

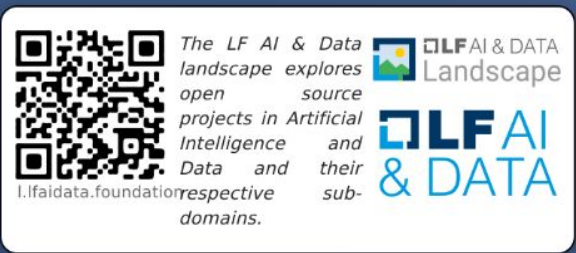
LF AI & Data

May 25, 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 and debating results and decision options

Lineage

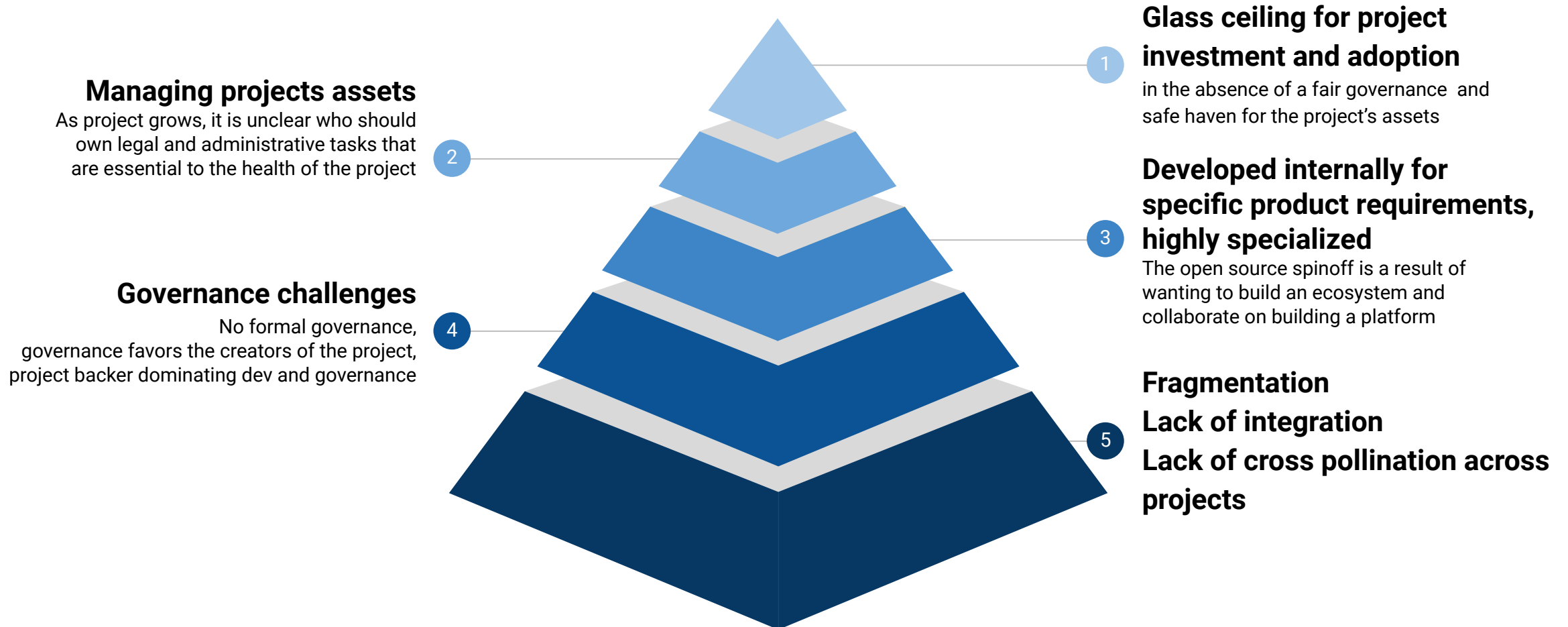
Methods to ensure provenance of datasets and AI models, including reproducibility of generated datasets and AI models

Open Data

Data Manipulation Tools

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

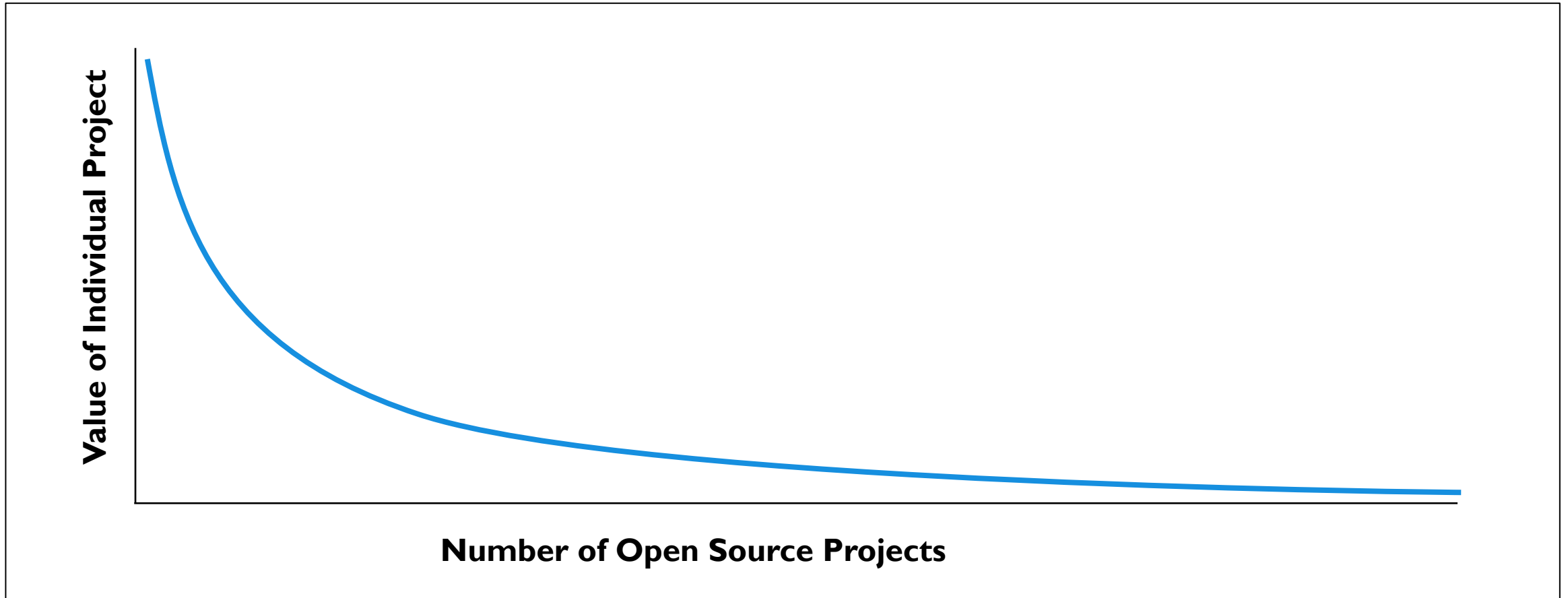
Ecosystem Challenges



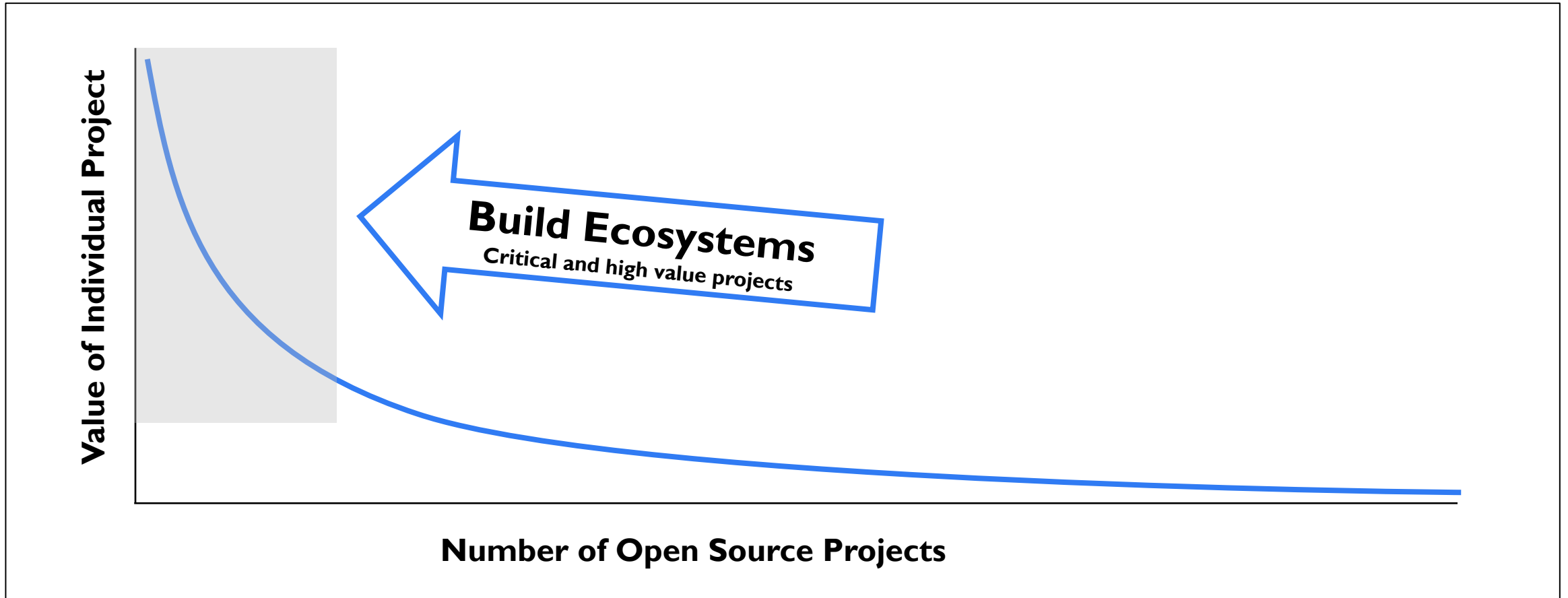
Enter LF AI & Data

 **LF** AI & DATA

Open source AI & Data isn't slowing anytime soon.
The real question is: Which projects really matter?



LF AI & Data seeks to accelerate projects to adoption and sustainability



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

Major Ongoing Efforts

Improve interoperability and capabilities across hosted projects, reduce fragmentation

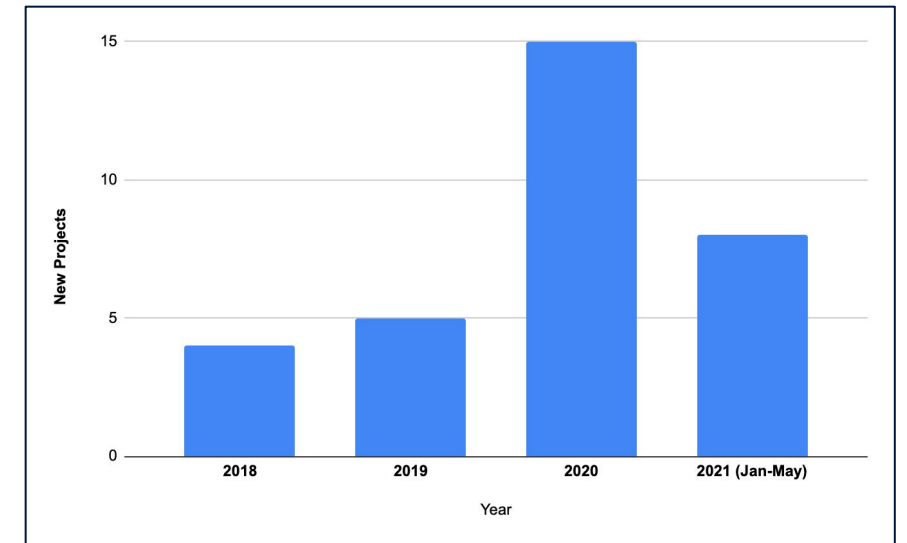
Enable closer collaboration and integration, and the support of a single Technical Advisory Council

Provide unified guidance to end users on tools, interoperability, integration, and standards

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

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)
- › >2000 subscribers to our main mailing list
- › Graduating 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

Active and growing developer community

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

8.92K

Contributors

+7.25%

34.65K

PRs/Changesets

+7.21%

100.03K

Commits

+11.03%

22.54K

Total issues

+5.19%

358

Repositories

+3.07%

2.08K

Slack messages

+68.75%

9.61K

Contributors

37.15K

PRs/Changesets

111.07K

Commits

23.71K

Total issues

369

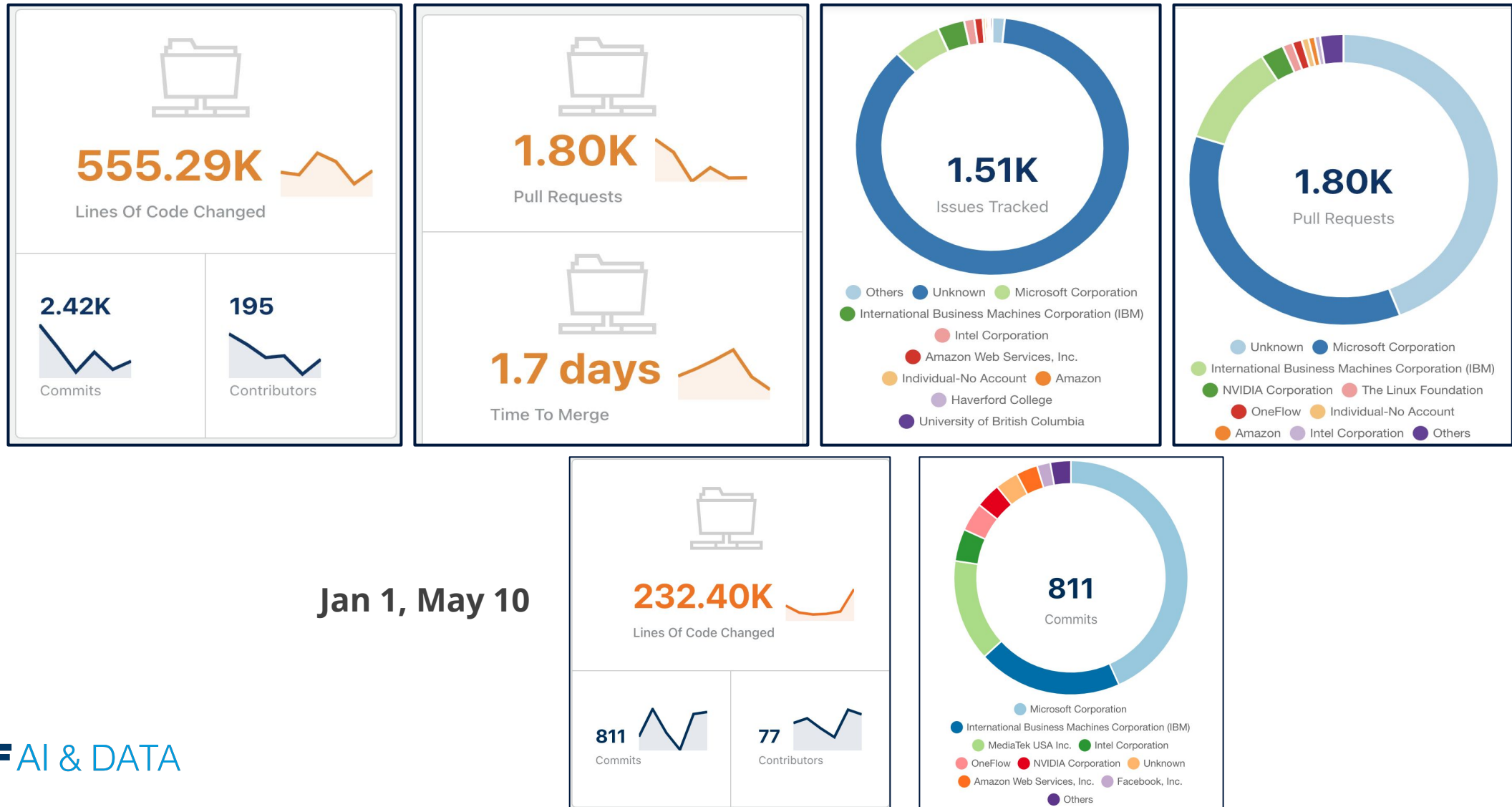
Repositories

3.51K

Slack messages

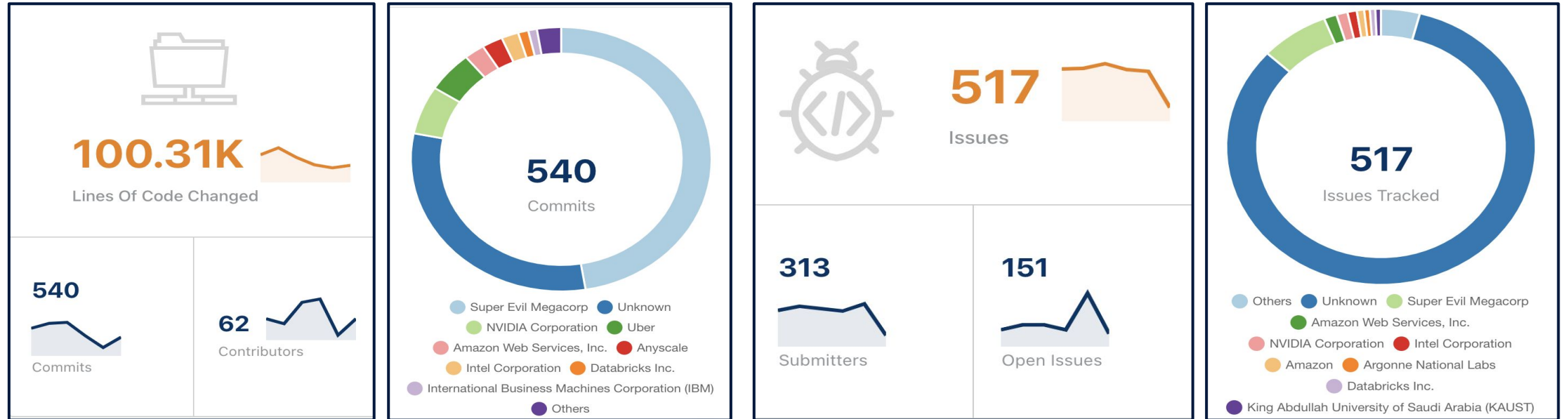
ONNX Dev Stats - Jan to Dec 2020

Open standard for machine learning interoperability

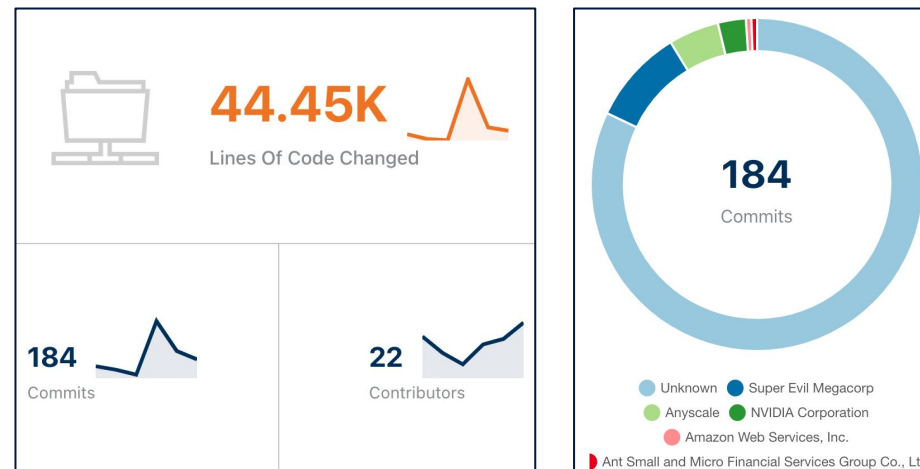


Horovod - Dev Stats - Jan to Dec 2020

Distributed training framework for TensorFlow, Keras and PyTorch



Jan 1, May 10



INSERT SLIDES FROM JIM -- from last GB

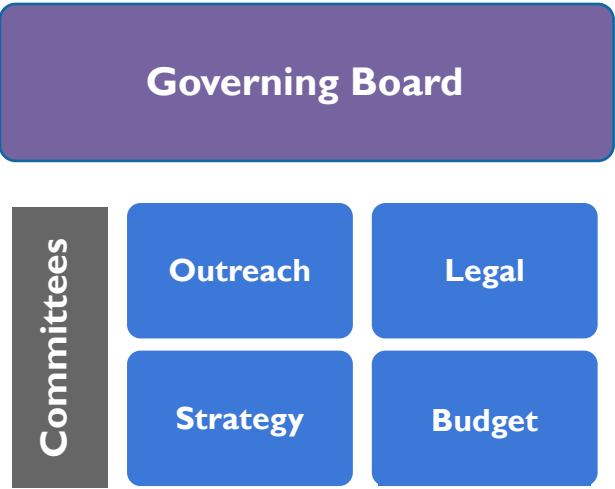
Add other examples... Projects that graduated in 2020 and 2021

go back to their proposal / see how many companies were involved and see how many are now

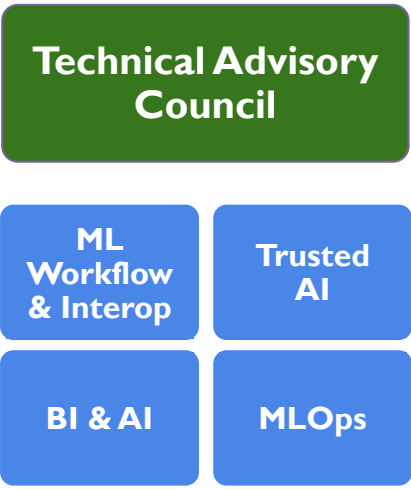
Structure and Governance

Foundation Governance

Funding effort to support hosted technical projects

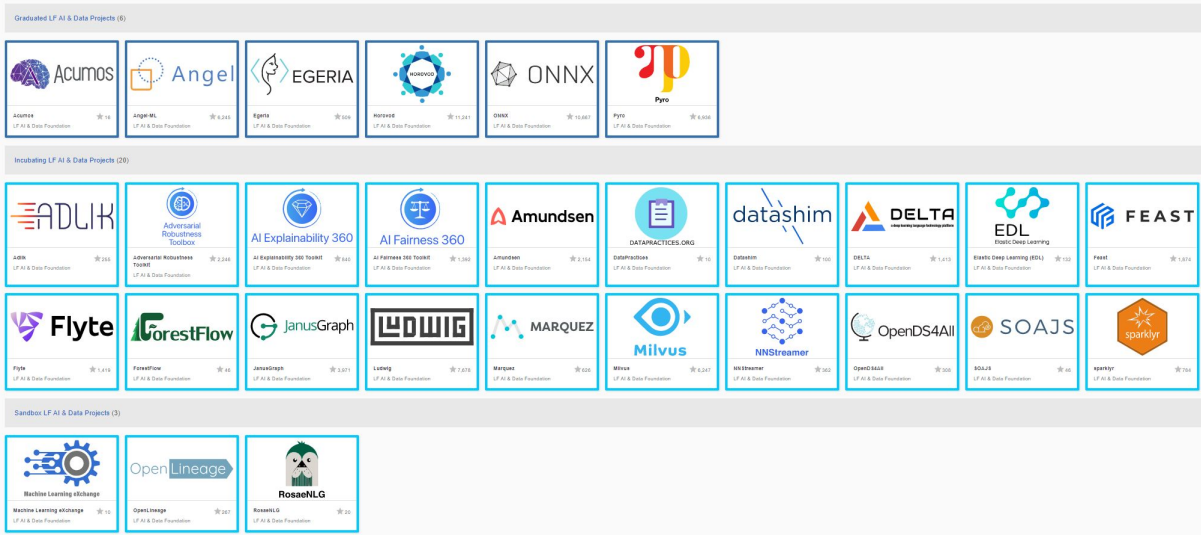


Technical Coordination



Hosted Projects

Projects have their own independent technical governance



Members

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

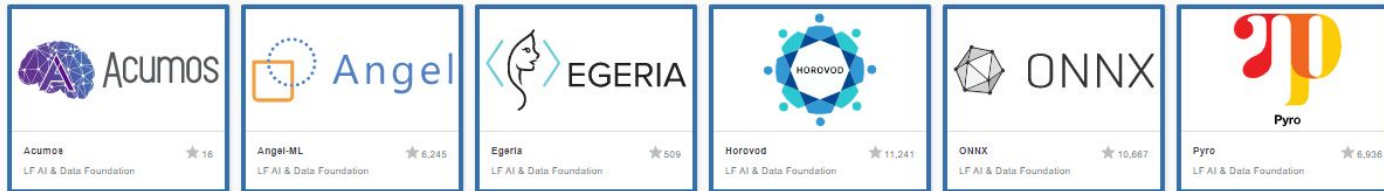
High Participation by Academia and Other Foundations

- › Visibility
- › Practical experience
- › Possible opportunities for internships
- › Possible collaboration opportunities
- › Publication opportunities
- › Open source working experience
- › Access to events

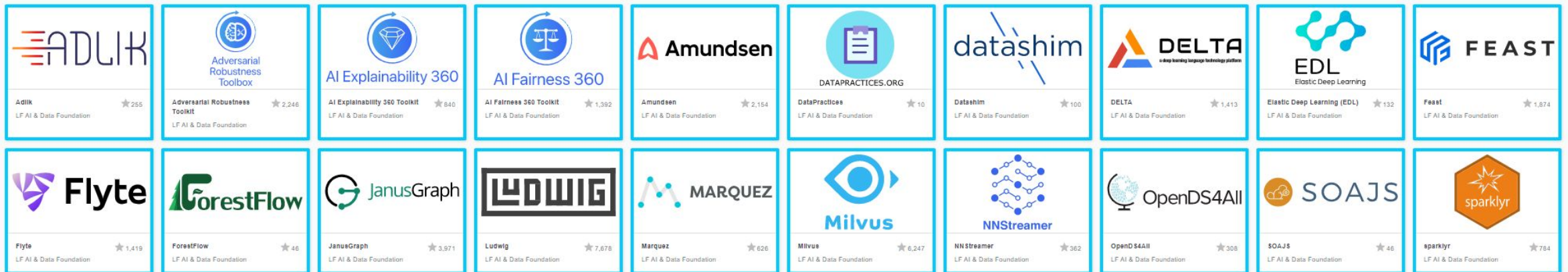
Projects (31)

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

Graduated LF AI & Data Projects (6)

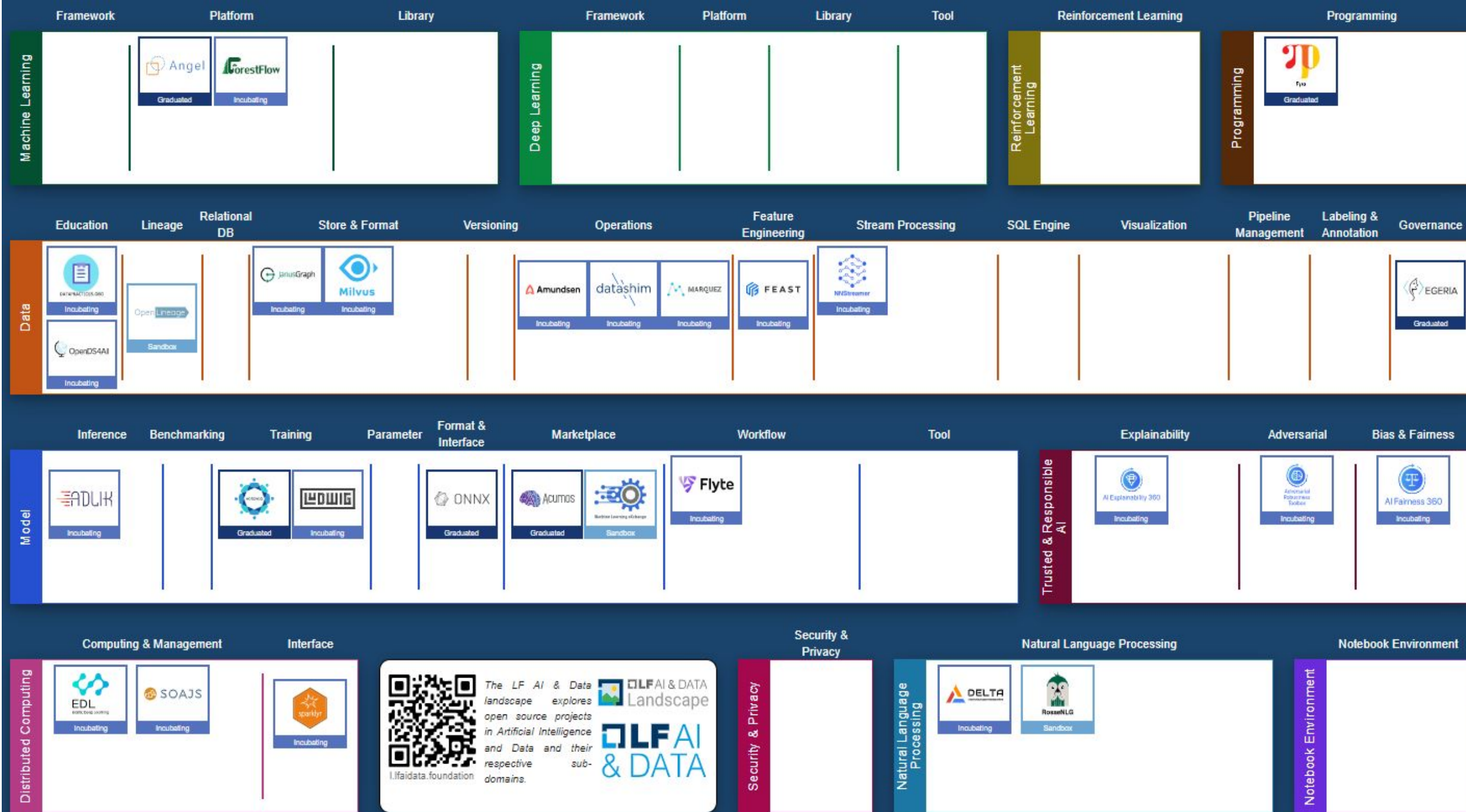


Incubating LF AI & Data Projects (20)



Sandbox LF AI & Data Projects (3)





Companies hosting projects in LF AI & Data

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

amazon

AT&T

Baidu 百度

data.world

datakin

DiDi

DREAMWORKS

FACEBOOK

Herron Tech

IBM

lyft

Microsoft

gojek

R Studio

SAMSUNG

Tech
Mahindra

Tencent 腾讯

Uber

wework

ZILLIZ
Reinvent Data Science

ZTE

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 Marketing/PR/Community Effort

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

37 announcements in 2019

61 announcements in 2020

XX announcements in 2021 (Jan - May)

- › Events announcements and promotion
- › Project release
- › Project incubation
- › New members joining
- › Project cross collaboration
- › Invited posts

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