

LF AI & Data: Accelerating Development and Innovation

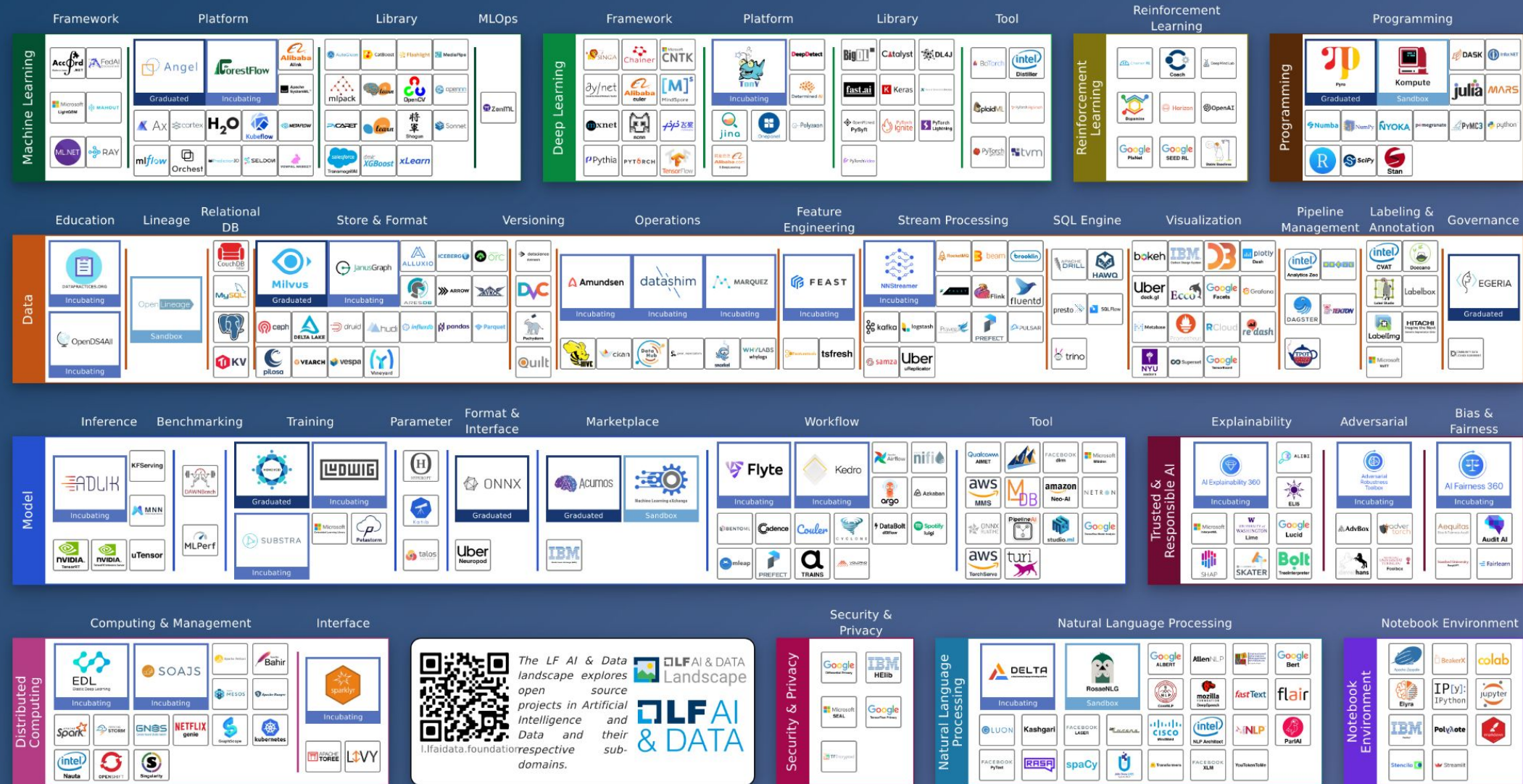
September 2021

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

 **LF** AI & DATA

Ecosystem Overview

**>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.

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

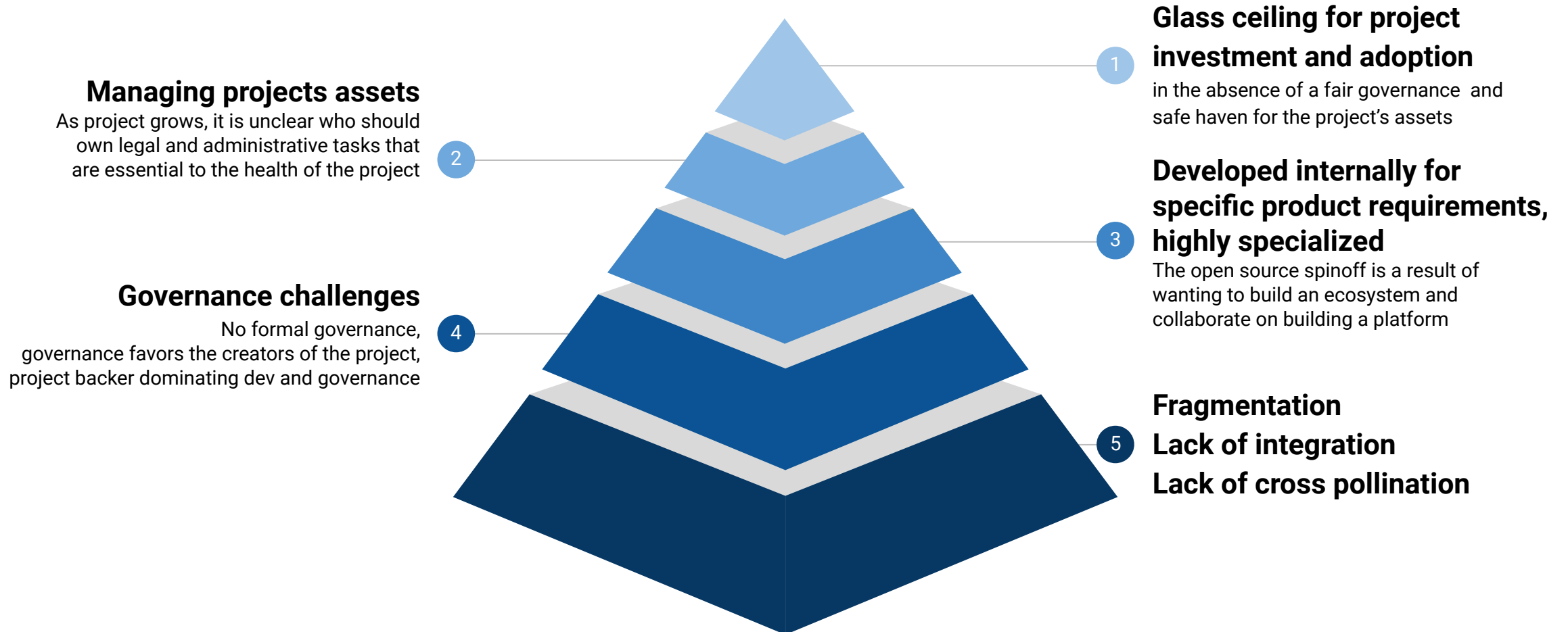
Open Data

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
- › +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

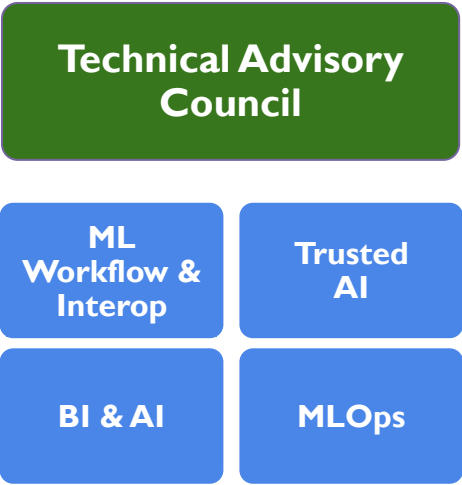
Structure and Governance

Foundation Governance

Funding effort

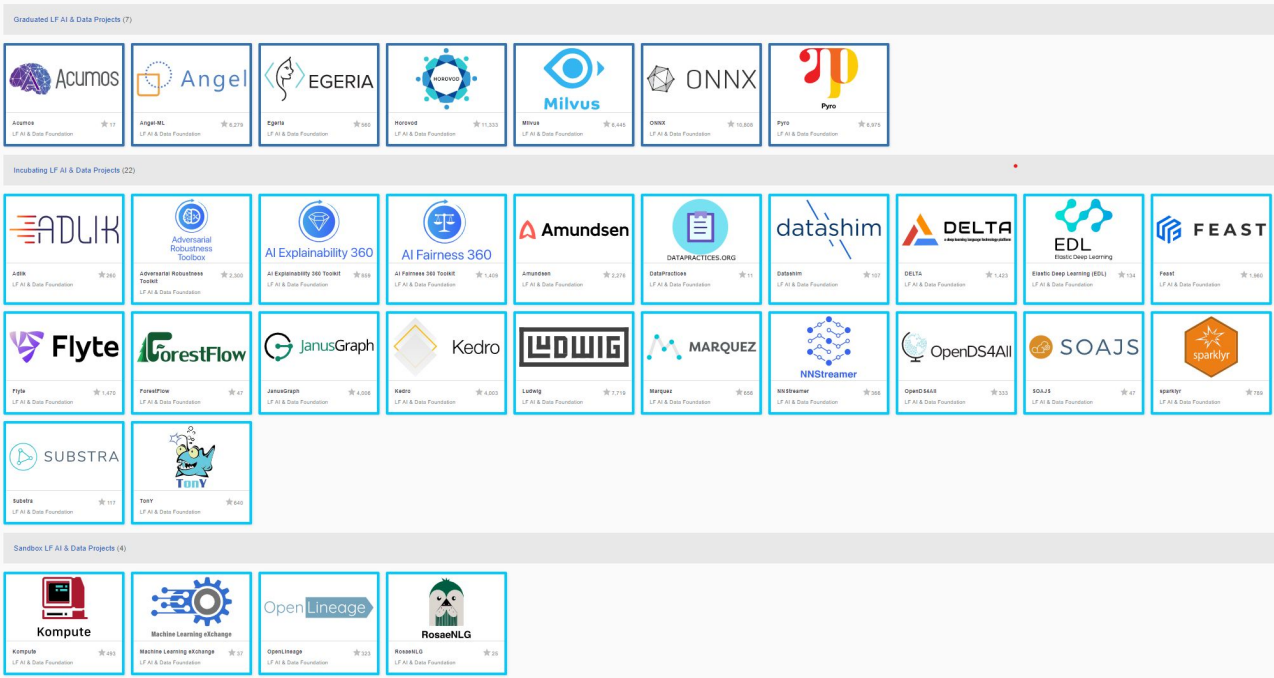


Technical Coordination



Hosted Projects

Projects have their own independent technical governance



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	 ZTE
General	 BROADCOM	 CLOUDERA	 D2 IQ	 databricks	 GRAVITI	 CS TECHNOLOGY
	 Herron Tech	 index analytics	 ING	 inwinSTACK	 orange	 precisely
	 Red Hat	 R Studio	 vmware	 XENONSTACK	 DiDi	
Associate	 AI for people	 aivancity	 AMBIANT	 BANQUE DE FRANCE EUROSYSTEME	 GALGOTIAS UNIVERSITY	 HSIS
	 NICE	 MAI	 NYU	 Open 启智 新一代人工智能开放平台	 OpenUK	 PennState Great Valley
	 SAHYADRI COLLEGE OF MANAGEMENT INNOVATION	 shopen 上海开放大学	 EIL	 UNIVERSITY OF WASHINGTON	 ULI University of Louisiana at Monroe	 PRIZE

Projects





Graduated LF AI & Data Projects (7)

 Acumos LF AI & Data Foundation ★ 17	 Angel-ML LF AI & Data Foundation ★ 6,279	 Egeria LF AI & Data Foundation ★ 580	 Horovod LF AI & Data Foundation ★ 11,333	 Milvus LF AI & Data Foundation ★ 6,445	 ONNX LF AI & Data Foundation ★ 10,808	 Pyro LF AI & Data Foundation ★ 6,975
--	---	---	---	--	--	---

Incubating LF AI & Data Projects (22)

 ADLIK LF AI & Data Foundation ★ 260	 Adversarial Robustness Toolbox LF AI & Data Foundation ★ 2,300	 AI Explainability 360 Toolkit LF AI & Data Foundation ★ 859	 AI Fairness 360 Toolkit LF AI & Data Foundation ★ 1,409	 Amundsen LF AI & Data Foundation ★ 2,276	 DataPractices LF AI & Data Foundation ★ 11	 Datashim LF AI & Data Foundation ★ 167	 DELTA LF AI & Data Foundation ★ 1,423	 Elastic Deep Learning (EDL) LF AI & Data Foundation ★ 134	 Feast LF AI & Data Foundation ★ 1,960
 Flyte LF AI & Data Foundation ★ 1,478	 ForestFlow LF AI & Data Foundation ★ 47	 JanusGraph LF AI & Data Foundation ★ 4,006	 Kedro LF AI & Data Foundation ★ 4,003	 Ludwig LF AI & Data Foundation ★ 7,719	 Marquez LF AI & Data Foundation ★ 656	 NNStreamer LF AI & Data Foundation ★ 386	 OpenDS4All LF AI & Data Foundation ★ 333	 SOAJS LF AI & Data Foundation ★ 47	 sparklyr LF AI & Data Foundation ★ 769
 Substra LF AI & Data Foundation ★ 117	 TonY LF AI & Data Foundation ★ 640								

Sandbox LF AI & Data Projects (4)

 Kompute LF AI & Data Foundation ★ 493	 Machine Learning eXchange LF AI & Data Foundation ★ 37	 OpenLineage LF AI & Data Foundation ★ 323	 RosaeNLG LF AI & Data Foundation ★ 25
--	---	--	--

Machine Learning				Deep Learning				Reinforcement Learning		Programming															
Framework		Platform		Library		MLOps		Framework		Platform		Library		Tool		Reinforcement Learning		Programming							
<div>Angel</div> <div>Graduated</div>		<div>ForestFlow</div> <div>Incubating</div>								<div>Tony</div> <div>Incubating</div>								<div>Pyro</div> <div>Graduated</div> <div>Kompute</div> <div>Sandbox</div>							
Education		Lineage		Relational DB		Store & Format		Versioning		Operations		Feature Engineering		Stream Processing		SQL Engine		Visualization		Pipeline Management		Labeling & Annotation		Governance	
<div><div>DATAFACT COLLAB</div><div>Incubating</div></div> <div><div>OpenDS4AI</div><div>Incubating</div></div>		<div>OpenLineage</div> <div>Sandbox</div>		<div>Milvus</div> <div>Graduated</div> <div>JanusGraph</div> <div>Incubating</div>						<div>Amundsen</div> <div>Incubating</div> <div>datashim</div> <div>Incubating</div> <div>MARQUEZ</div> <div>Incubating</div>		<div>FEAST</div> <div>Incubating</div>		<div>NNStreamer</div> <div>Incubating</div>										<div>EGERIA</div> <div>Graduated</div>	
Inference		Benchmarking		Training		Parameter		Format & Interface		Marketplace		Workflow		Tool		Trusted & Responsible AI		Explainability		Adversarial		Bias & Fairness			
<div>ADLIX</div> <div>Incubating</div>				<div><div>ACCORD</div><div>Graduated</div></div> <div><div>LDWIG</div><div>Incubating</div></div> <div><div>SUBTRA</div><div>Incubating</div></div>				<div>ONNX</div> <div>Graduated</div>		<div>Acumos</div> <div>Graduated</div> <div>Machine Learning exchange</div> <div>Sandbox</div>		<div>Flyte</div> <div>Incubating</div> <div>Kedro</div> <div>Incubating</div>						<div>AI Explainability 360</div> <div>Incubating</div>		<div>Adversarial Robustness Toolbox</div> <div>Incubating</div>		<div>AI Fairness 360</div> <div>Incubating</div>			
Computing & Management		Interface		Security & Privacy		Natural Language Processing		Notebook Environment																	
<div>EDL</div> <div>Incubating</div> <div>SOAJS</div> <div>Incubating</div>		<div>sparklyr</div> <div>Incubating</div>				<div>DELTA</div> <div>Incubating</div> <div>RosaeNLG</div> <div>Sandbox</div>																			
<div><div></div><div>The LF AI & Data landscape explores open source projects in Artificial Intelligence and Data and their respective sub-domains.</div><div></div></div>				<div>Security & Privacy</div>				<div>Natural Language Processing</div>				<div>Notebook Environment</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

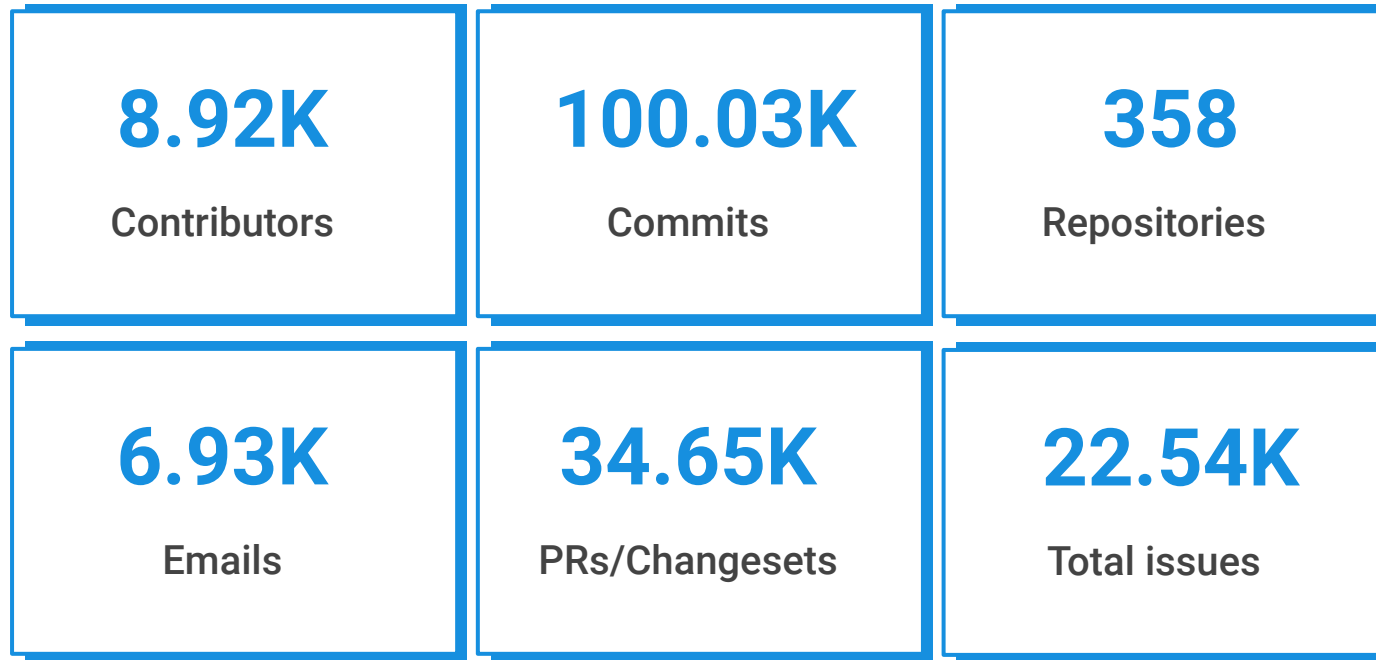
Get community exposed to how MLOps tools work together and where to use in the pipeline; Create template architecture for managing ML project lifecycle; Provide R&D collaboration opps; Fill in the gaps in use case implementation

ML Workflow & Interop

Improve interoperability and integration capabilities across hosted and external projects

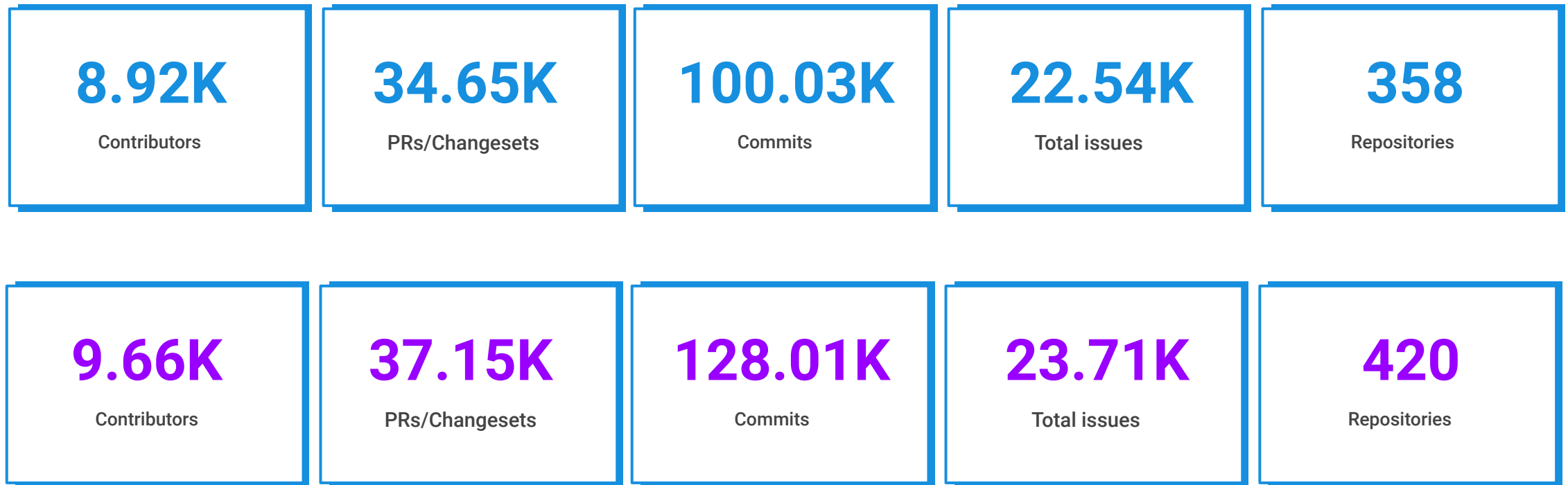
Active and growing developer community

Jan 1 - Dec 31, 2020



Active and growing developer community

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





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
16.87K

Contributions
282.21K

Commits
128.09K

Repositories
429



10.06M
Lines Of Code



138.90K
Pull Requests /
Changesets



No Data
Builds



10.13K
Emails



13.31K
Total Mentions



59.85K
Total Issues



1.65K
Pages



6.90K
Messages



15.28K
50th Percentile Of
Median Pulls By
Image



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

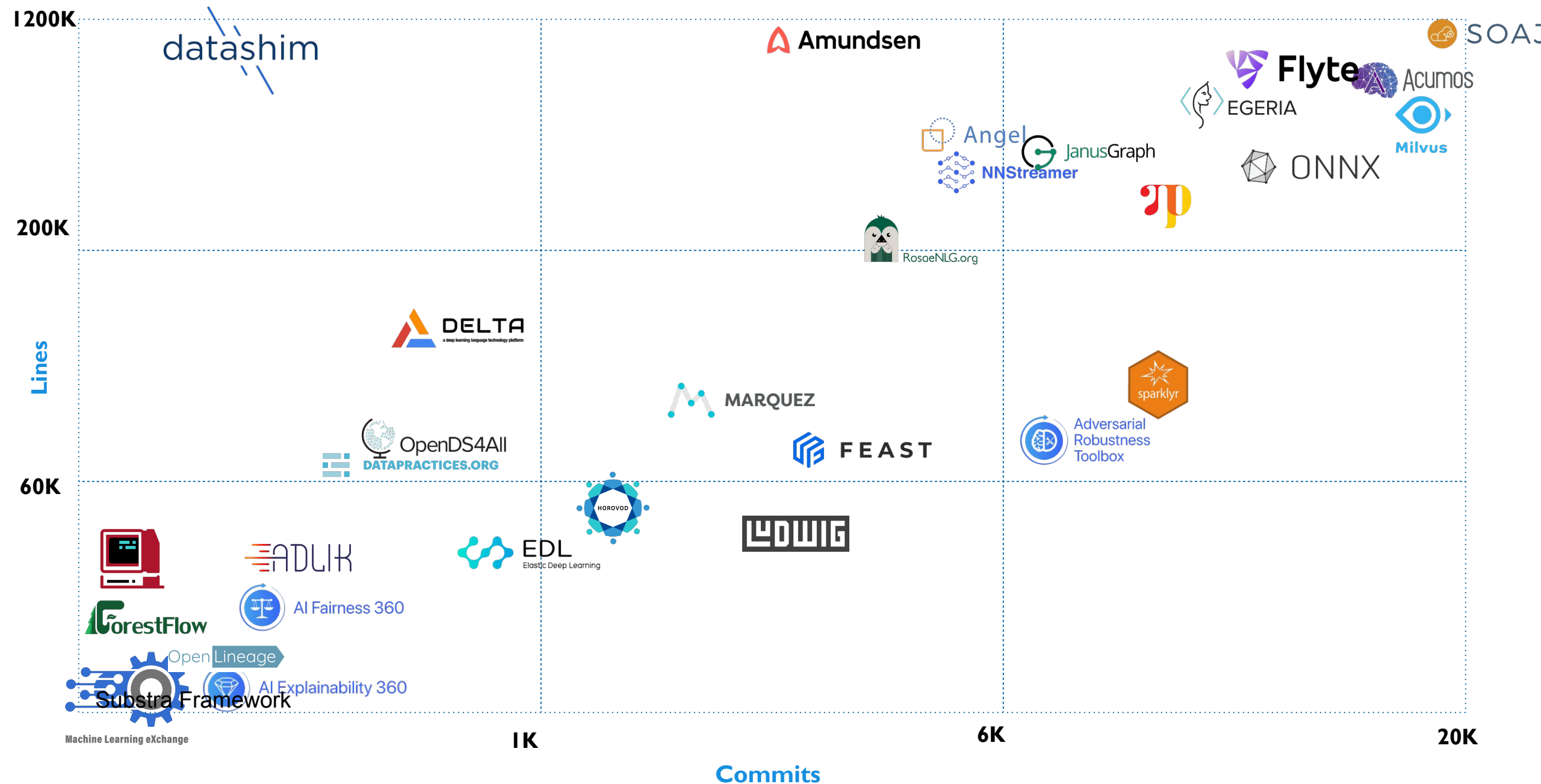
AI Track, Open Source Summit

Kubernetes + AI

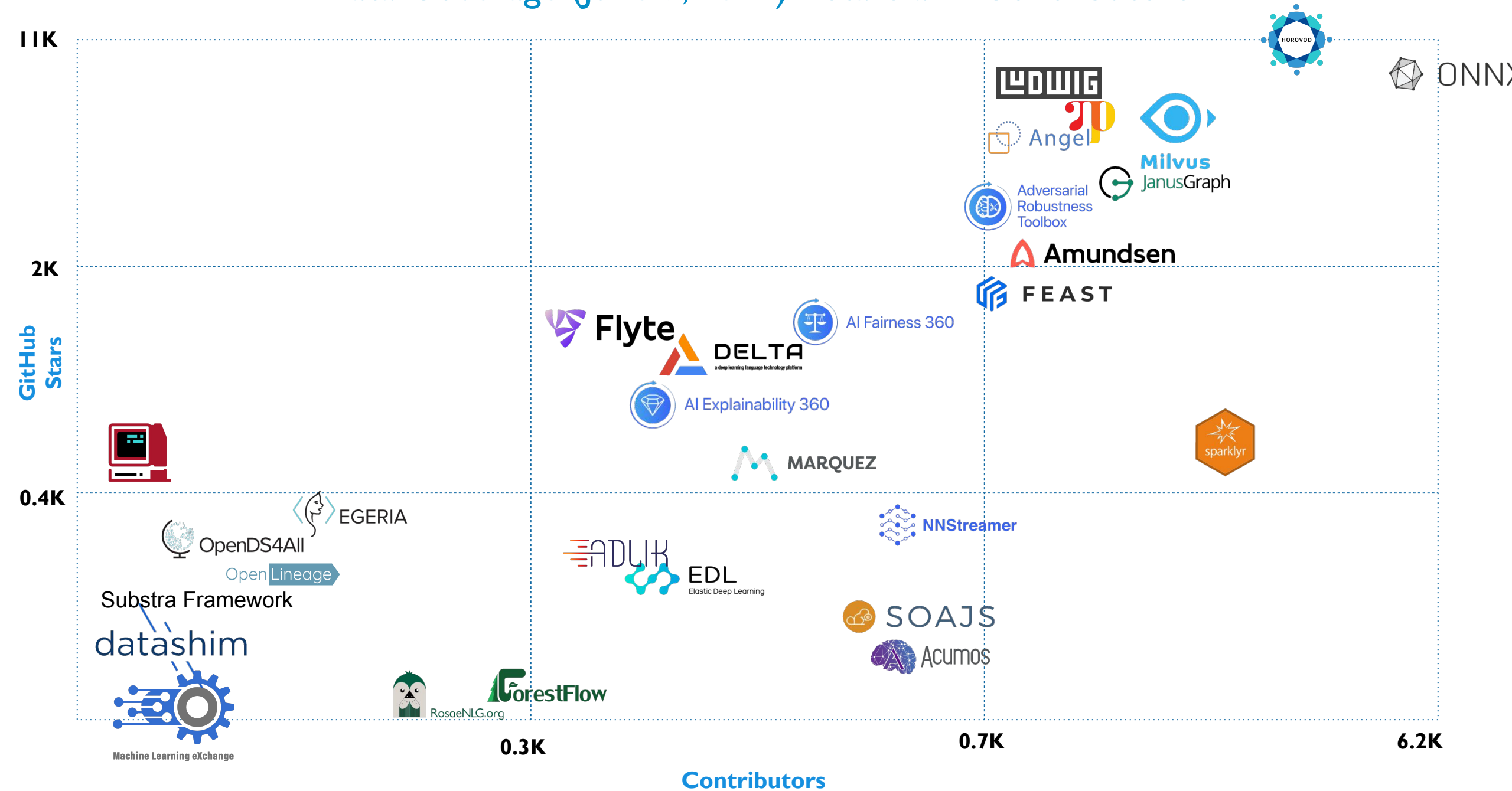
Open Networking & Edge Summit

LF AI & Data Day (multiples/year)

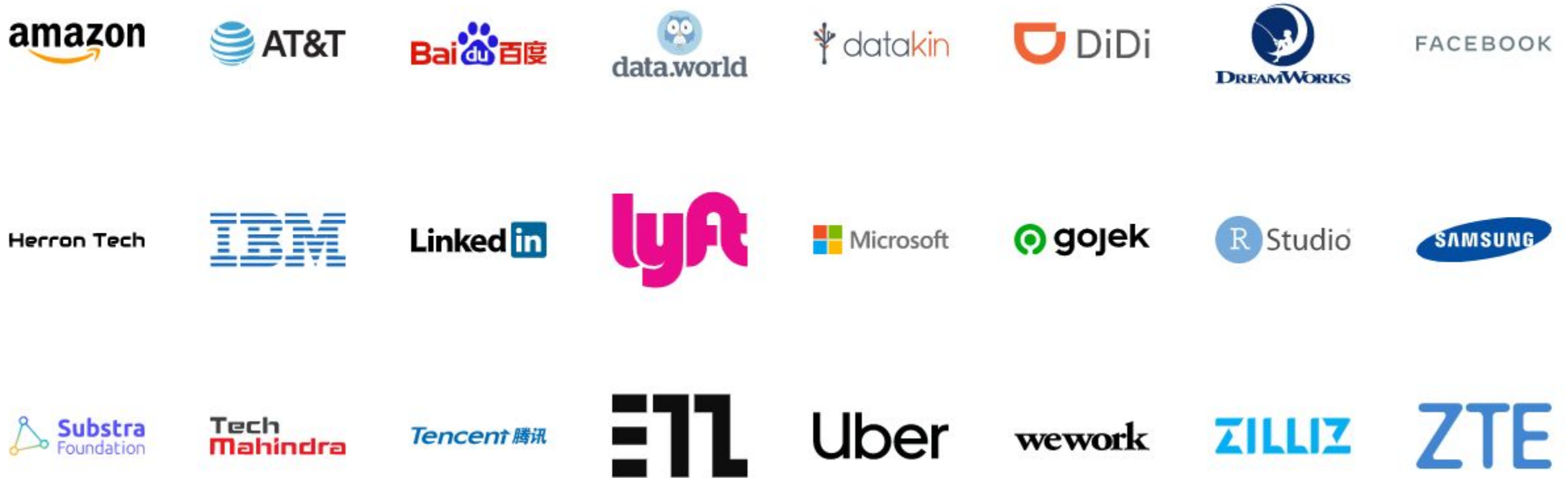
Data Coverage (June 7, 2021) - Line of Code and Commits



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, its 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