LF AI & Data

June 2021

Ibrahim Haddad, Ph.D.
Executive Director, LF AI & Data
Ibrahim@LinuxFoundation.org
Ecosystem Overview
305 projects
2.2M+ GitHub Stars
450M+ LoC
40K+ developers

130+ founding orgs
10+ universities
13 licenses
>450M LoC
>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 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

1. Glass ceiling for project investment and adoption in the absence of a fair governance and safe haven for the project's assets

2. 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

3. Developed internally for specific product requirements, highly specialized
   The open source spinoff is a result of wanting to build an ecosystem and collaborate on building a platform

4. Governance challenges
   No formal governance, governance favors the creators of the project, project backer dominating dev and governance

5. Fragmentation
   Lack of integration
   Lack of cross pollination

LFAI & DATA
Enter LF AI & Data
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
History & Growth

› Launched in 03/2018 with 9 members and 1 hosted project
› +3 projects in 2018
› +5 projects in 2019
› +15 projects in 2020
› +8 projects (Jan - May) 2021
› Supported by 49 member companies (all geos)
› Build an ecosystem around critical and high value projects and accelerate the adoption and sustainability of hosted projects
› >2000 subscribers to our main mailing list
Structure and Governance

Foundation Governance
Funding effort to support hosted technical projects

Technical Coordination

Hosted Projects
Projects have their own independent technical governance

Governing Board

Committees
Outreach
Legal
Strategy
Budget

Technical Advisory Council
ML Workflow & Interop
Trusted AI
BI & AI
MLOps

Hosted Projects
Members
https://landscape.lfai.foundation/card-mode?project=company
High Participation by Academia and Other Member Foundations

- Visibility
- Practical experience working with open source foundations
- Internships opportunities
- Collaboration opportunities
- Publication opportunities
- Open source working experience
- Access to events
Projects (31)
https://landscape.lfai.foundation/card-mode?project=company

* Missing: Substra Framework, Kompute (pending logo redesign)
## Major Ongoing Efforts

<table>
<thead>
<tr>
<th>Effort</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Develop an ecosystem of developers, users, expertise adopting and relying on the same projects</td>
<td>Enable closer collaboration with the support of a single Technical Advisory Council across all technical efforts</td>
</tr>
<tr>
<td>Provide unified guidance to end users on open source software solutions, tools, interoperability, integration, and standards</td>
<td>Increase efforts to harmonize AI/Data open source projects, reduce fragmentation; Rally the industry to adopt Foundation hosted projects</td>
</tr>
<tr>
<td>Increase efforts to harmonize AI/Data open source projects, reduce fragmentation; Rally the industry to adopt Foundation hosted projects</td>
<td>Provide a funding model for key projects; Increase awareness via a number of marketing and outreach activities;</td>
</tr>
</tbody>
</table>

### 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
Very active and growing developer community
Jan 1- Dec 31, 2020

<table>
<thead>
<tr>
<th>Contributors</th>
<th>Contributions</th>
<th>Commits</th>
<th>Repositories</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.92K</td>
<td>217K</td>
<td>100.03K</td>
<td>358</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Emails</th>
<th>PRs/Changesets</th>
<th>Slack messages</th>
<th>Total issues</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.93K</td>
<td>34.65K</td>
<td>2.08K</td>
<td>22.54K</td>
</tr>
</tbody>
</table>

Source: [https://insights.lfx.linuxfoundation.org/projects/lfai](https://insights.lfx.linuxfoundation.org/projects/lfai)
Active and growing developer community

<table>
<thead>
<tr>
<th>Metric</th>
<th>Jan 1- Dec 31, 2020</th>
<th>Jan 1, 2020 to Apr 2, 2021</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contributors</td>
<td>8.92K</td>
<td>9.61K</td>
<td>+7.25%</td>
</tr>
<tr>
<td>PRs/Changesets</td>
<td>34.65K</td>
<td>37.15K</td>
<td>+7.21%</td>
</tr>
<tr>
<td>Commits</td>
<td>100.03K</td>
<td>111.07K</td>
<td>+11.03%</td>
</tr>
<tr>
<td>Total issues</td>
<td>22.54K</td>
<td>23.71K</td>
<td>+5.19%</td>
</tr>
<tr>
<td>Repositories</td>
<td>358</td>
<td>369</td>
<td>+3.07%</td>
</tr>
<tr>
<td>Slack messages</td>
<td>2.08K</td>
<td>3.51K</td>
<td>+68.75%</td>
</tr>
</tbody>
</table>

+7.21%  +11.03%  +5.19%  +3.07%  +68.75%
ONNX Dev Stats - Jan to Dec 2020
Open standard for machine learning interoperability

555,290
Lines Of Code Changed

1,800
Pull Requests

1,510
Issues Tracked

1,800
Pull Requests

2,420
Commits

195
Contributors

1.7 days
Time To Merge

Jan 1, May 10

232,400
Lines Of Code Changed

811
Commits

77
Contributors

811
Commits
Horovod - Dev Stats - Jan to Dec 2020
Distributed training framework for TensorFlow, Keras and PyTorch

100.31K Lines Of Code Changed
540 Commits
62 Contributors

313 Submitters
151 Open Issues
517 Issues
517 Issues Tracked

44.45K Lines Of Code Changed
184 Commits
22 Contributors
Companies hosting projects in LF AI & Data

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

- LF AI & Data Summit
- LF AI & Data Day
- Project Workshops
Joining LF AI & Data is Easy!

General Inquiries:
info@lfaidata.foundation