LF AI & Data: Accelerating Development and Innovation

September 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

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

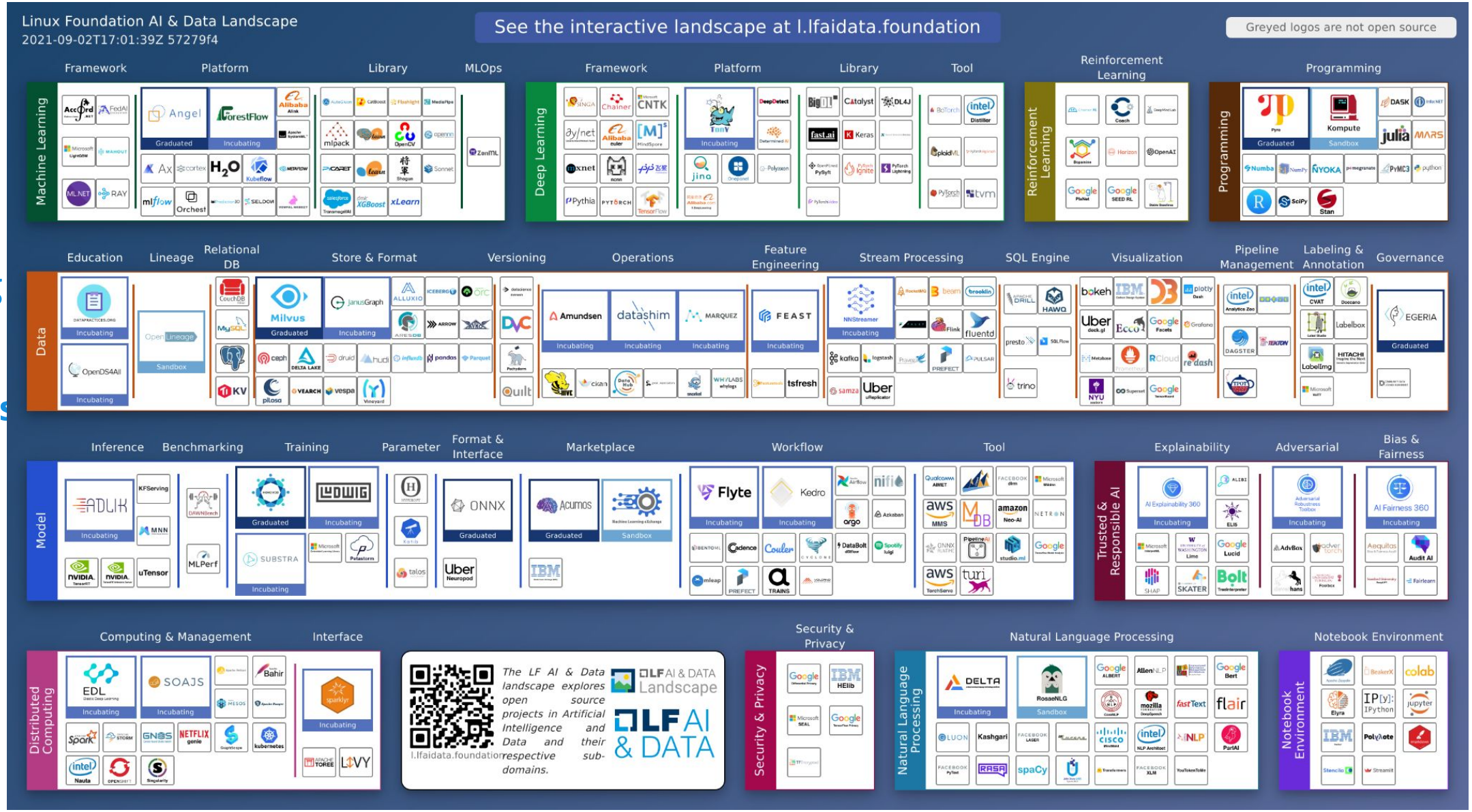
140+ founding orgs

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 & 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 licensed under an open source data specific license making it freely available for use without mechanisms of control

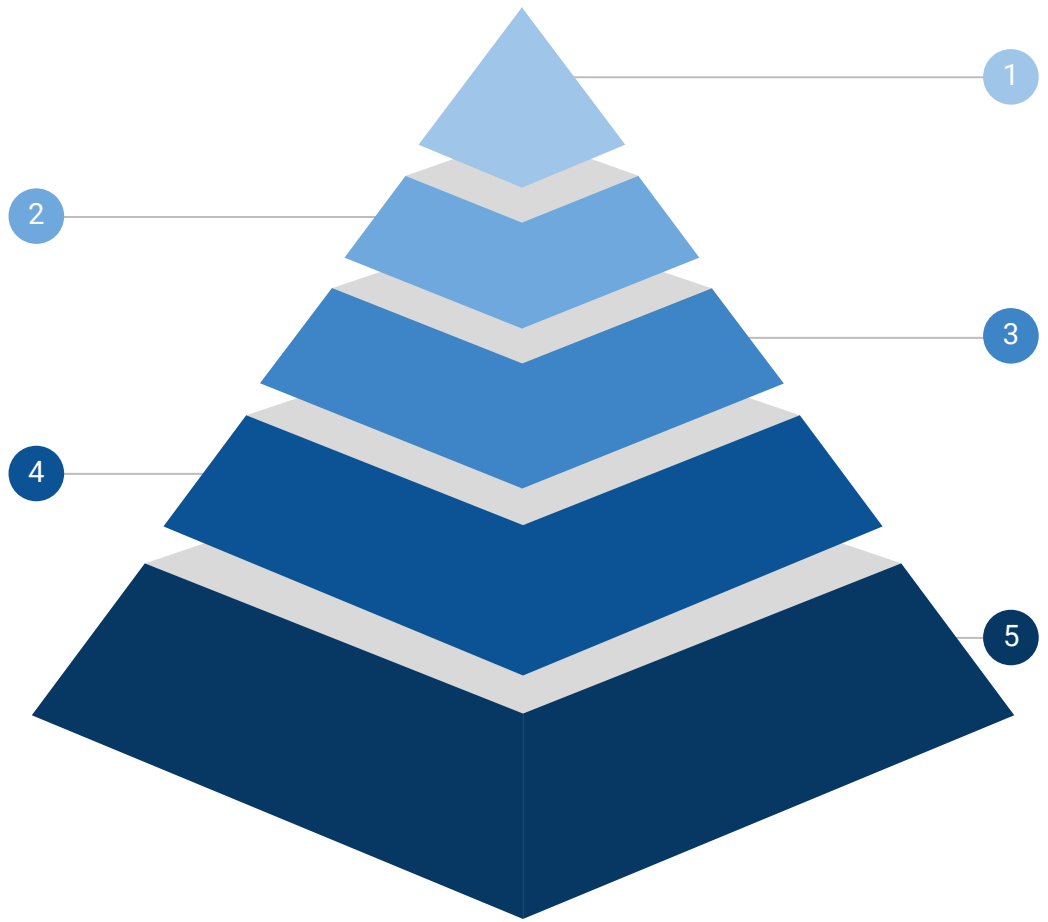
Data Manipulation Tools

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
- Accelerating innovation and development in open source AI & Data
- Enabling 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
- +9 projects (Jan - July) 2022
- Supported by 49 member companies (all geos)
- Focused on building an ecosystem around high value and critical projects, accelerate the adoption and ensure sustainability

Structure and Governance

Foundation Governance

Funding effort

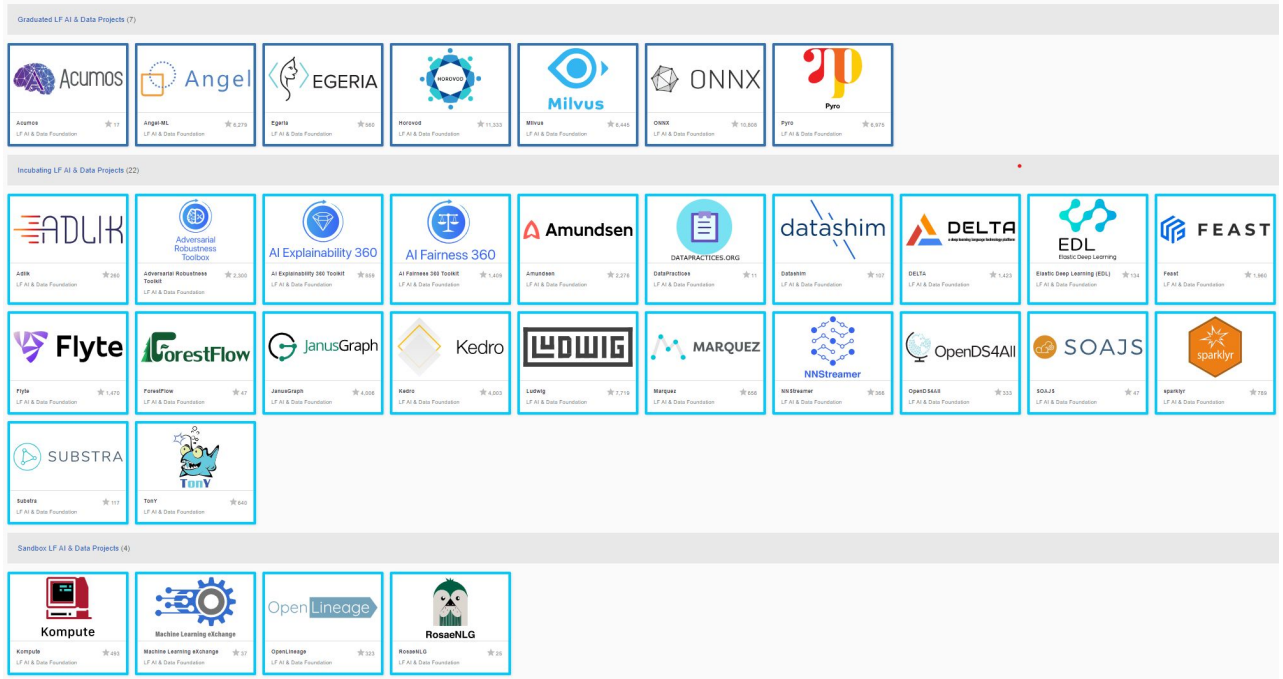
- Governing Board
- Outreach
- Legal
- Strategy
- Budget

Technical Coordination

- Technical Advisory Council
- ML Workflow & Interop
- Trusted AI
- BI & AI
- MLOps

Hosted Projects

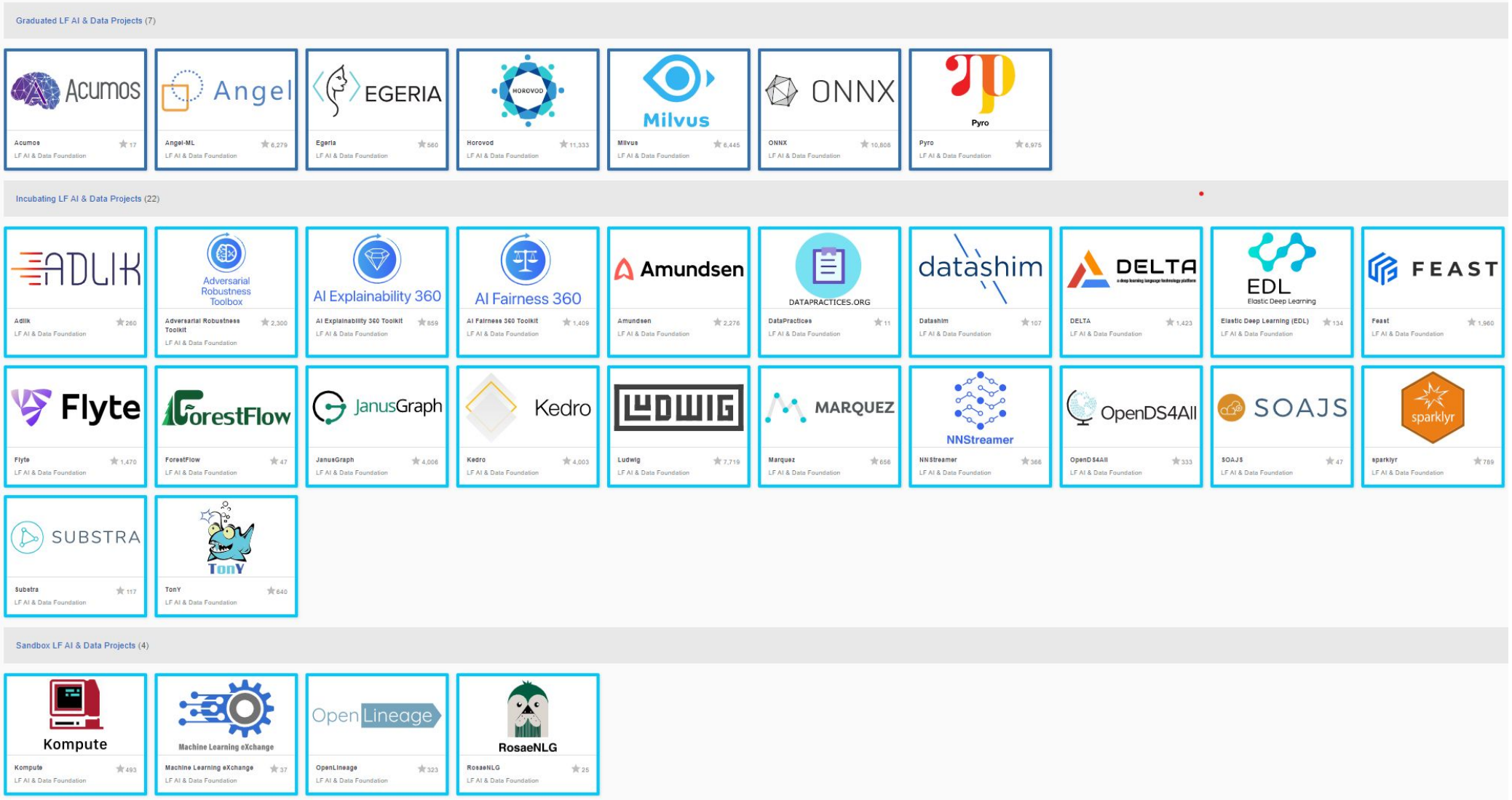
Projects have their own independent technical governance

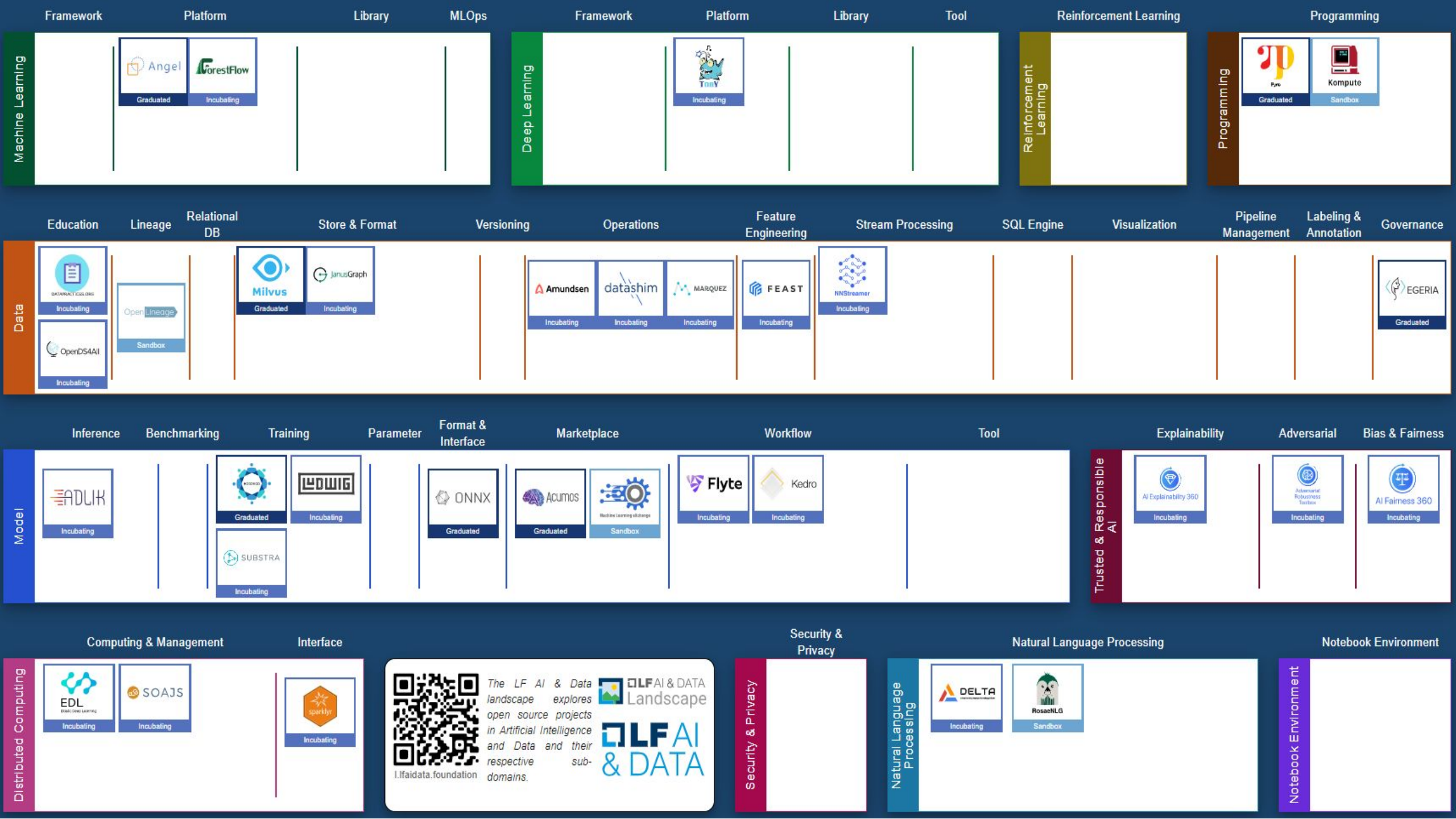


Members

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

Projects


Graduated LF AI & Data Projects (7)
Acumos
Angel
EGERIA
HOROVOD
Milvus
ONNX
Pyro
Incubating LF AI & Data Projects (22)
ADLIK
Adversarial Robustness Toolbox
AI Explainability 360
AI Fairness 360
Amundsen
DATAPRACTICES.ORG
datashim
DELTA
EDL
Elastic Deep Learning
FEAST
Flyte
ForestFlow
JanusGraph
Kedro
LUDWIG
MARQUEZ
NNStreamer
OpenDS4All
SOAJS
sparklyr
SUBSTRA
TonY
Sandbox LF AI & Data Projects (4)
Kompute
Machine Learning eXchange
OpenLineage
RosaeNLG


Machine Learning
Framework
Platform
Library
MLOps
Angel
Graduated
ForestFlow
Incubating
Deep Learning
Framework
Platform
Library
Tool
Tony
Incubating
Reinforcement Learning
Reinforcement Learning
Programming
Programming
Pyro
Graduated
Kompute
Sandbox
Data
Education
Lineage
Relational DB
Store & Format
Versioning
Operations
Feature Engineering
Stream Processing
SQL Engine
Visualization
Pipeline Management
Labeling & Annotation
Governance
Incubating
OpenDS4All
Incubating
OpenLineage
Sandbox
Milvus
Graduated
JanusGraph
Incubating
Amundsen
Incubating
datashim
Incubating
MARQUEZ
Incubating
FEAST
Incubating
NNStreamer
Incubating
EGERIA
Graduated
Model
Inference
Benchmarking
Training
Parameter
Format & Interface
Marketplace
Workflow
Tool
ADLIK
Incubating
Graduated
LUDWIG
Incubating
SUBSTRA
Incubating
ONNX
Graduated
Acumos
Graduated
Sandbox
Flyte
Incubating
Kedro
Incubating
Trusted & Responsible AI
Explainability
Adversarial
Bias & Fairness
AI Explainability 360
Incubating
Adversarial Robustness Toolbox
Incubating
AI Fairness 360
Incubating
Distributed Computing
Computing & Management
Interface
EDL
Incubating
SOAJS
Incubating
sparklyr
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

Major Ongoing Efforts

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

Enable 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

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

Provide a funding model for 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

Active and growing developer community

Jan 1- Dec 31, 2020

8.92K Contributors	**100.03K** Commits	**358** Repositories
6.93K Emails	**34.65K** PRs/Changesets	**22.54K** Total issues

LF AI & DATA

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

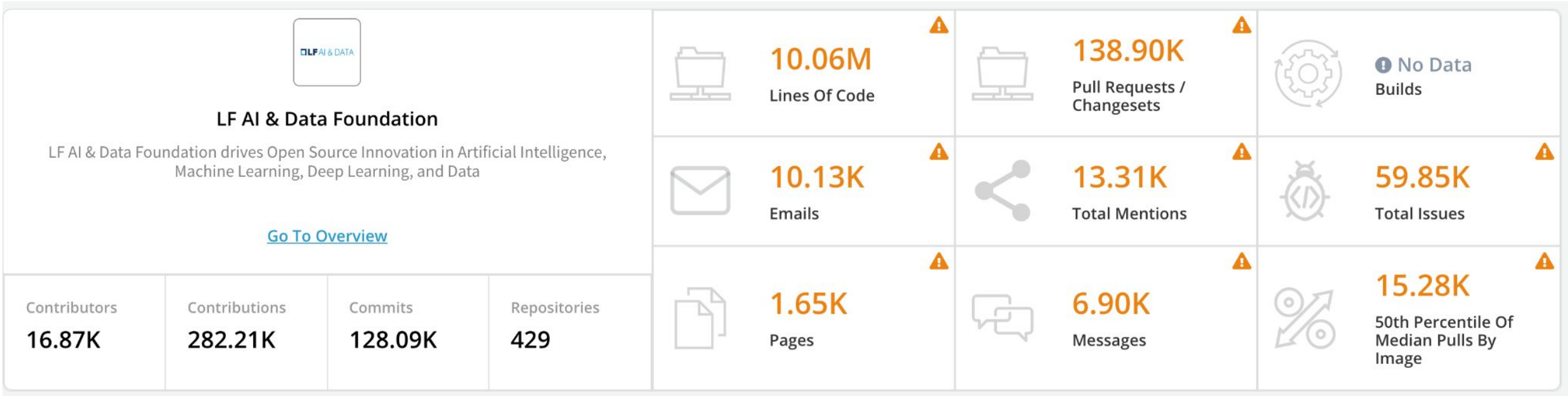
Active and growing developer community

Cumm. Jan 1- Dec 31, 2020 vs. Jan 1, 2021 to June 7, 2021

8.92K	34.65K	100.03K	22.54K	358
Contributors	PRs/Changesets	Commits	Total issues	Repositories

9.66K	37.15K	128.01K	23.71K	420
Contributors	PRs/Changesets	Commits	Total issues	Repositories

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



LF AI & DATA

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

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.

AI & Data Track, Open Source Summit

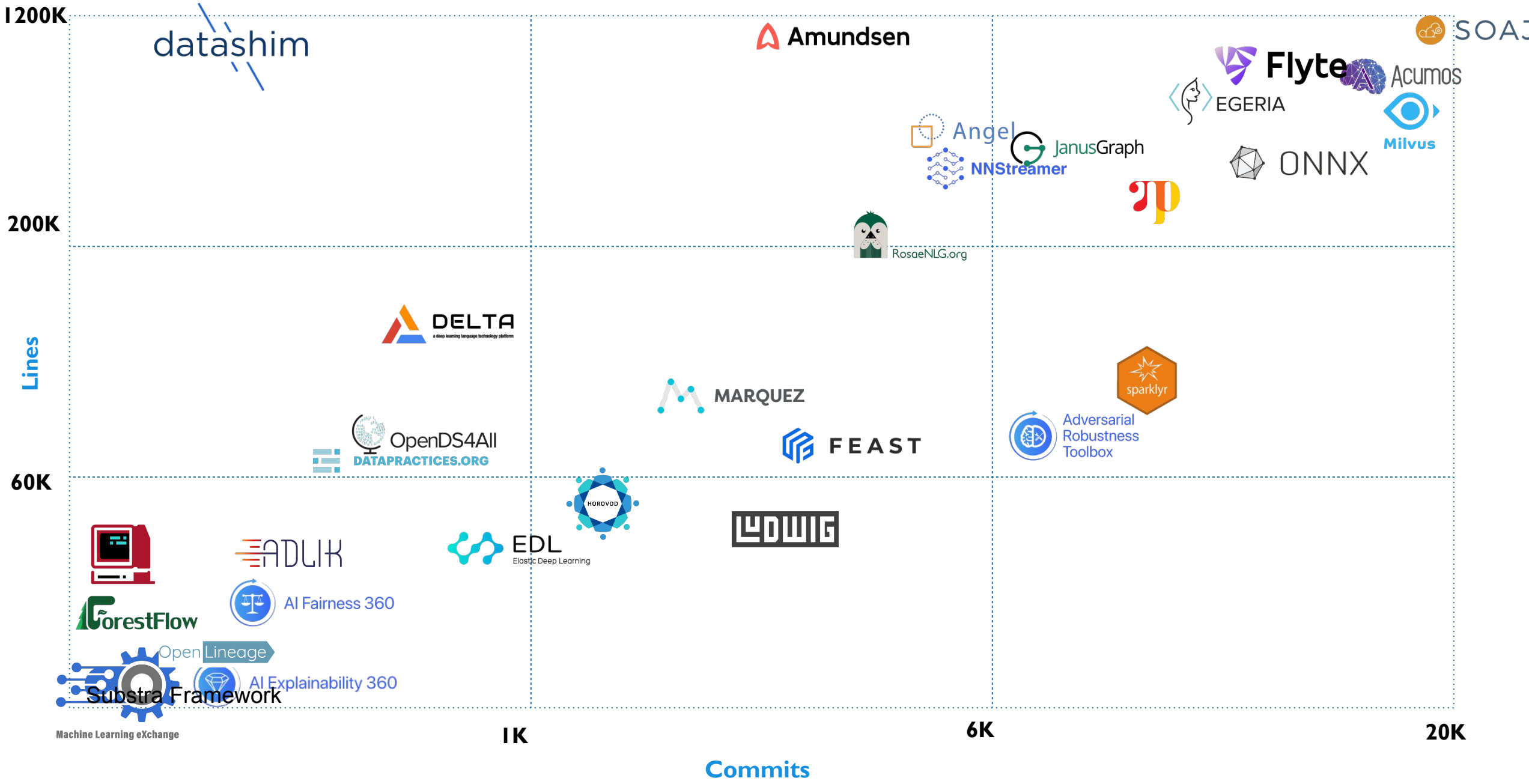
AI Track, Open Source Summit

Kubernetes + AI

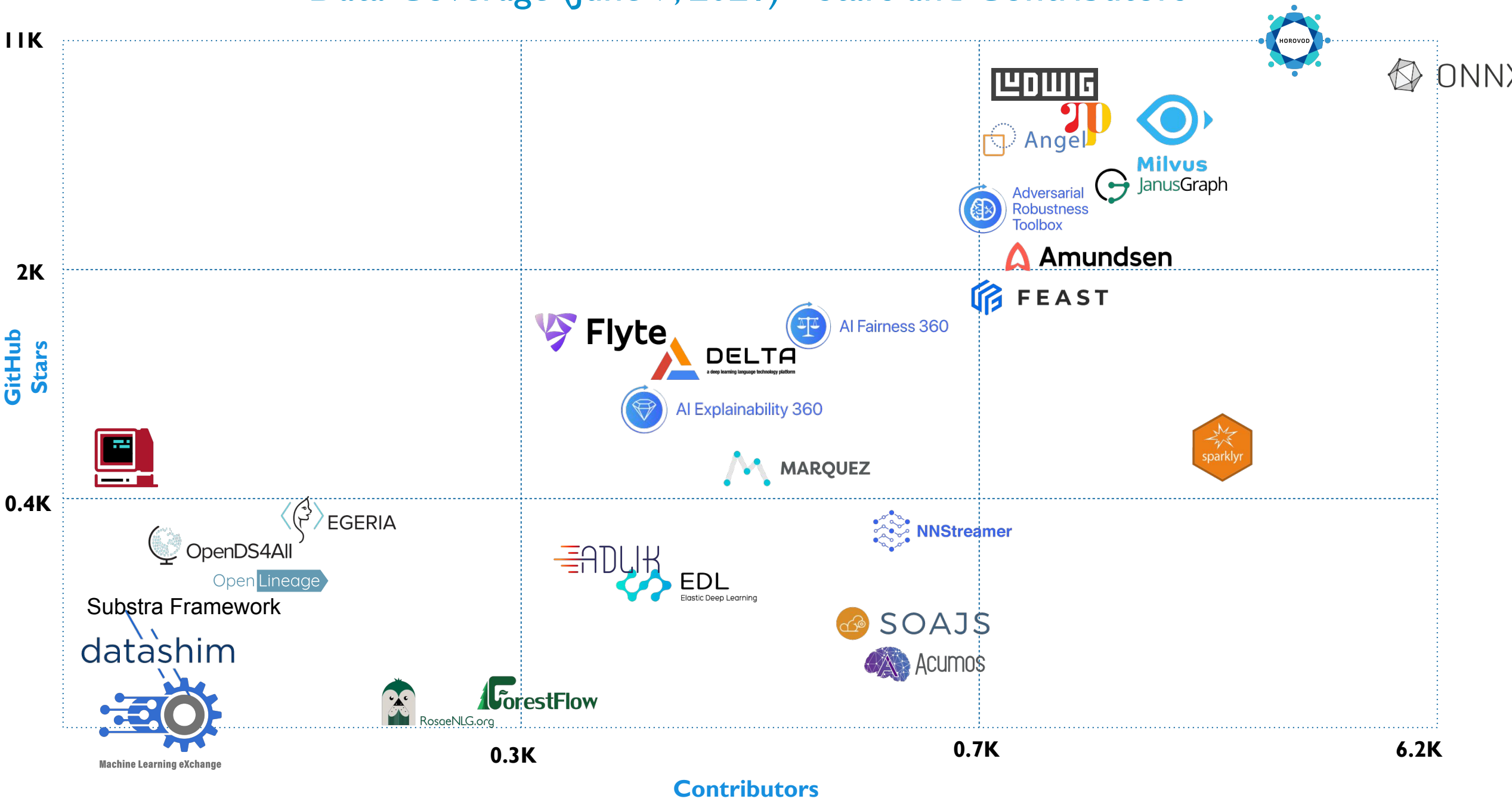
Open Networking & Edge Summit

LF AI & Data Day (multiples/year

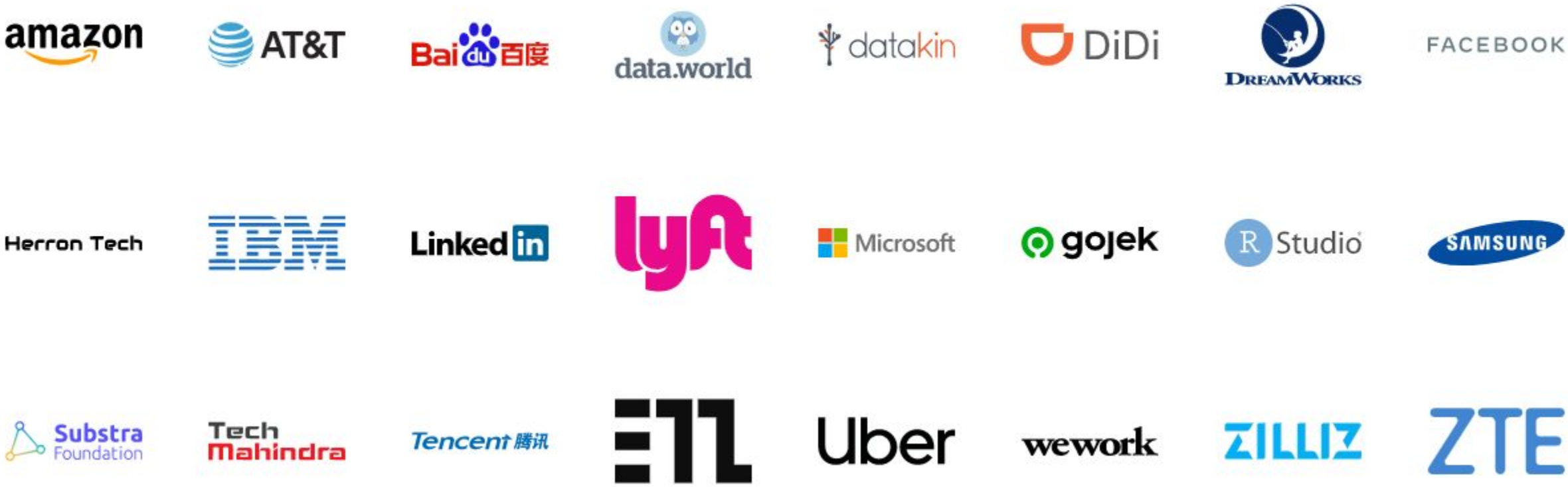
Data Coverage (June 7, 2021) - Line of Code and Commits



Data Coverage (June 7, 2021) - Stars and Contributors



Companies hosting projects in LF AI & Data



Why incubate with us?

Building large, sustainable ecosystems requires collective resources

Neutral Hosting with technical open and neutral governance model

Dedicated staff - Access to full-time staff (ED, PM, Tech) with program and project management services

Training & Certification - We build training classes, design and execute and testing and certification programs in support of hosted projects

Events team - Access to a full range of events services to help the project build and grow a community around it

Developer Operations - LFX platform, IT infrastructure, release management, IT ops, support, security audits, etc.

Endorsement & mentorship by members of the LF AI & Data Technical Advisory Committee

Marketing and communications services to support community and ecosystem engagement

Legal support - including hosting the trademark, CLA and DCO integration with GitHub, compliance scans, export control filings, etc.

Organizations join LF AI & Data to take an active role in supporting the growth and evolution of the Open Source AI & Data ecosystem

Support our mission, programs, and the community by helping fund services that hosted projects rely on

Demonstrate thought leadership by participating in a wide reaching networking and marketing programs

Be part of defining and maintaining the technologies that are at the forefront of the industry

Host projects with us and benefit from the services & support we offer our hosted projects to increase their adoption and footprint in the ecosystem

Provide technical leadership to the open source AI and Data community via the TAC, it various sub-committees and efforts

Receive greater insight into the foundation's strategy, efforts, projects and initiatives through increased engagement with the ED, staff and committees' leads

Join us!

Incubate your project with LF AI & Data

info@lfaidata.foundation

LF AI & DATA

High Participation by Academia and Other Foundations

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