

# LF AI & Data

March 8, 2021

Ibrahim Haddad, Ph.D.  
Jacqueline Cardoso  
John Mertic

Executive Director, LF AI & Data  
Head of Program Management  
Director of Technical Projects

[Ibrahim@LinuxFoundation.org](mailto:Ibrahim@LinuxFoundation.org)  
[jzcardoso@linuxfoundation.org](mailto:jzcardoso@linuxfoundation.org)  
[jmertic@linuxfoundation.org](mailto:jmertic@linuxfoundation.org)

 LF AI & DATA

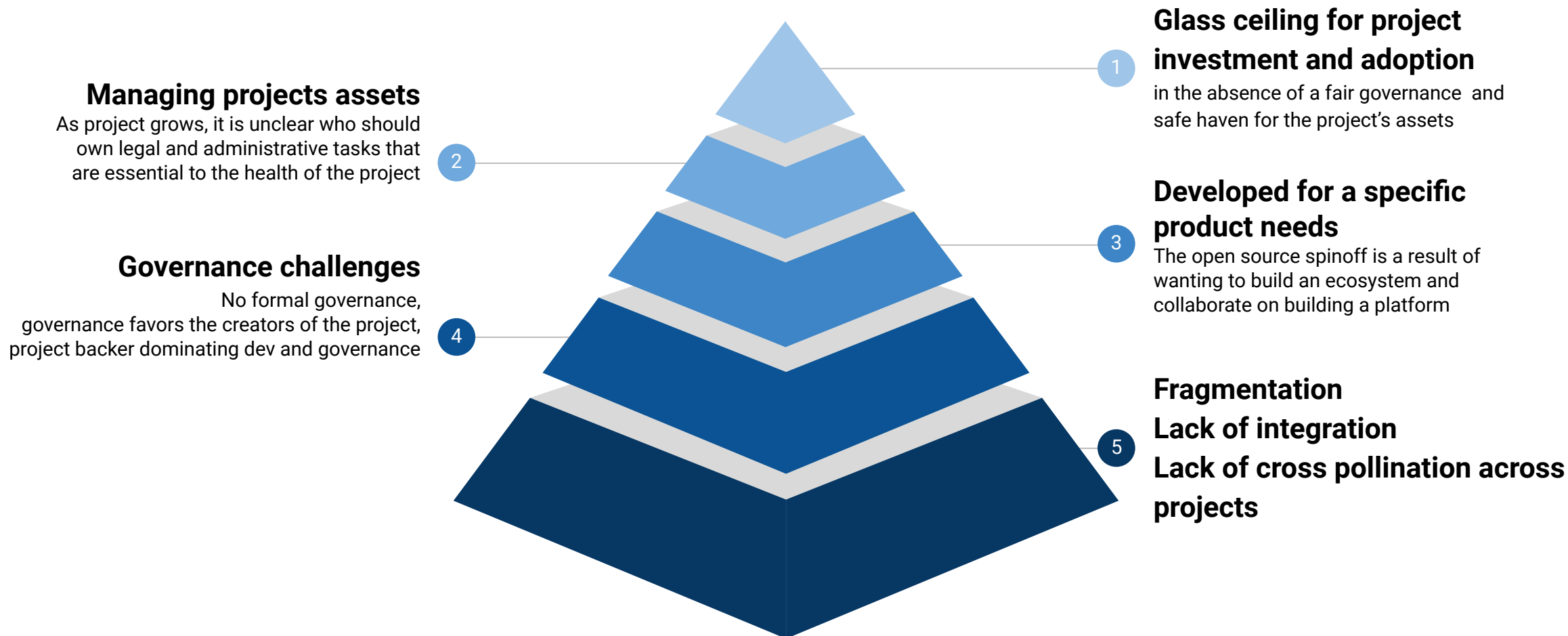
# LF AI & Data

Collaborating to support an open and 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

<https://lfaidata.foundation>

# Ecosystem Overview

# Open Source AI & Data - Ecosystem Challenges



# Enter LF AI & Data

 **LF** AI & DATA

# What is LF AI & Data?

- › Non-profit
- › Open source collaborative effort focused on AI & Data
- › Hosted at the Linux Foundation
- › Focused on supporting open source AI & data projects and accelerating innovation and their trajectory to adoption

# 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

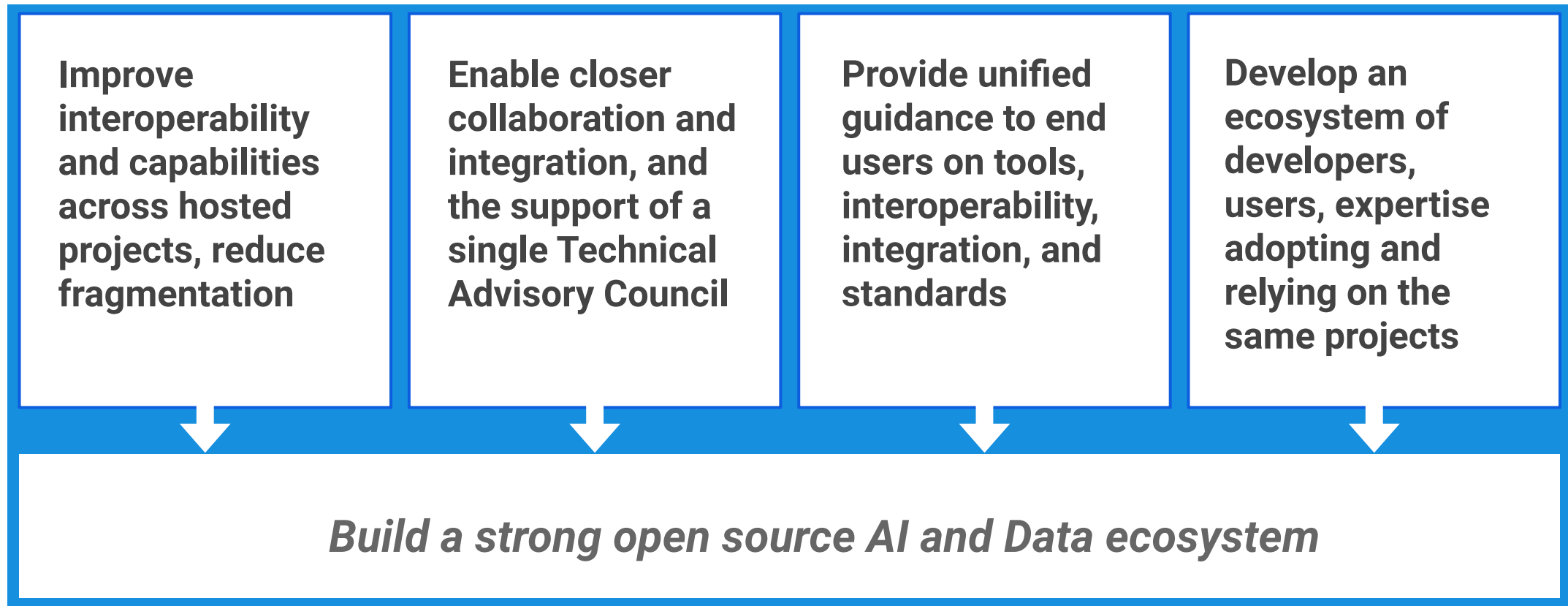
## Lineage

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

# Goals of LF AI & Data





# Why is LF AI & Data important?

1

## OPEN & NEUTRAL

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

2

## HARMONIZATION & INTEROPERABILITY

Increase efforts to harmonize AI & Data open source projects and reduce fragmentation; increase interoperability among technical projects

3

## TRUSTED & 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

4

## ECOSYSTEM DEVELOPMENT

Fund key projects; Increase awareness via a number of marketing and outreach activities; Training courses published in edX

5

## DATA

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

# LF AI & Data Growth

- › Launched in 03/2018 with 9 members and 1 hosted project
- › +3 projects in 2018
- › +5 projects in 2019
- › +15 projects, +16 members in 2020
- › Currently supported by 46 member companies
- › >2000 subscribers to our main mailing list

# Very active 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

# Structure and Governance

## Foundation Governance

Funding effort to support hosted technical projects

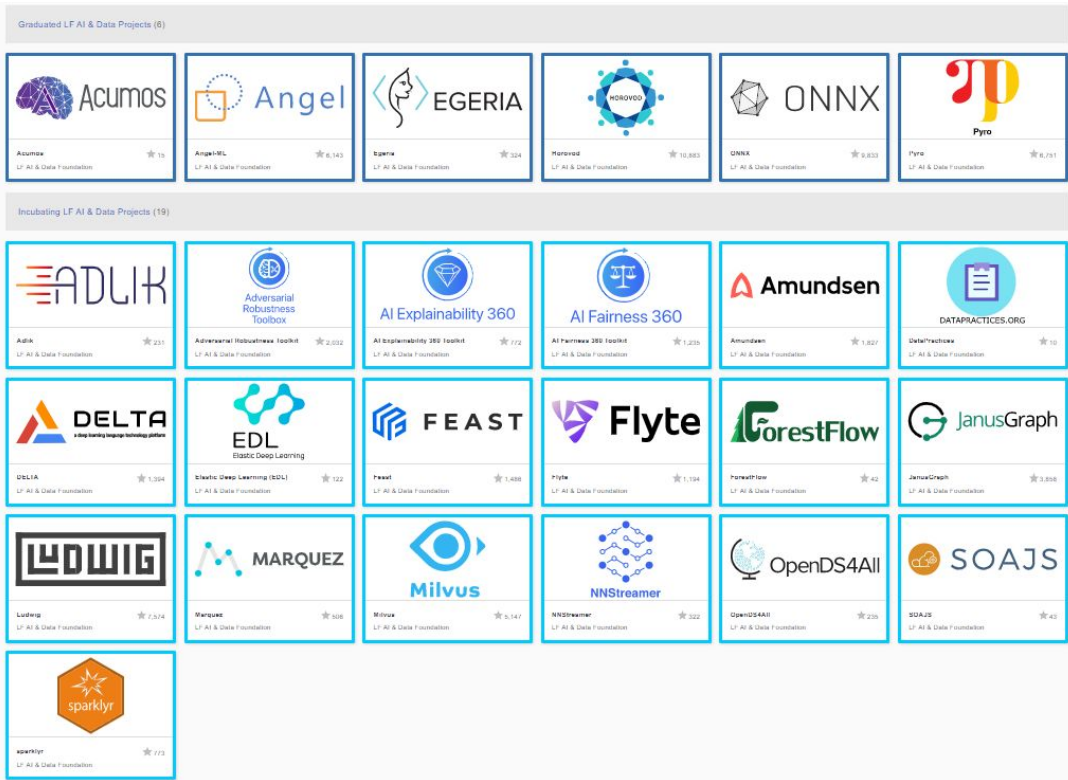


## Technical Coordination, New Initiatives



## Hosted Projects

Projects have their own independent technical governance



# Members

<https://landscape.lfai.foundation/members?project=hosted>

LF AI & Data Member Company - Premier



LF AI & Data Member Company - General














LF AI & Data Member Company - Associate



# Projects

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

 <b>Acumos</b> ★ 15 LF AI & Data Foundation	 <b>Angel</b> ★ 6,146 Angel-ML LF AI & Data Foundation	 <b>EGERIA</b> ★ 326 Egeria LF AI & Data Foundation	 <b>HOROVOD</b> ★ 10,909 Horovod LF AI & Data Foundation	 <b>ONNX</b> ★ 9,862 ONNX LF AI & Data Foundation	 <b>Pyro</b> ★ 6,783 Pyro LF AI & Data Foundation	
 <b>ADLIK</b> ★ 236 Adlik LF AI & Data Foundation	 <b>Adversarial Robustness Toolkit</b> ★ 2,047 Adversarial Robustness Toolkit LF AI & Data Foundation	 <b>AI Explainability 360</b> ★ 780 AI Explainability 360 Toolkit LF AI & Data Foundation	 <b>AI Fairness 360</b> ★ 1,245 AI Fairness 360 Toolkit LF AI & Data Foundation	 <b>Amundsen</b> ★ 1,860 Amundsen LF AI & Data Foundation	 <b>DATA PRACTICES.ORG</b> ★ 10 DataPractices LF AI & Data Foundation	 <b>DELTA</b> ★ 1,393 DELTA LF AI & Data Foundation
 <b>EDL</b> ★ 125 Elastic Deep Learning (EDL) LF AI & Data Foundation	 <b>FEAST</b> ★ 1,531 Feast LF AI & Data Foundation	 <b>Flyte</b> ★ 1,216 Flyte LF AI & Data Foundation	 <b>ForestFlow</b> ★ 43 ForestFlow LF AI & Data Foundation	 <b>JanusGraph</b> ★ 3,867 JanusGraph LF AI & Data Foundation	 <b>LUDWIG</b> ★ 7,585 Ludwig LF AI & Data Foundation	 <b>MARQUEZ</b> ★ 516 Marquez LF AI & Data Foundation
 <b>Milvus</b> ★ 5,251 Milvus LF AI & Data Foundation	 <b>NNStreamer</b> ★ 324 NNStreamer LF AI & Data Foundation	 <b>OpenDS4All</b> ★ 237 OpenDS4All LF AI & Data Foundation	 <b>SOAJS</b> ★ 43 SOAJS LF AI & Data Foundation	 <b>sparklyr</b> ★ 775 sparklyr LF AI & Data Foundation		

Not listed yet:  
Datashim

# 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](#)

# Companies hosting projects in LF AI & Data

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

amazon

AT&T

Baidu 百度

data.world

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 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

# Strong Marketing/PR/Community Effort

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

**37 announcements in 2019**

**61 announcements in 2020**

- › 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.**



# Joining LF AI & Data is Easy!



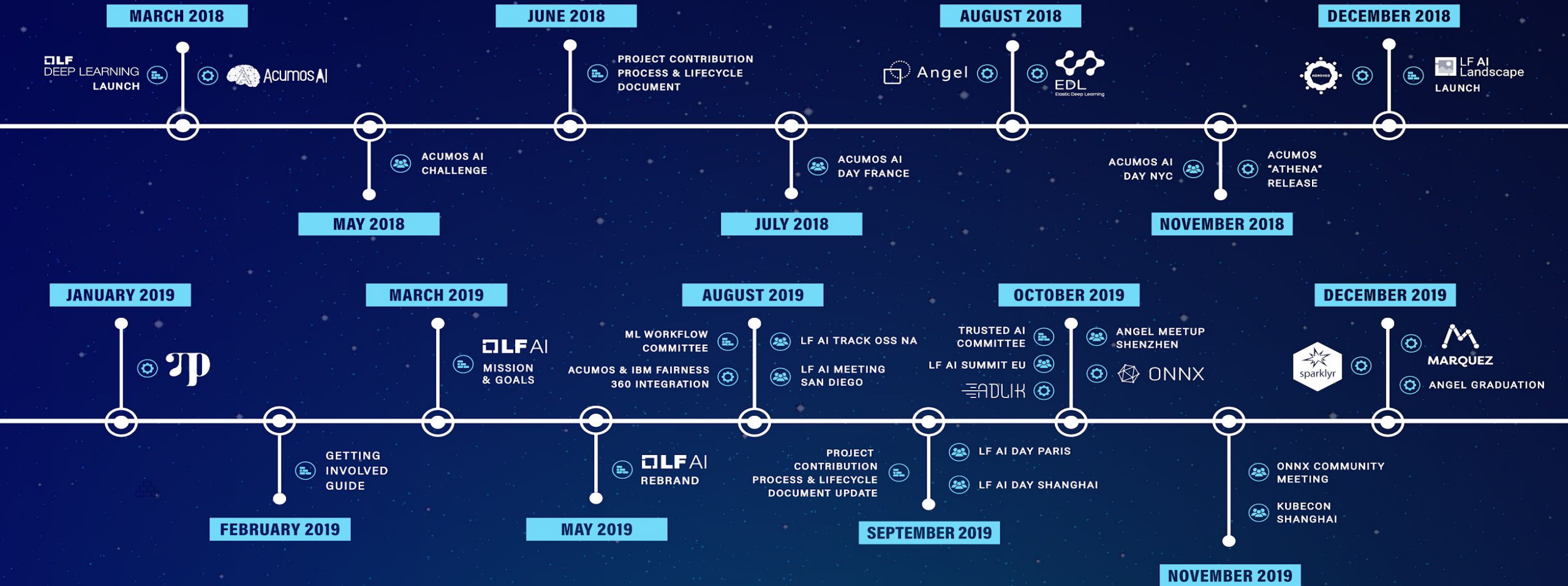
General Inquiries:

[info@lfaidata.foundation](mailto:info@lfaidata.foundation)

 **LF** AI & DATA

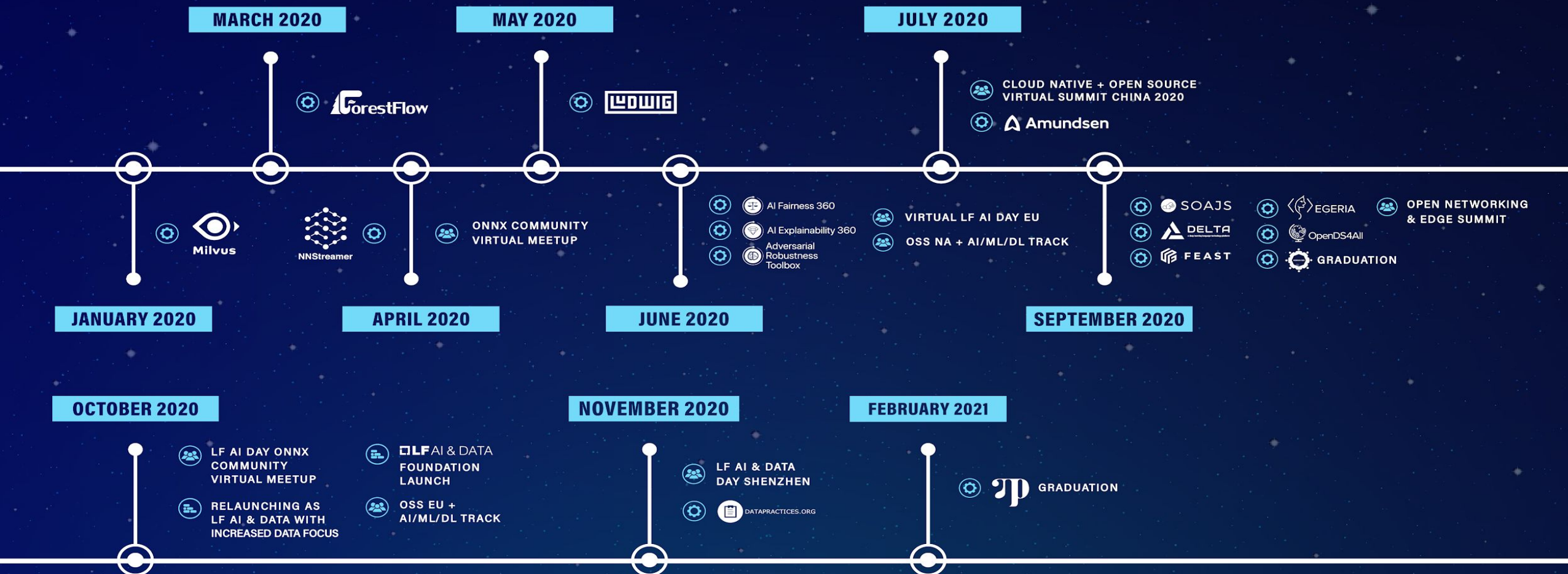
# Backup

# LF AI TIMELINE 2018-2019





# LF AI & DATA TIMELINE 2020/2021



# 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.



# 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**