

LF AI & Data

June 24, 2021

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

 LF AI & DATA

Ecosystem Overview

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

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.

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

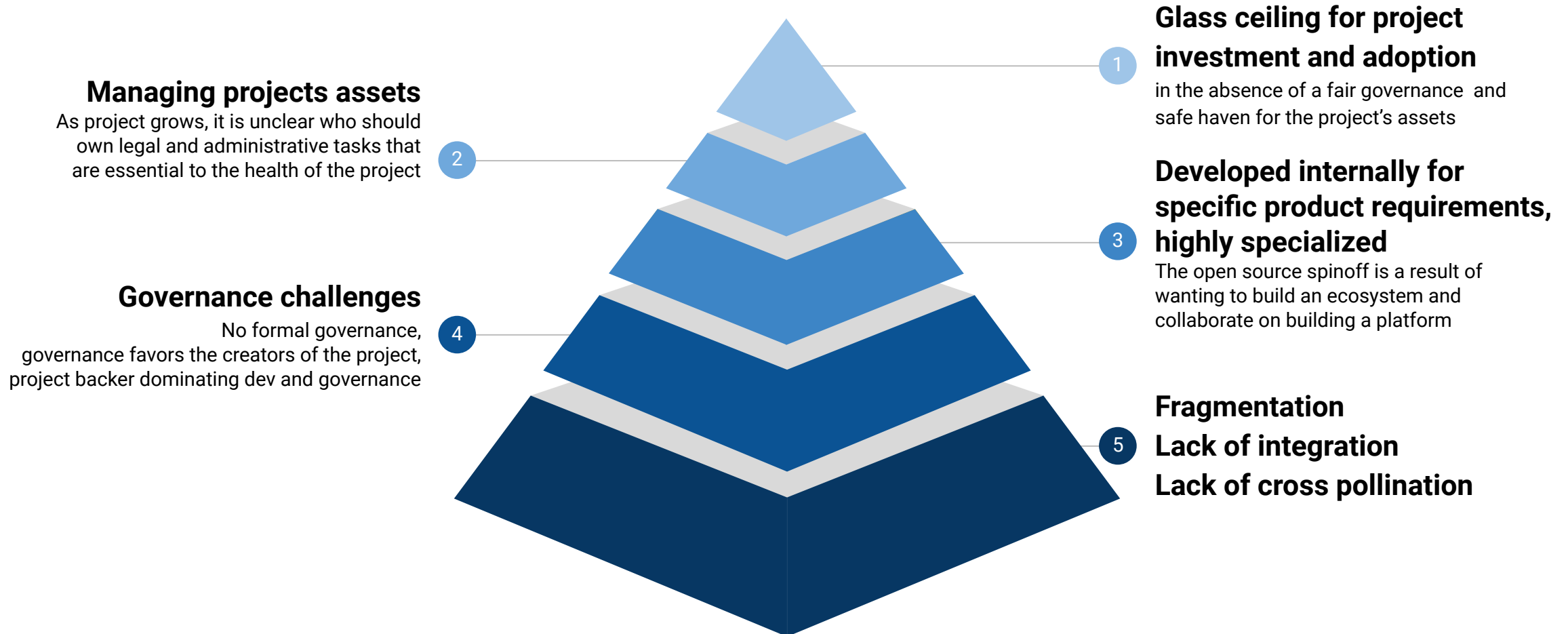
Open Data

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



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
- › +8 projects (Jan - May) 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

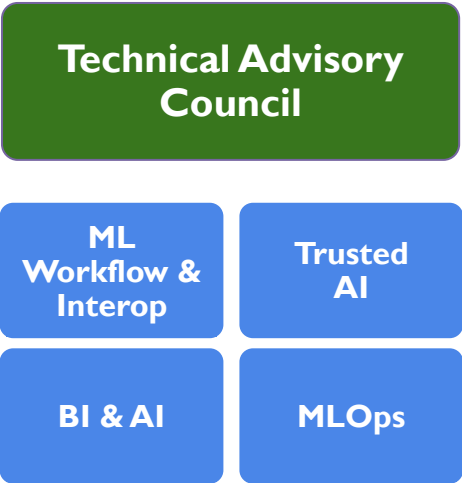
Structure and Governance

Foundation Governance

Funding effort to support hosted technical projects



Technical Coordination



Hosted Projects

Projects have their own independent technical governance

The screenshot shows the LF AI & Data Projects page, categorized into three sections: Graduated, Incubating, and Sandbox. Each project card includes a logo, name, and a star rating.

Project Name	Star Rating
Acumos	11
Angel ML	9,275
EGERIA	916
Horovod	11,335
Milvus	9,445
ONNX	10,884
Pyras	9,975

Project Name	Star Rating
ADLIR	389
Adversarial Robustness Toolkit	2,300
AI Explainability 360	369
AI Fairness 360	1,409
Amundsen	2,279
DATA PRACTICES.ORG	11
datashim	107
DELTA	1,423
Elastic Deep Learning (EDL)	134
FEAST	1,360
Flyte	1,475
ForestFlow	47
JanusGraph	4,008
LUDWIG	7,719
MARQUEZ	436
NNStreamer	368
OpenDS4All	333
SOAJS	97
sparklyr	709
SUBSTRA	117

Project Name	Star Rating
Machine Learning Exchange	17
OpenLineage	123
RosaeNLG	25

Members

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





















Premier	 AT&T	 BAIDU AI CLOUD	 ERICSSON	 oppo	 HUAWEI	 IBM		
	 NOKIA	 sas	 Tech Mahindra	 Tencent 腾讯	 ZILLIZ Reinvent Data Science	 ZTE		
General	 BROADCOM	 CLOUDERA	 D2 IQ	 databricks	 FS TECHNOLOGY	 Herron Tech	 index analytics	 ING
	 inwinSTACK	 orange	 precisely	 Red Hat	 R Studio	 vmware	 XENONSTACK	 DiDi
Associate	 AI for people	 aivancity PARIS-CACHAN	 AMBIANT	 BANQUE DE FRANCE EUROSYSTEME	 UNIVERSITY INSTITUTE OF TECHNOLOGY	 GALGOTIAS UNIVERSITY	 HIS	
	 The International Society of Service Innovation Professionals	 RICE UNIVERSITY	 MAIEI	 NYU	 Open 启智 新一代人工智能开源开放平台	 OpenUK	 PSIT Kanpur	
	 SAHYADRI UNIVERSITY OF SCIENCE & TECHNOLOGY	 shopen 上海市开放大学	 EIL	 UNIVERSITY OF WASHINGTON	 ULB Université Libre de Bruxelles	 PRIZE		

Projects

Graduated LF AI & Data Projects (7)
























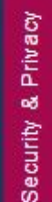

 <p>Acumos LF AI & Data Foundation ★ 17</p>	 <p>Angel Angel-ML LF AI & Data Foundation ★ 8,279</p>	 <p>EGERIA Egeria LF AI & Data Foundation ★ 515</p>	 <p>Horovod LF AI & Data Foundation ★ 11,333</p>	 <p>Milvus LF AI & Data Foundation ★ 8,445</p>	 <p>ONNX LF AI & Data Foundation ★ 10,808</p>	 <p>Pyro LF AI & Data Foundation ★ 6,975</p>
---	--	---	---	--	---	--

Incubating LF AI & Data Projects (20)

 <p>ADLIK Adlik LF AI & Data Foundation ★ 260</p>	 <p>Adversarial Robustness Toolkit Adversarial Robustness Toolkit LF AI & Data Foundation ★ 2,300</p>	 <p>AI Explainability 360 AI Explainability 360 Toolkit LF AI & Data Foundation ★ 859</p>	 <p>AI Fairness 360 AI Fairness 360 Toolkit LF AI & Data Foundation ★ 1,409</p>	 <p>Amundsen Amundsen LF AI & Data Foundation ★ 2,276</p>	 <p>DataPractices DATAPRACTICES.ORG LF AI & Data Foundation ★ 11</p>	 <p>datashim Datashim LF AI & Data Foundation ★ 107</p>	 <p>DELTA delta learning language technology platform LF AI & Data Foundation ★ 1,423</p>	 <p>EDL Elastic Deep Learning LF AI & Data Foundation ★ 134</p>	 <p>FEAST Feast LF AI & Data Foundation ★ 1,960</p>
 <p>Flyte Flyte LF AI & Data Foundation ★ 1,470</p>	 <p>ForestFlow ForestFlow LF AI & Data Foundation ★ 47</p>	 <p>JanusGraph JanusGraph LF AI & Data Foundation ★ 4,006</p>	 <p>LUDWIG Ludwig LF AI & Data Foundation ★ 7,719</p>	 <p>MARQUEZ Marquez LF AI & Data Foundation ★ 656</p>	 <p>NNStreamer NNStreamer LF AI & Data Foundation ★ 366</p>	 <p>OpenDS4All OpenDS4All LF AI & Data Foundation ★ 333</p>	 <p>SOAJS SOAJS LF AI & Data Foundation ★ 47</p>	 <p>sparklyr sparklyr LF AI & Data Foundation ★ 769</p>	 <p>SUBSTRA Substra LF AI & Data Foundation ★ 117</p>

Sandbox LF AI & Data Projects (3)

 <p>Machine Learning eXchange Machine Learning eXchange LF AI & Data Foundation ★ 37</p>	 <p>OpenLineage OpenLineage LF AI & Data Foundation ★ 323</p>	 <p>RosaeNLG RosaeNLG LF AI & Data Foundation ★ 25</p>
--	---	--

Machine Learning	Framework		Platform		Library		Framework		Platform		Library		Tool		Reinforcement Learning		Programming									
		<div> Graduated</div>	<div> Incubating</div>														<div> Graduated</div>									
Data	Education		Lineage		Relational DB		Store & Format		Versioning		Operations		Feature Engineering		Stream Processing		SQL Engine		Visualization		Pipeline Management		Labeling & Annotation		Governance	
	<div> Incubating</div> <div> Incubating</div>	<div> Sandbox</div>		<div> Graduated</div> <div> Incubating</div>					<div> Incubating</div>	<div> Incubating</div>	<div> Incubating</div>	<div> Incubating</div>	<div> Incubating</div>											<div> Graduated</div>		
Model	Inference		Benchmarking		Training		Parameter		Format & Interface		Marketplace		Workflow		Tool		Explainability		Adversarial		Bias & Fairness					
	<div> Incubating</div>		<div> Graduated</div> <div> Incubating</div> <div> Incubating</div>		<div> Graduated</div>	<div> Graduated</div> <div> Sandbox</div>	<div> Incubating</div>									<div>Trusted & Responsible AI</div>	<div> Incubating</div>		<div> Incubating</div>		<div> Incubating</div>					
Distributed Computing	Computing & Management		Interface		<div> The LF AI & Data landscape explores open source projects in Artificial Intelligence and Data and their respective sub-domains. lfaidata.foundation</div>		<div> DLFAI & DATA</div>		Security & Privacy		Natural Language Processing		Notebook Environment													
	<div> Incubating</div> <div> Incubating</div>	<div> Incubating</div>	<div> Incubating</div> <div> Sandbox</div>																							

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

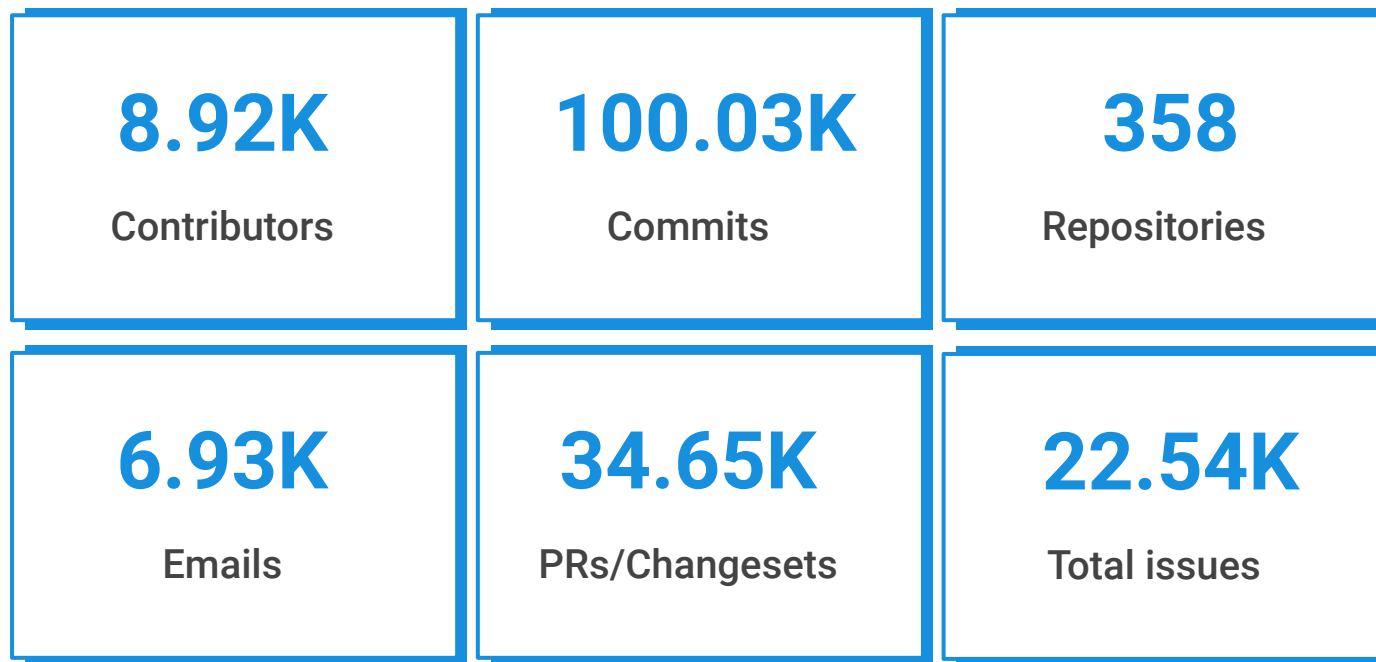
Get community exposed to how MLOps tools work together and where to use in the pipeline; Create template architecture for managing ML project lifecycle; Provide R&D collaboration opps; Fill in the gaps in use case implementation

ML Workflow & Interop

Improve interoperability and integration capabilities across hosted and external projects

Active and growing developer community

Jan 1 - Dec 31, 2020



Active and growing developer community

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

8.92K

Contributors

34.65K

PRs/Changesets

100.03K

Commits

22.54K

Total issues

358

Repositories

9.66K

Contributors

37.15K

PRs/Changesets

128.01K

Commits

23.71K

Total issues

420

Repositories

LF AI & Data Foundation

LF AI & Data Foundation drives Open Source Innovation in Artificial Intelligence, Machine Learning, Deep Learning, and Data

[Go To Overview](#)

Contributors

16.43K

Contributions

270.07K

Commits

125.89K

Repositories

420

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

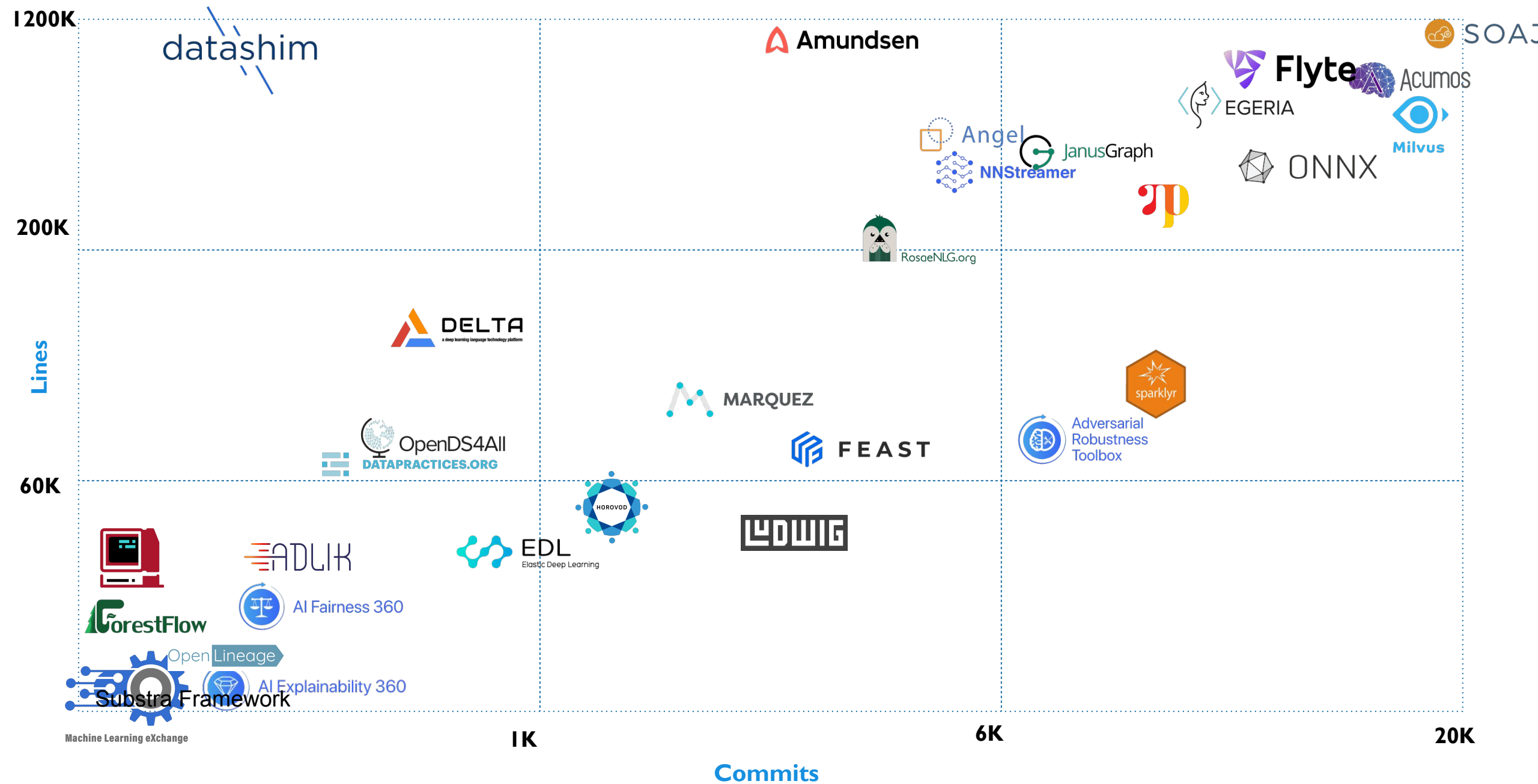
AI Track, Open Source Summit

Kubernetes + AI

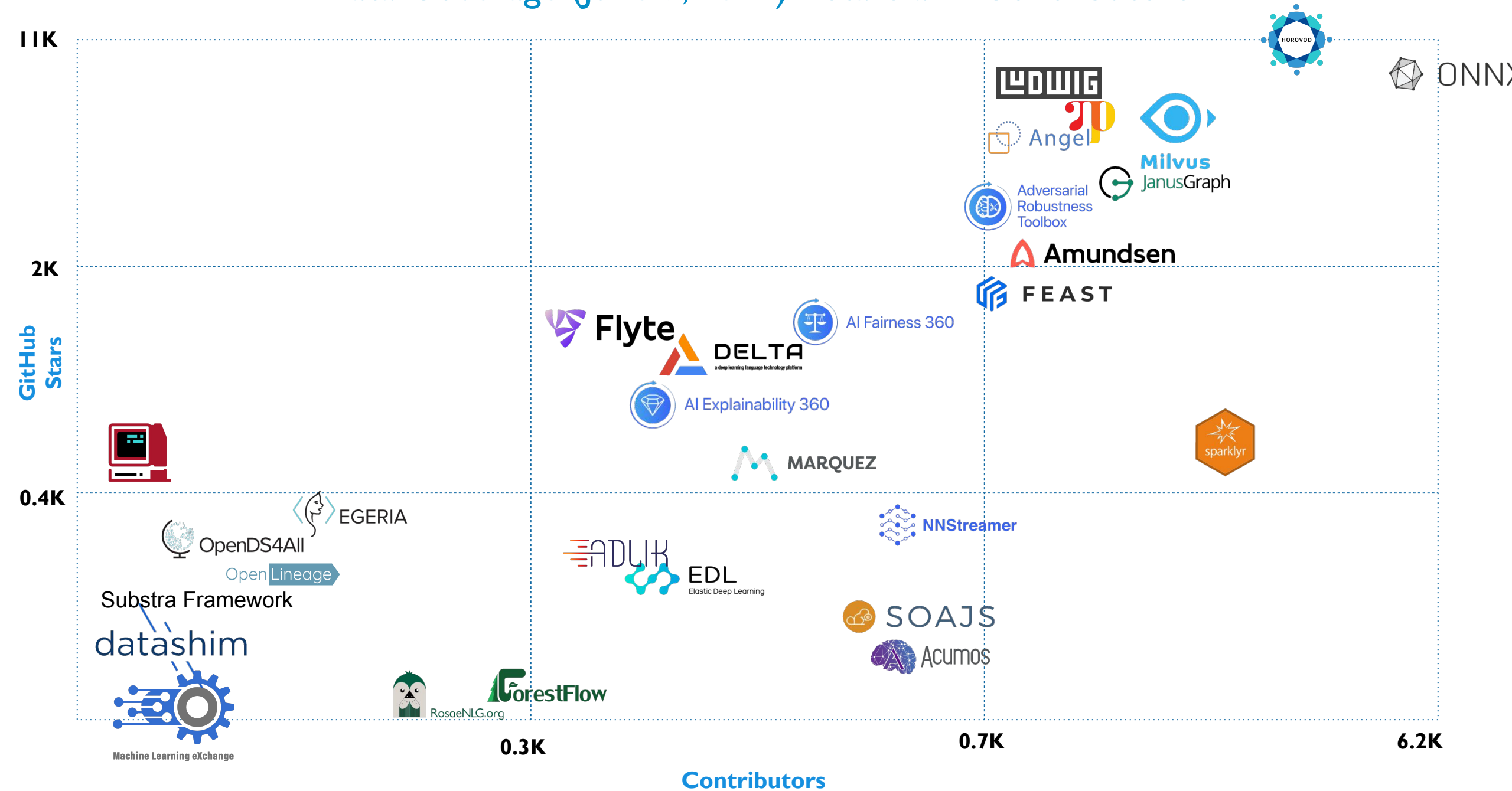
Open Networking & Edge Summit

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 community via the TAC

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