LF AI & Data

March 1, 2021

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

LF AI & DATA

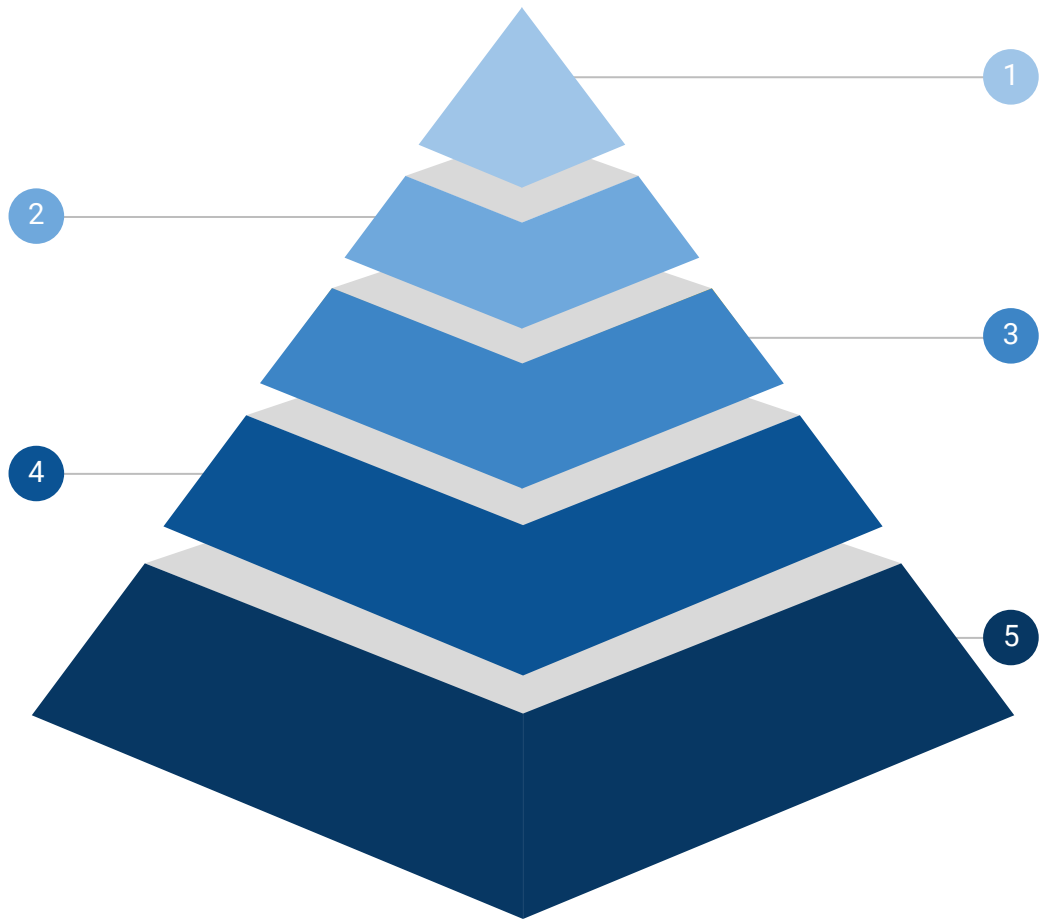
Ecosystem Overview

LF AI & DATA

Open Source AI & Data - 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 for a specific product needs
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 across projects

We all are aware of the advantages of open source ...

Access to source code	**Ability to customize**	**Rapid development**	**Peer review**	**Improved quality and security**
Reduce dev time and TTM	**Flexible licensing model**	**Community driven innovation**	**Influence roadmap via participation**	**Transparent technical discussions**

But, 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

Methods to clean, sort, tagg and track provenance and a governance structure for doing these things

Enter LF AI & Data

LF AI & DATA

LF AI & Data - Mission

To build and support an open community and a growing ecosystem of open source AI, data and analytics projects, by accelerating innovation, enabling collaboration and the creation of new opportunities for all the members of the community

LF AI TIMELINE 2018-2019

MARCH 2018

- LF DEEP LEARNING LAUNCH
- Acumos AI

MAY 2018

- ACUMOS AI CHALLENGE

JUNE 2018

- PROJECT CONTRIBUTION PROCESS & LIFECYCLE DOCUMENT

JULY 2018

- ACUMOS AI DAY FRANCE

AUGUST 2018

- Angel
- EDL Elastic Deep Learning

NOVEMBER 2018

- ACUMOS AI DAY NYC
- ACUMOS "ATHENA" RELEASE

DECEMBER 2018

- Horovod
- LF AI Landscape LAUNCH

JANUARY 2019

- yp

FEBRUARY 2019

- GETTING INVOLVED GUIDE

MARCH 2019

- LF AI MISSION & GOALS

MAY 2019

- LF AI REBRAND

AUGUST 2019

- ML WORKFLOW COMMITTEE
- ACUMOS & IBM FAIRNESS 360 INTEGRATION
- LF AI TRACK OSS NA
- LF AI MEETING SAN DIEGO

SEPTEMBER 2019

- PROJECT CONTRIBUTION PROCESS & LIFECYCLE DOCUMENT UPDATE

OCTOBER 2019

- TRUSTED AI COMMITTEE
- LF AI SUMMIT EU
- ADLIK
- ANGEL MEETUP SHENZHEN
- ONNX
- LF AI DAY PARIS
- LF AI DAY SHANGHAI

NOVEMBER 2019

- ONNX COMMUNITY MEETING
- KUBECON SHANGHAI

DECEMBER 2019

- sparklyr
- MARQUEZ
- ANGEL GRADUATION

TECHNICAL PROJECTS · FOUNDATION · COMMUNITY

INFO@LFAI.FOUNDATION

LF AI & DATA TIMELINE 2020/2021

JANUARY 2020

- Milvus

MARCH 2020

- ForestFlow
- NNStreamer

APRIL 2020

- ONNX COMMUNITY VIRTUAL MEETUP

MAY 2020

- LUDWIG

JUNE 2020

- AI Fairness 360
- AI Explainability 360
- Adversarial Robustness Toolbox
- VIRTUAL LF AI DAY EU
- OSS NA + AI/ML/DL TRACK

JULY 2020

- CLOUD NATIVE + OPEN SOURCE VIRTUAL SUMMIT CHINA 2020
- Amundsen

SEPTEMBER 2020

- SOAJS
- DELTA
- FEAST
- EGERIA
- OpenDS4All
- GRADUATION
- OPEN NETWORKING & EDGE SUMMIT

OCTOBER 2020

- LF AI DAY ONNX COMMUNITY VIRTUAL MEETUP
- RELAUNCHING AS LF AI & DATA WITH INCREASED DATA FOCUS
- LF AI & DATA FOUNDATION LAUNCH
- OSS EU + AI/ML/DL TRACK

NOVEMBER 2020

- LF AI & DATA DAY SHENZHEN
- DATAPRACTICES.ORG

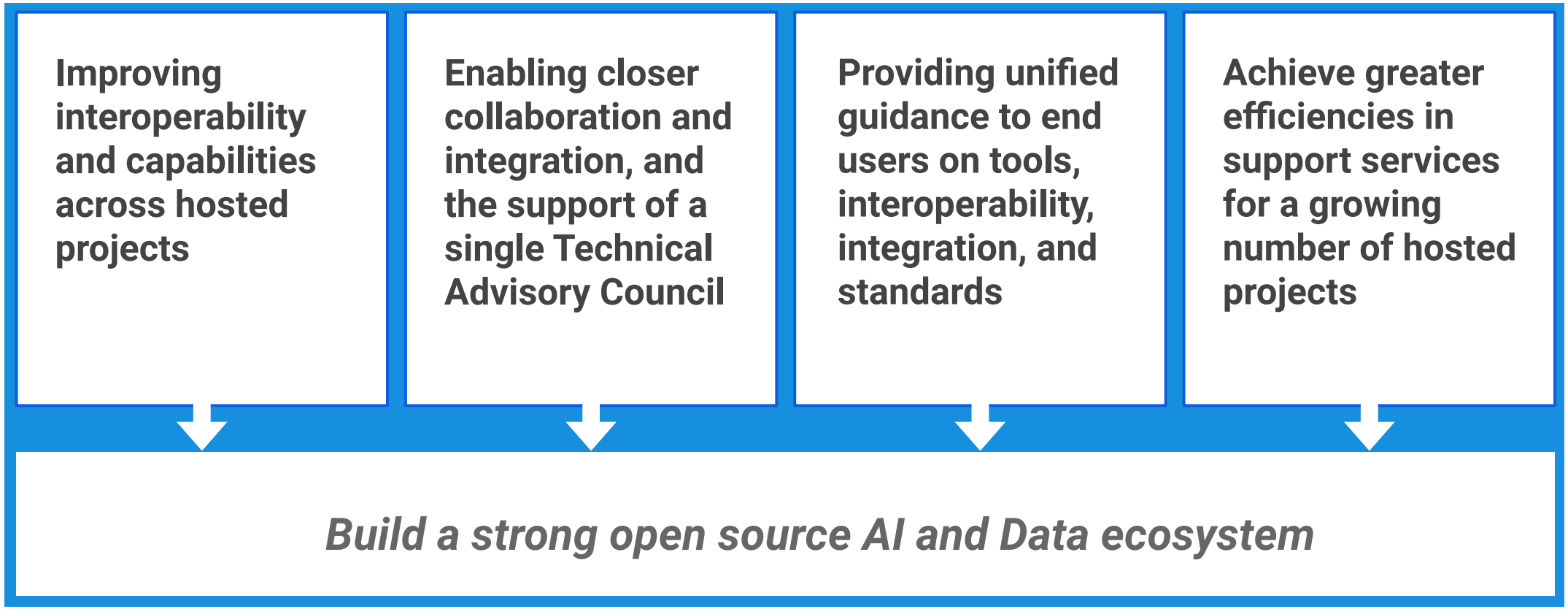
FEBRUARY 2021

- GRADUATION

INFO@LFAI.FOUNDATION

TECHNICAL PROJECTS | FOUNDATION | COMMUNITY

Motivations for Harmonization



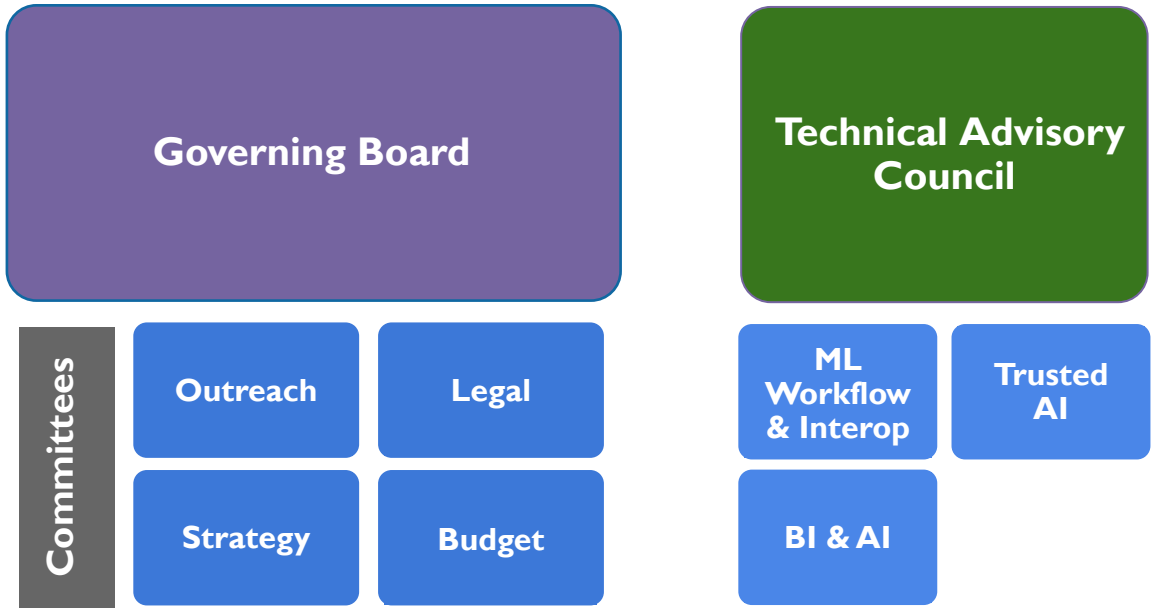
Structure and Governance

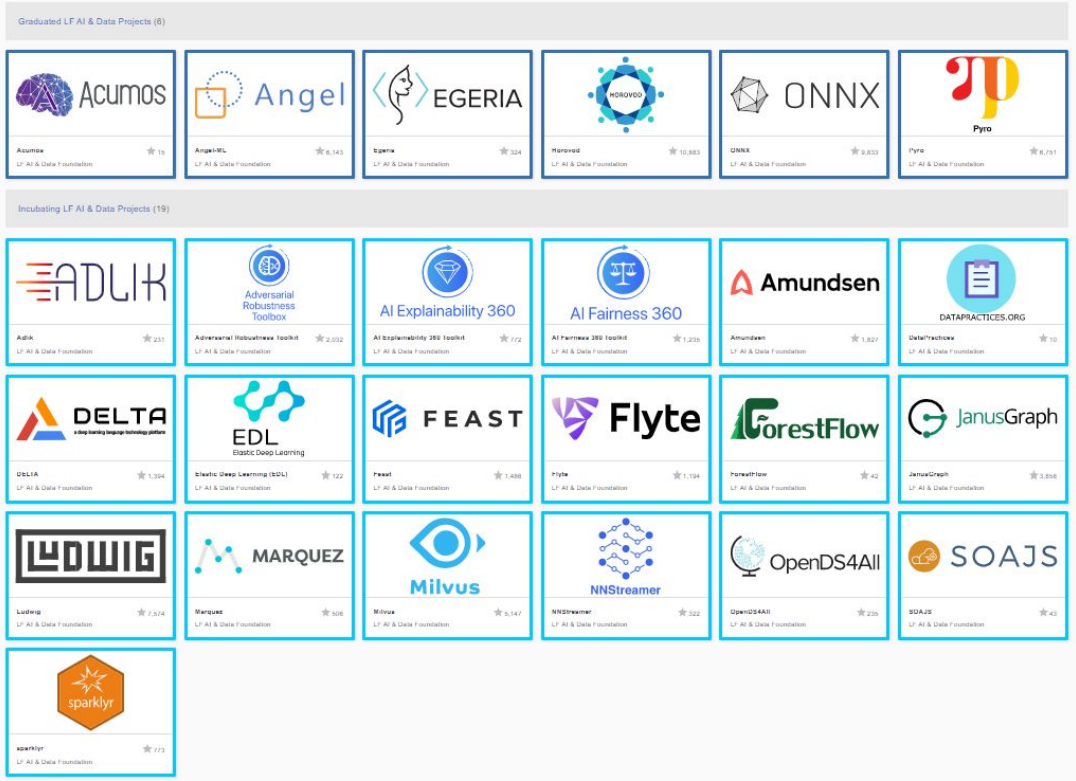
Foundation Governance
Funding effort to support hosted technical projects

Technical Coordination, New Initiatives

Hosted Projects

Projects have their own independent technical governance





LF AI & DATA

Major Efforts

Neutral Environment

Provide an neutral environment and an open governance that fosters collaboration and cross-pollination across projects and communities

Vendol neutral
#
Not-for profit
#
Open governance
#
IP clarity

Harmonization, Interoperability

Increase efforts to harmonize AI/ML/DL open source projects and reduce fragmentation; increase interoperability among technical projects

TAC
ML Workflow & Interop
#
Cross Foundation Collaboration
#
Integration efforts

Trusted and Responsible AI

Become the home of open source trusted and responsible AI, both at the level of technical projects and also at the level of principles and guidelines

Trusted AI Committee
#
3 Core Projects
#
edX Training

Data

Develop the only vendor neutral, open source standards to enable best practices for data governance, connectivity, business intelligence, and analytics.

BI & AI Committee
#
Large number of Data hosted projects

Model Marketplace

Host open source models and provide supporting tools; publish best practices; offer a marketplace for AI/ML/DL models

Open Source Model marketplace

Funding and Awareness

Provide a funding model for key projects

Increase awareness via a number of marketing and outreach activities

Training courses published in edX

Funding model
#
LF AI & Data Days, LF AI & Data Summit
#
Services: Marketing, Legal, Events and PM

Members

LF AI & DATA

2020 in numbers

+15

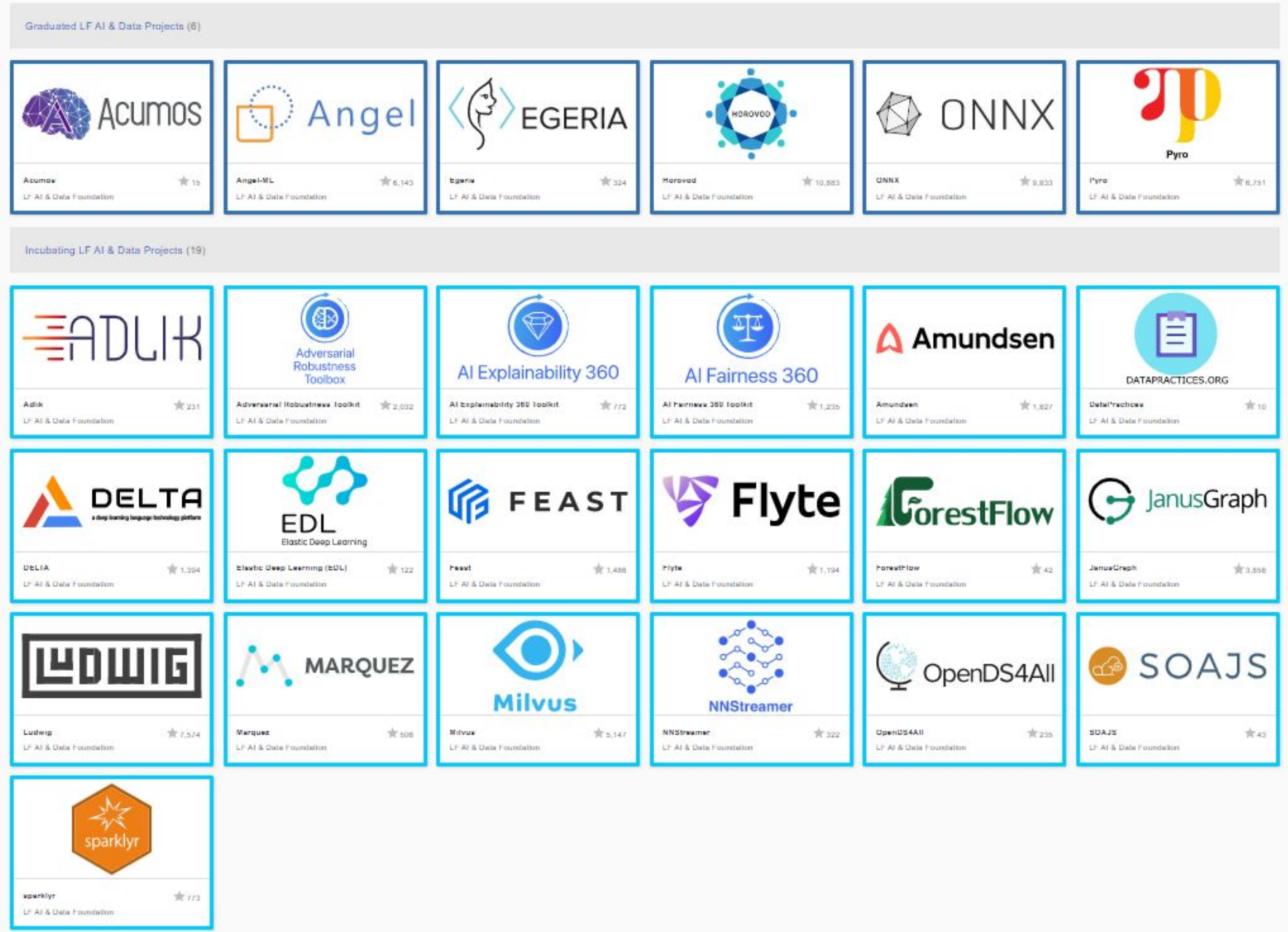
Projects

+16

Members

Projects

<u>Not listed yet:</u>
Data Lifecycle Framework



LF AI & DATA

Very active community
Jan 1 - Nov 24

8.01K	205.89K	97.60K	232
Contributors	Contributions	Commits	Repositories

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

What does it mean to be a Foundation project?

- **A neutral home for the project** increases the willingness of developers from organizations to collaborate, contribute, and become committers
- **Endorsement by members of the LF AI & Data Technical Advisory Committee** (TAC) is an independent signal of the quality of your project
- **Open and neutral governance model**
- **Program and project management services**
- **Access to full-time staff** who are eager to assist your project
- **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, etc.
- **Presence across multiple geographies** including China and Japan

Companies hosting projects in LF AI & Data

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

Why these top companies host with LF AI & Data?

The Linux Foundation is the leading commons for community assets

- The Linux Foundation (LF) is a recognized and respected brand and can activate communications across tech, business, and social media channels.
- The LF is the largest open source foundation, providing management of community assets for over 400 projects:
  - Over 3,400 source code repositories.
  - Over 700 trademark registrations and applications, plus hundreds of unregistered marks.
  - Over 700 project domains and DNS records.
  - Over 12,000 Contributor License Agreements signed for project communities that decide to require one.
- The LF offers neutral sponsorship for managing fundraising.
  - Over 1,600 unique member company relationships.
- Our communities raised over $100 million in 2019 via 6,500+ invoices.
  - Our communities spent funds in 180,000+ separate transactions in 2019.
  - The LF reports its audited financials to its board.

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.

Strong LF AI & Data Marketing/PR Effort

37 announcements between LF AI and ODPi in 2019

27 announcements between LF AI and ODPi in 1H 2020

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

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

- Content ranging from the LF AI blog, white papers, posters, presentations, and more
- Oversee blog editorial calendars and writing, editing, and/or identifying and funding freelancers to develop content
- Help develop the project website and other required creative work

Events and Community

- **LF AI & Data Summit** - co-located event with major LF conferences + Booth
- **LF AI & Data Track at Open Source Summit** (NA, EU, China) - Supporting 18-20 talks over 3 days + Booth
- **LF AI & Data Days -** a whole day (in person and virtual) hosted in collaboration with an LF AI & Data member with audience ranging from 50-250. Multiple/year.
- Talks at other leading industry events and regional events (All Things Open, Data Con LA, Big Data Conference Europe)
- Project specific events
- Sponsorship and booth at KubeCon
- Strong presence & event participation in China

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

Joining LF AI & Data is Easy!

General Inquiries:

info@lfaidata.foundation

LF AI & DATA