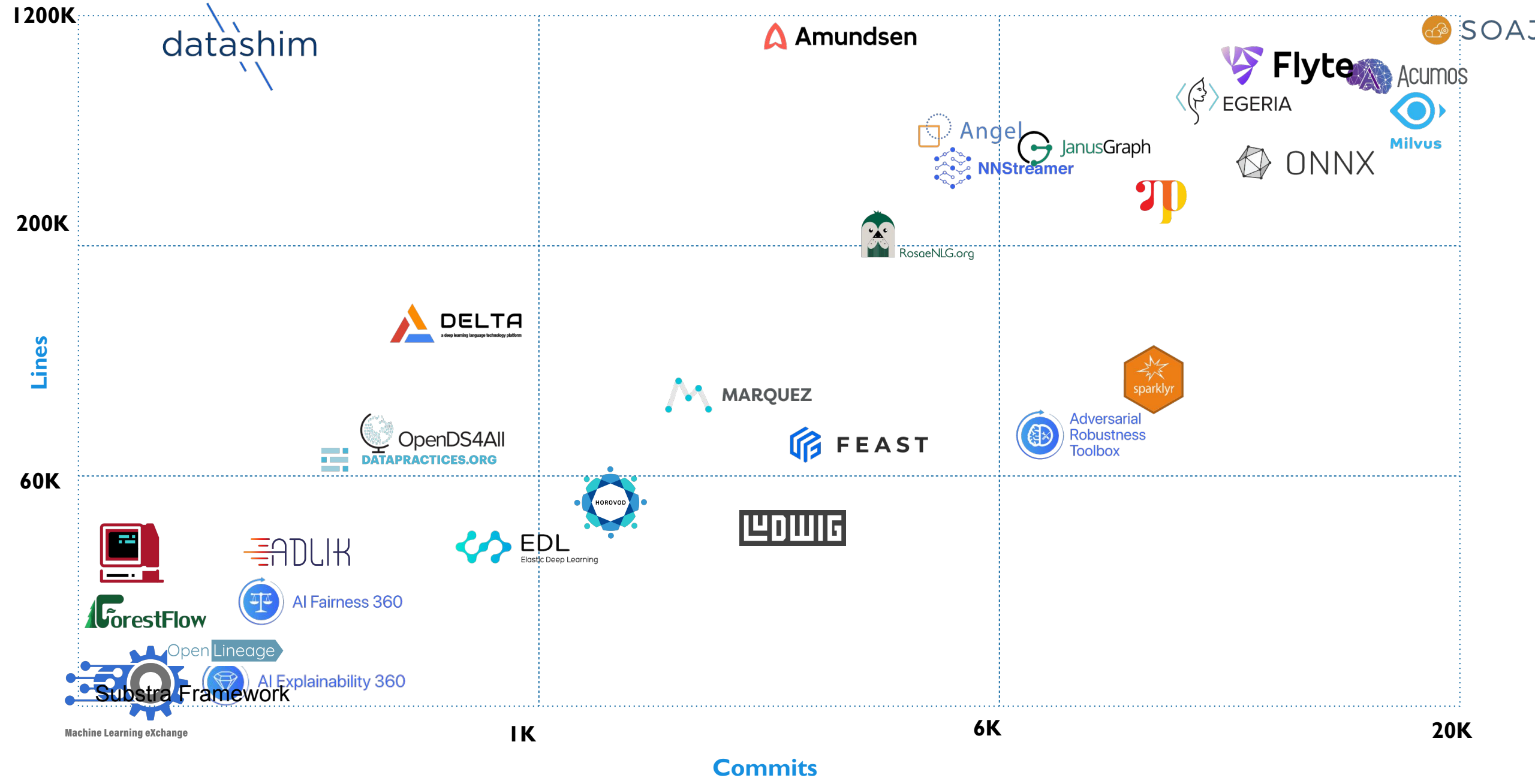




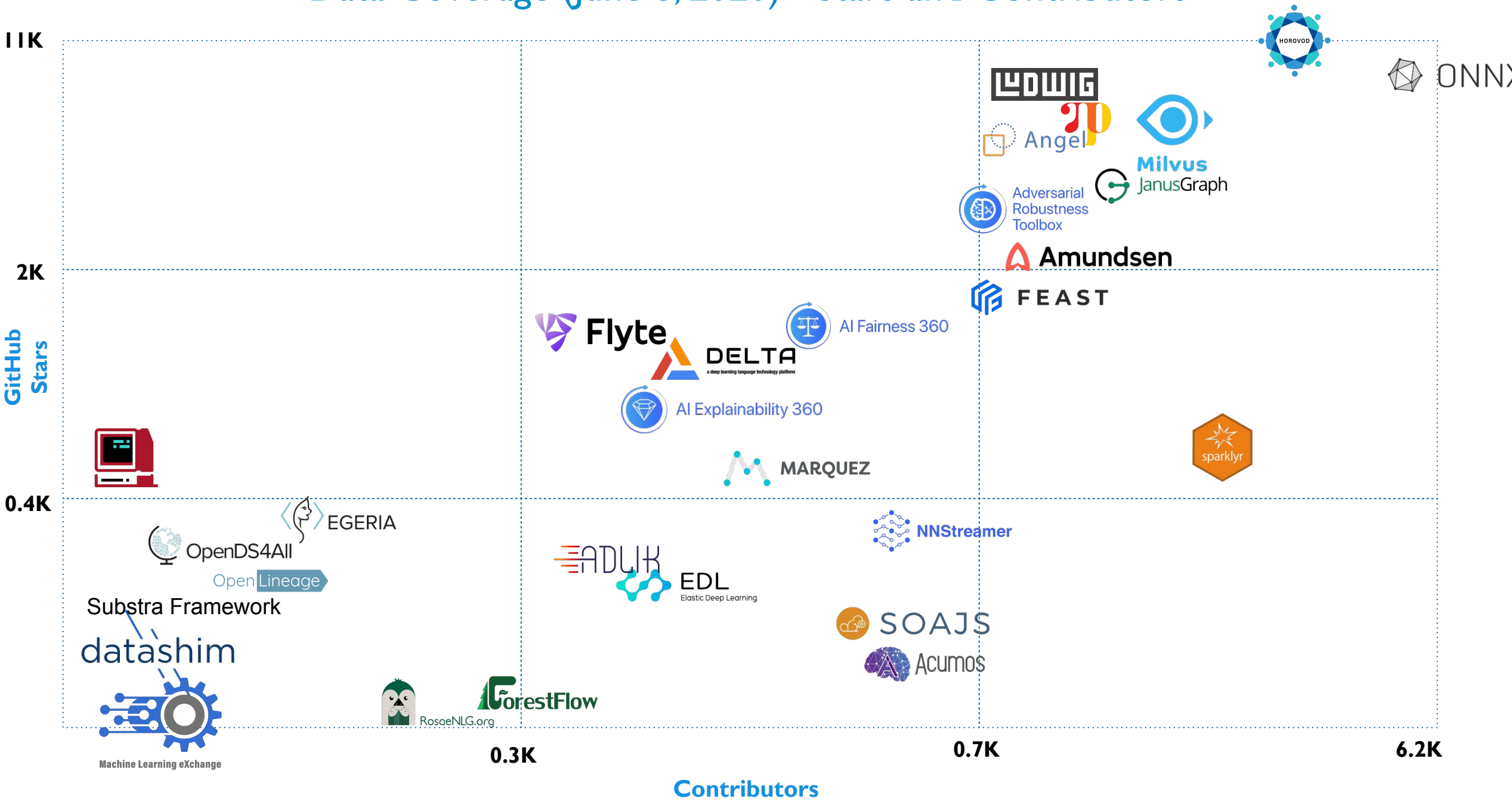
LF AI & DATA


Framework
Platform
Library
Machine Learning
Angel
Graduated
ForestFlow
Incubating
Framework
Platform
Library
Tool
Deep Learning
Reinforcement Learning
Reinforcement Learning
Programming
Programming
Graduated
Education
Lineage
Relational DB
Store & Format
Versioning
Operations
Feature Engineering
Stream Processing
SQL Engine
Visualization
Pipeline Management
Labeling & Annotation
Governance
Data
Incubating
OpenDS4AI
Incubating
OpenLineage
Sandbox
JanusGraph
Incubating
Milvus
Incubating
Amundsen
Incubating
datashim
Incubating
MARQUEZ
Incubating
FEAST
Incubating
Incubating
EGERIA
Graduated
Inference
Benchmarking
Training
Parameter
Format & Interface
Marketplace
Workflow
Tool
Model
ADLIK
Incubating
Graduated
Incubating
ONNX
Graduated
Acumos
Graduated
Sandbox
Flyte
Incubating
Explainability
Adversarial
Bias & Fairness
Trusted & Responsible AI
AI Explainability 360
Incubating
Incubating
AI Fairness 360
Incubating
Computing & Management
Interface
Distributed Computing
EDL
Incubating
SOAJS
Incubating
Incubating
l.lfaidata.foundation
The LF AI & Data landscape explores open source projects in Artificial Intelligence and Data and their respective sub-domains.
LF AI & DATA Landscape
LF AI & DATA
Security & Privacy
Security & Privacy
Natural Language Processing
Natural Language Processing
DELTA
Incubating
RosaeNLG
Sandbox
Notebook Environment
Notebook Environment

Data Coverage (June 8, 2021) - Line of Code and Commits



Data Coverage (June 8, 2021) - Stars and Contributors



Companies hosting projects in LF AI & Data

https://landscape.lfai.foundation/format=hosting

Why incubate with us?

Building large, sustainable ecosystems requires collective resources

Events - We gathered over 45,000 attendees from over 12,000 organizations across 113 countries in 2019.

Legal - We manage IP for the world's most important tech and have one of the world's top open source legal teams in house.

Training - We trained millions of students (free and paid online), online skills certification, and on-site e-learning.

Certification - We designed and executed software and hardware testing and certification programs.

Developer Marketing - We have the largest share of voice of any open source foundation and a proven method to build large scale developer programs.

Developer Operations - We host the infrastructure that develops the world's largest software communities and provide release management, IT ops, and support.

Application Security- our projects are regularly audited and pen tested. We offer bug bounties, dependency analysis, and code scanning.

What does it mean to be an LF AI & Data project?

- **A neutral home for the project** increases the willingness of developers from organizations to collaborate, contribute, and become committers
- **Endorsement & mentorship by members of the LF AI & Data Technical Advisory Committee (TAC)**
- **Open and neutral governance model**
- **Program and project management services**
- **Access to full-time staff (ED, PM, Tech)**
- **Access to marketing and communications services** to support community and ecosystem engagement
- **Access to a full range of events services** to help the project build and grow a community around it
- **Legal support** ranging from hosting the trademark in a neutral organization to CLA and DCO system integration with GitHub, regular compliance scans, export control filings, etc.
- **Presence across multiple geos**

Incubating Projects

- TAC committee is responsible to voting in new incubation projects.
- The TAC meets bi-weekly and generally booked 4-6 weeks in advance.
- Proposing projects for incubation

Strong Developer/Community Marketing Effort

LF AI provides marketing services to support community and ecosystem engagement.

37 announcements in 2019

61 announcements in 2020

30 announcements in 2021 (Jan - May)

- Events announcements and promotion
- Project release
- Project incubation
- Project cross collaboration

LF AI provides event and community ecosystem building services.

LF AI & Data Summit

LF AI & Data Day

Project Workshops

Joining LF AI & Data is Easy!

General Inquiries:

info@lfaidata.foundation

LF AI & DATA