LF AI & DATA

LF AI & Data Foundation

Ibrahim Haddad, PhD
Executive Director

Decentralized innovation. Built on trust.

The Linux Foundation provides a neutral, trusted hub for developers and organizations to code, manage, and scale open technology projects and ecosystems.

900
open source projects >

3M+
developers trained >

777K
developers contributing code >

51M
lines of code added weekly >

17K
contributing organizations >

70+
upcoming events >

LF AI & DATA

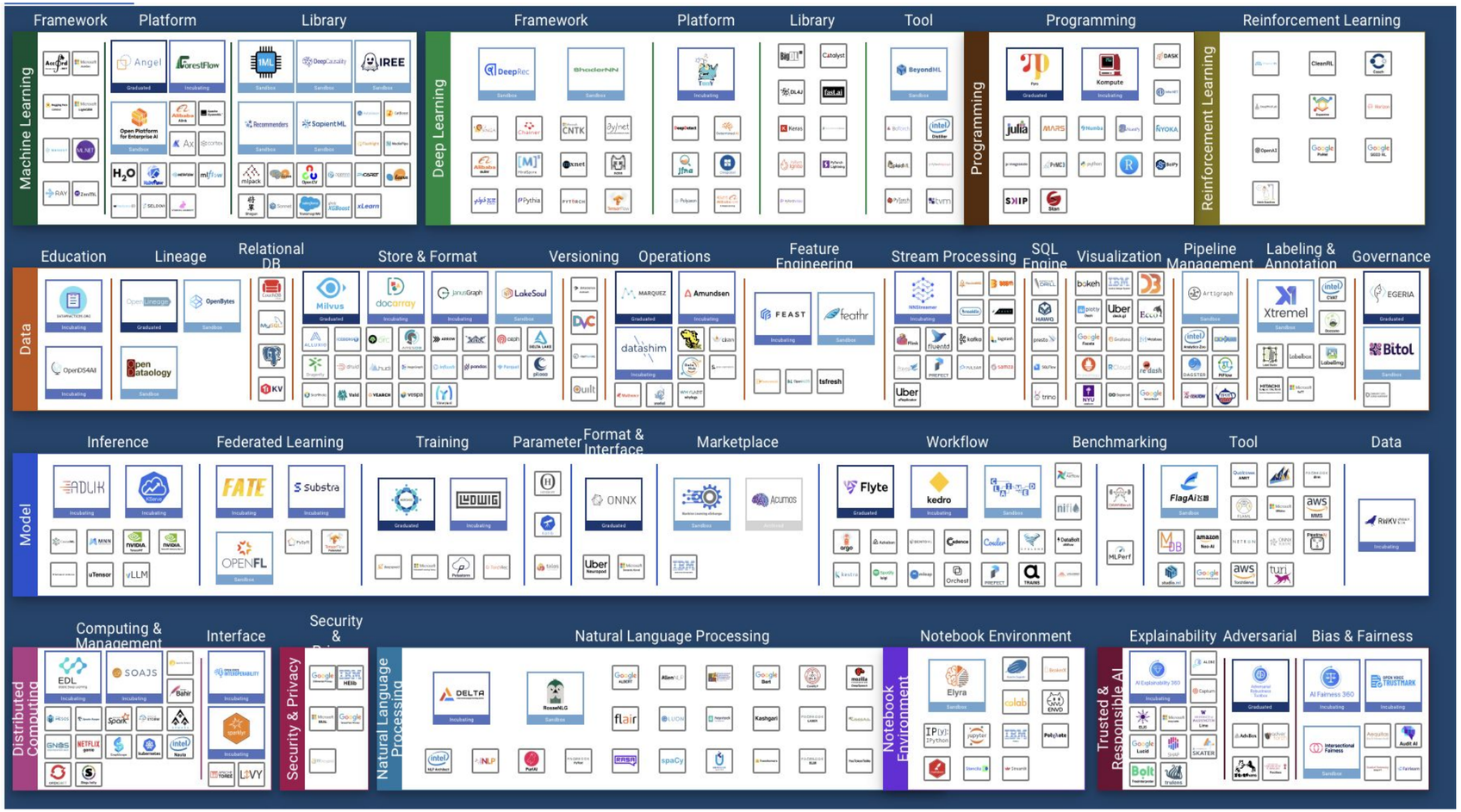
Who is LF AI & Data?

The **LF AI & Data** is a global not for profit foundation that hosts critical components of the global AI & Data technology infrastructure. It brings together the world's top developers, end users, and vendors to identify and contribute to the projects and initiatives that address industry challenges for the benefit of all participants.

Evolution & Mission

Our mission is to build and support an open AI community, and drive open source innovation in the AI, ML, DL and Data domains by enabling collaboration, sharing best practices, supporting development efforts, and the creation of new opportunities for all the members of the community.

A growing ecosystem: The barrier to entry in AI is lower than ever before, thanks to open source software



LF AI & DATA Landscape

359+ Projects

3.3M+ GitHub Stars

100K+ Developers

200+ Founding Org

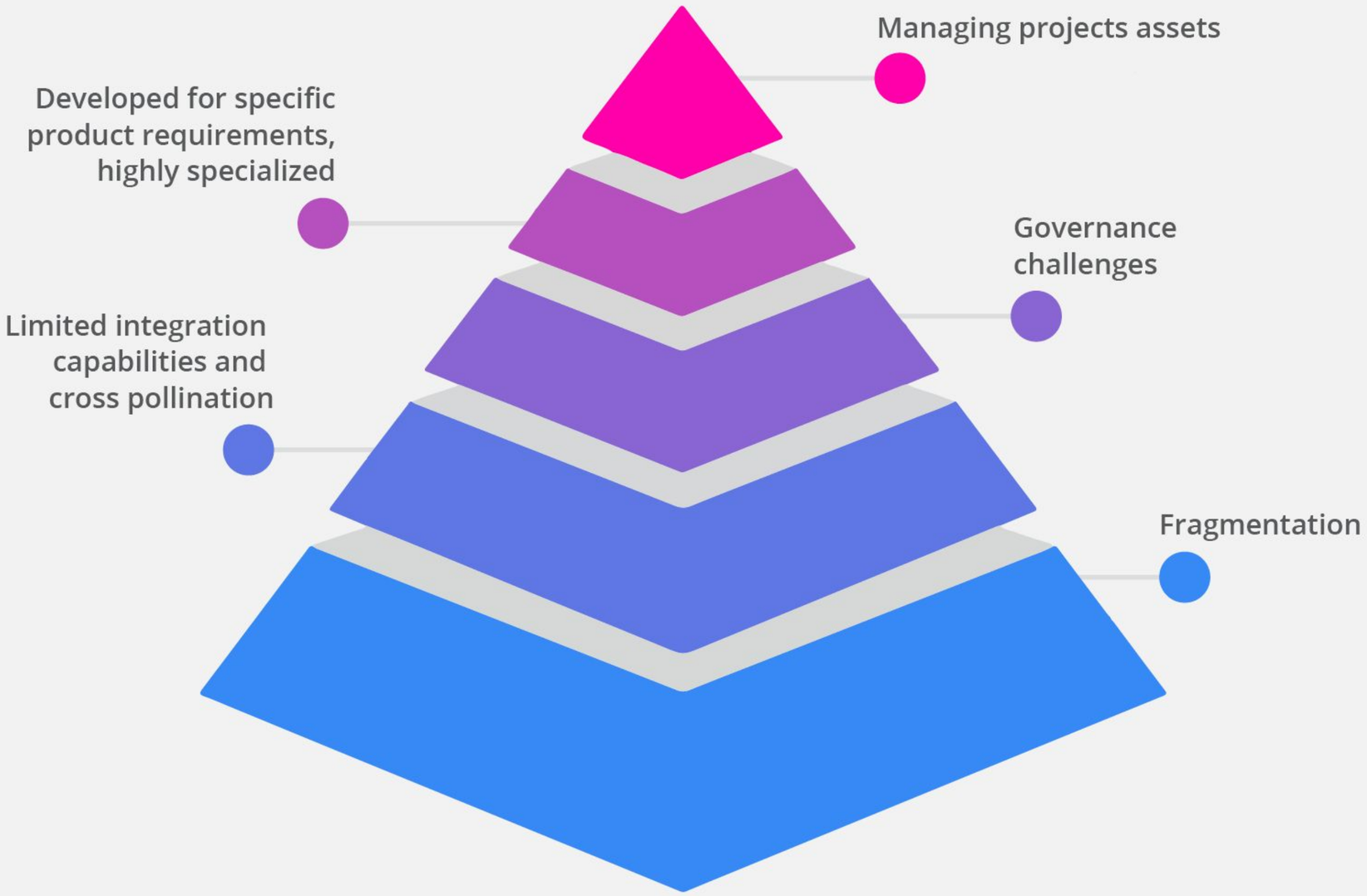
600M+ LoC

1M+ LoC / Week

1000s of Contributing Orgs

LF AI & DATA

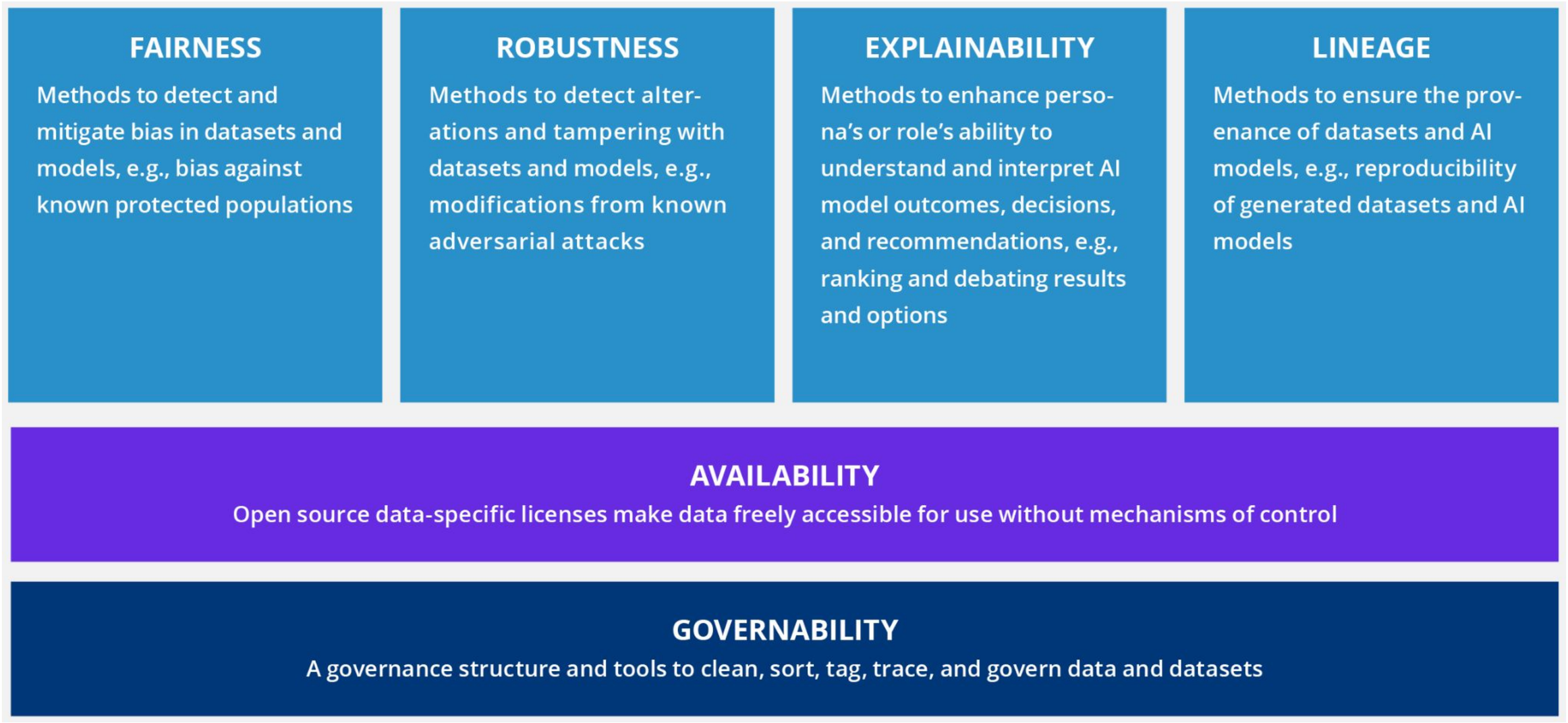
But with challenges…



Open source libraries, frameworks, platforms, and tools are a two-way street: they make AI accessible to everyone, and companies benefit from the collective efforts of the community which helps accelerate open source AI applied R&D.

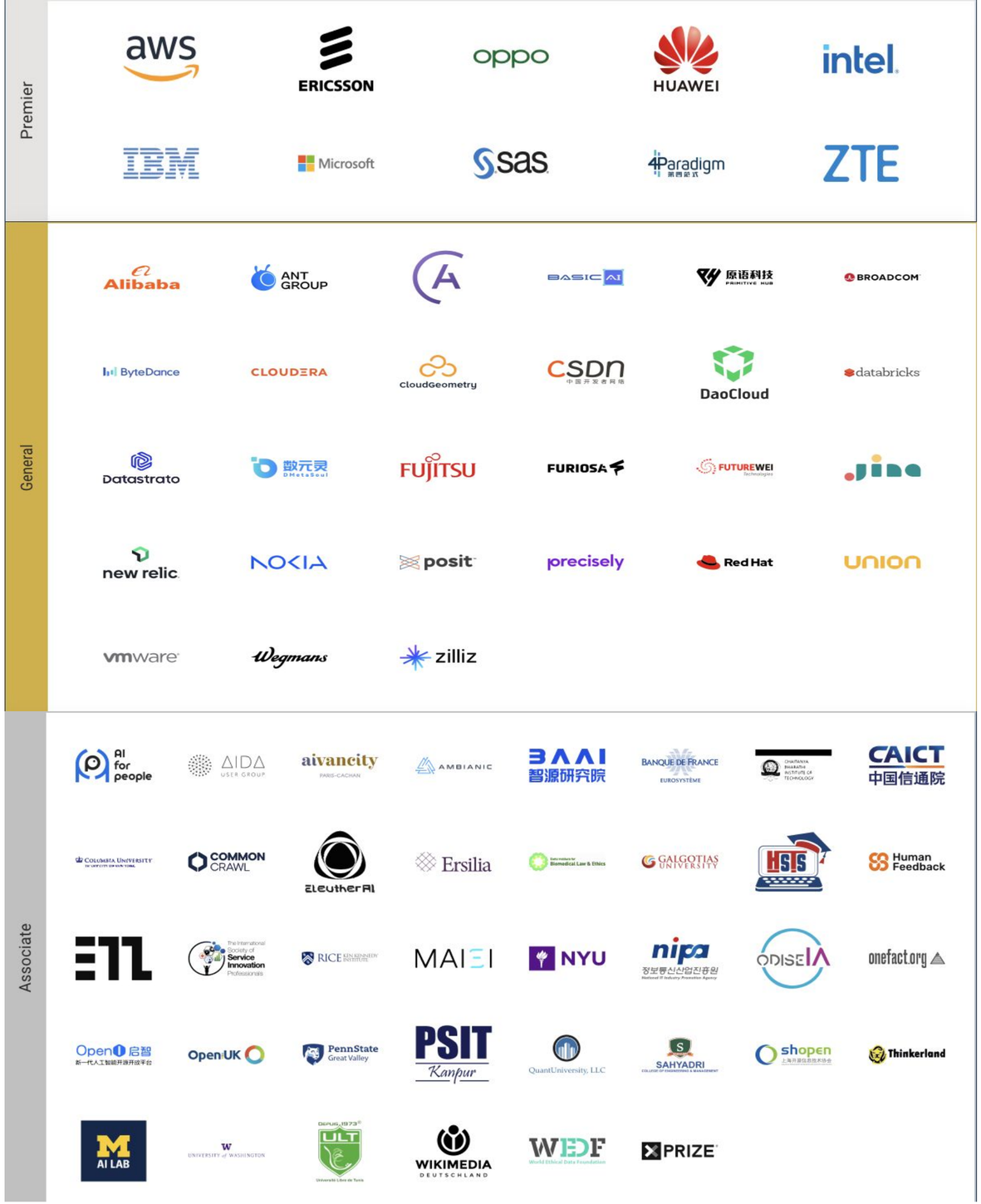
Transparency	**Open Governance**	**Neutrality**	**Openness**	**Long term sustainability**	**Collective Testing**
High Value	**Lower Cost**	**Faster TTM**	**Better Interoperability**	**Access to Talent**	**Ease of Integration**

Open source benefits for AI & Data



Members

February 27, 2024



Committees and Sign-Up

Foundation Governance

- Governing Board
- Outreach
- Legal
- Strategy
- Budget

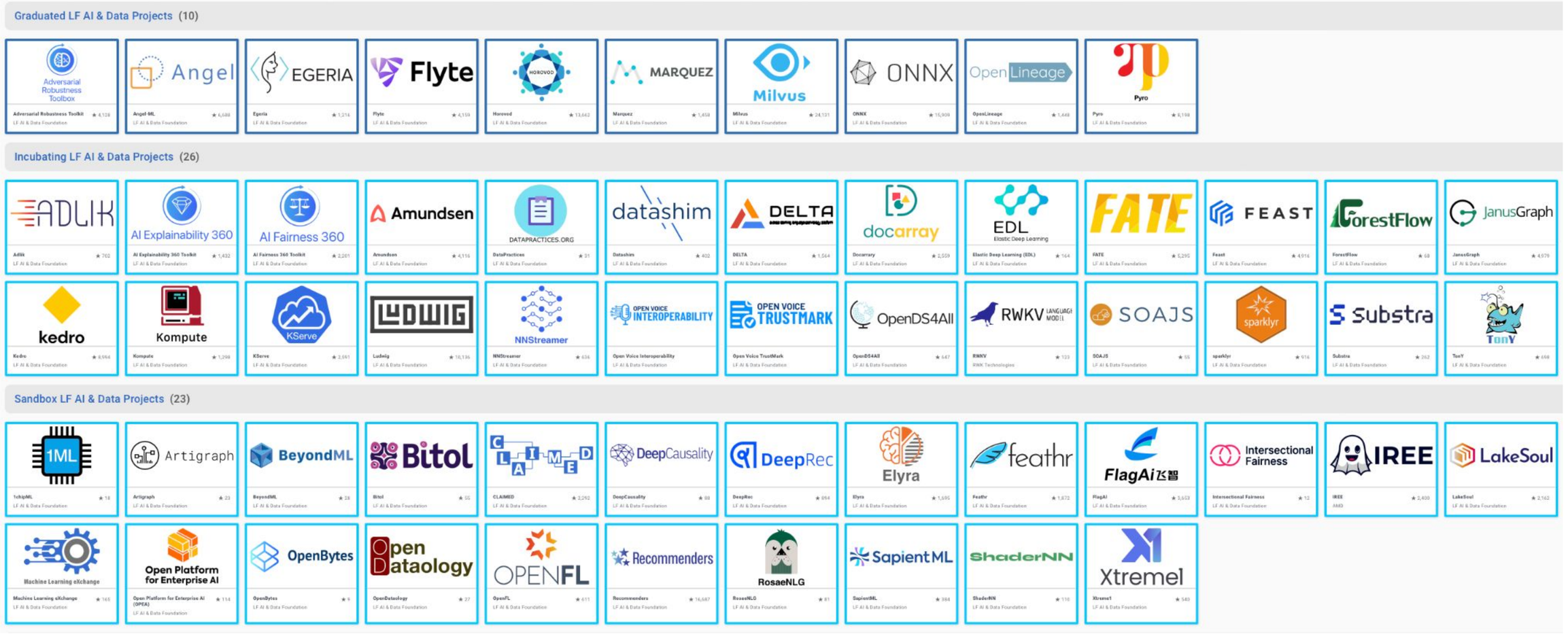
Technical Coordination

- Technical Advisory Council
- DataOps
- Trusted AI
- BI & AI
- DataOps
- ML Ops
- MLSecOps

Generative AI Commons

- Model Workstream
- Data Workstream
- Framework Workstream
- Education Workstream

Projects Hosted in LF AI & Data



Hallmarks of a successful open source project

- **Thriving community:** Projects should have (or be on track to building) an active, broad-based community.
- **Well-tested:** Users should be confident that ecosystem projects will work well with the project, and include support for CI to ensure that testing is occurring on a continuous basis.
- **Clear utility / function:** Users should understand where each project fits within the project's ecosystem and the value it brings. Enables easy curation.
- **Permissive licensing:** Users must be able to utilize ecosystem projects without licensing concerns (OSI-approved licenses)
- **Easy onboarding:** Projects need to have support for binary installation options, clear documentation and a rich set of tutorials.
- **Ongoing maintenance:** Project authors need to be committed to supporting and maintaining their projects.
- **Neutral hosting for project assets:** All project's assets are hosted and managed by a neutral party.

Why should you participate in open source AI development?

- **Alignment with upstream** development for projects you depend on
- **Minimize technical debt** and reduce development costs
- **Product enablement** via project contributions (features, security, performance, etc.)
- **Influence technology direction** via active participation in the projects
- **Access to talent**, diverse and specialized
- **Enhanced interoperability** and compatibility: OSS tends to adhere to standards, which can lead to better interoperability with other systems and reduce the risk of vendor lock-in.
- **Faster innovation** and development cycles: Active participation in open source communities will enable you to tap into a global pool of talent and resources, accelerating the development cycle of its own software products
- **Boost your reputation**, demonstrating a commitment to collaboration and innovation
- **Long-term sustainability** of the projects you depend on
- **Engage with communities** unlocks valuable partnerships and collaborations with other organizations, potentially opening new business opportunities

Open source and data

Data licenses

Open source data specific **licenses** such as CDLA make data freely accessible for use without mechanisms of control.

Data governance

Open source offers a **governance structure** and tools to help govern data and data sets.

Diverse data sets

Open sourcing high-quality datasets allows others to reproduce and validate research, compare models, and develop new approaches. The availability of **diverse and representative open source datasets** is essential for training robust and unbiased models.

Data tools

Open sourcing custom code or scripts for data preprocessing tasks is valuable to the community to **understand used data preparation techniques and reproduce results**.

Key Stats

75 Members	**65** Projects	**9** Committees	**650+** Contributing Organizations
22.58 M LoC created since 2018	**250,000+** GitHub PRs	**100,000+** Contributors across all projects	**25+** Integrations

Code Stats

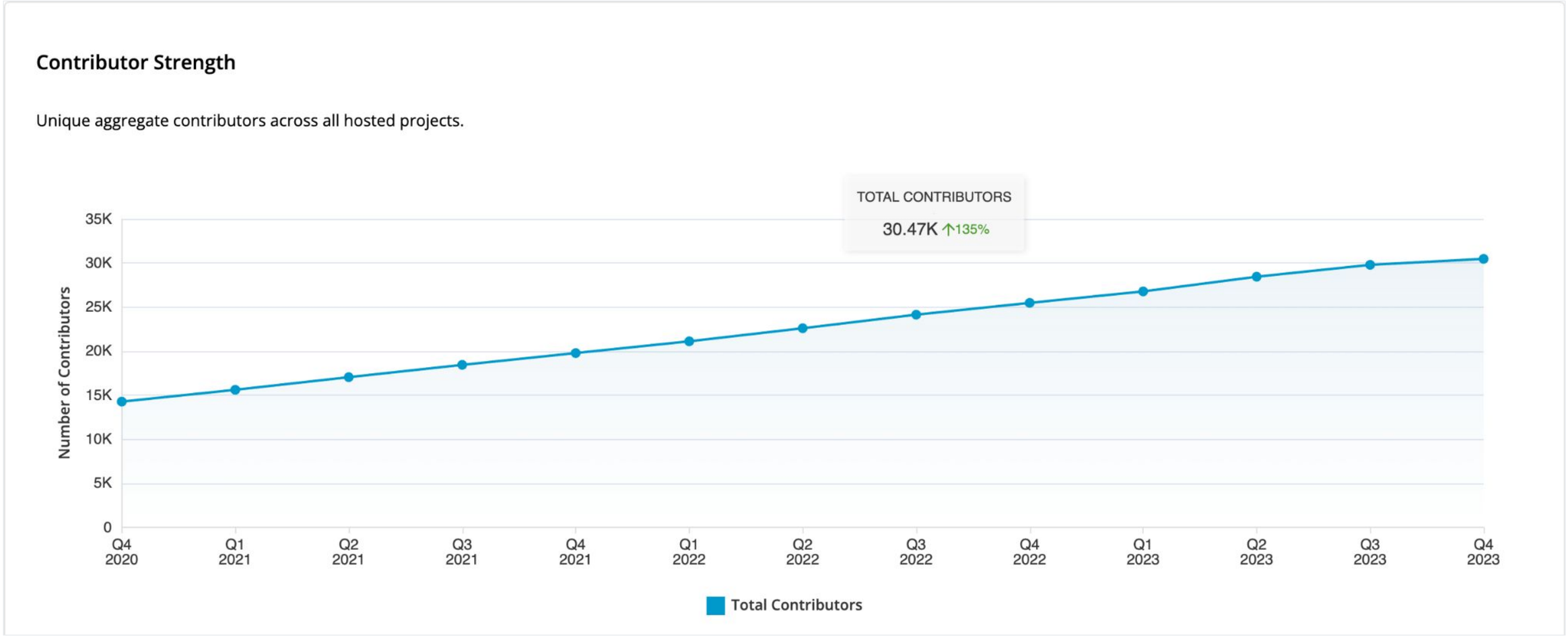
May 19, 2023, Source: LFX Insights

Across **297 total repositories**, an average of 35.64K LOC were added per repository on a weekly basis.

→

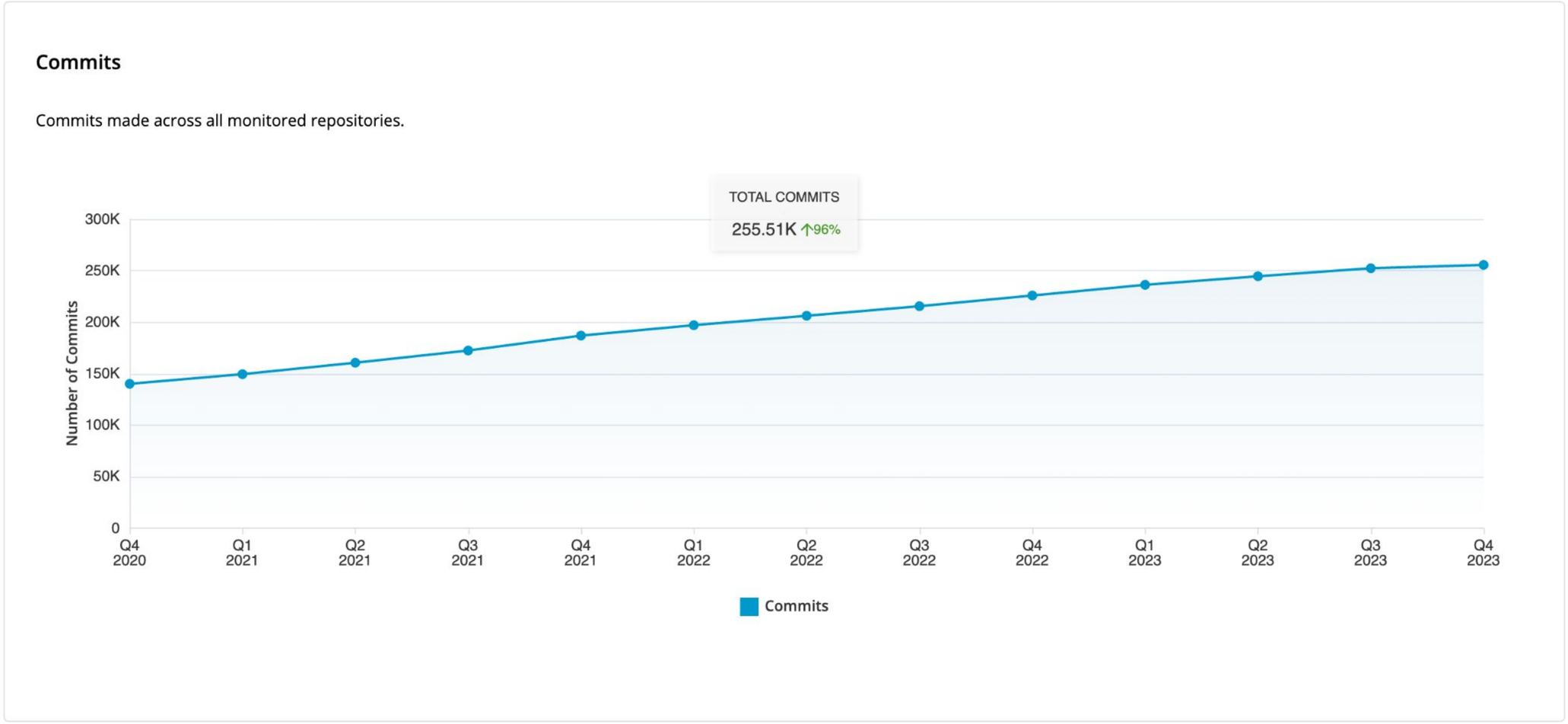
An average of **275.20M LOC** were added in the last 5 years.

Total Number of Contributor

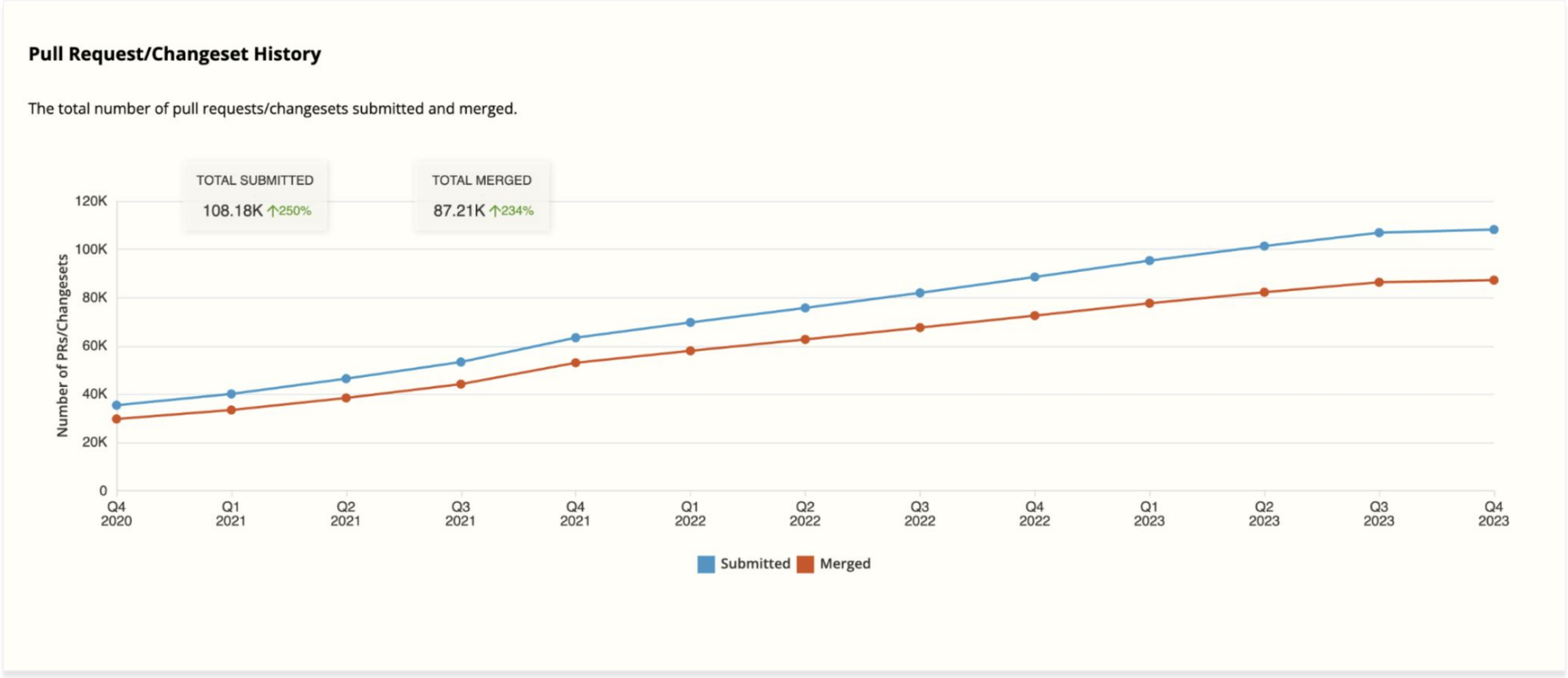


Our programs are enabling our projects to increase the contributions from existing developers and welcome new developers into projects at an unprecedented rate.

Commit Growth

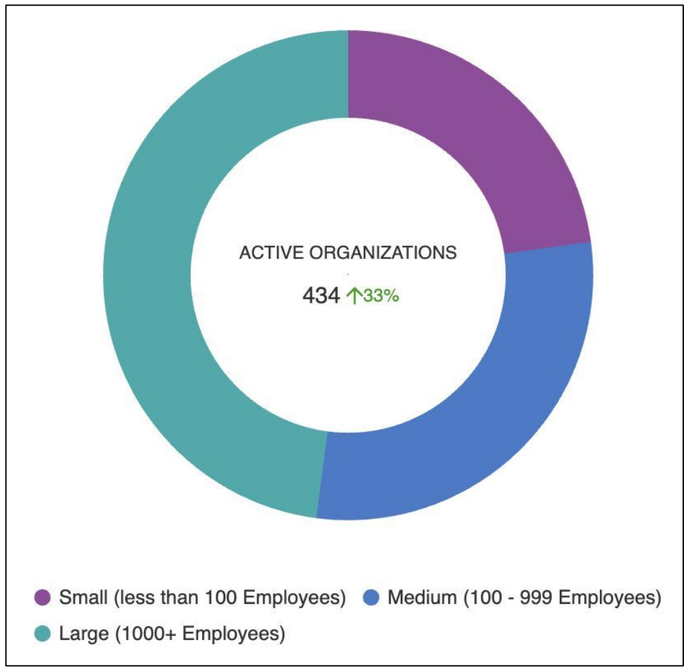


PR / Changeset History



Organization Engagement

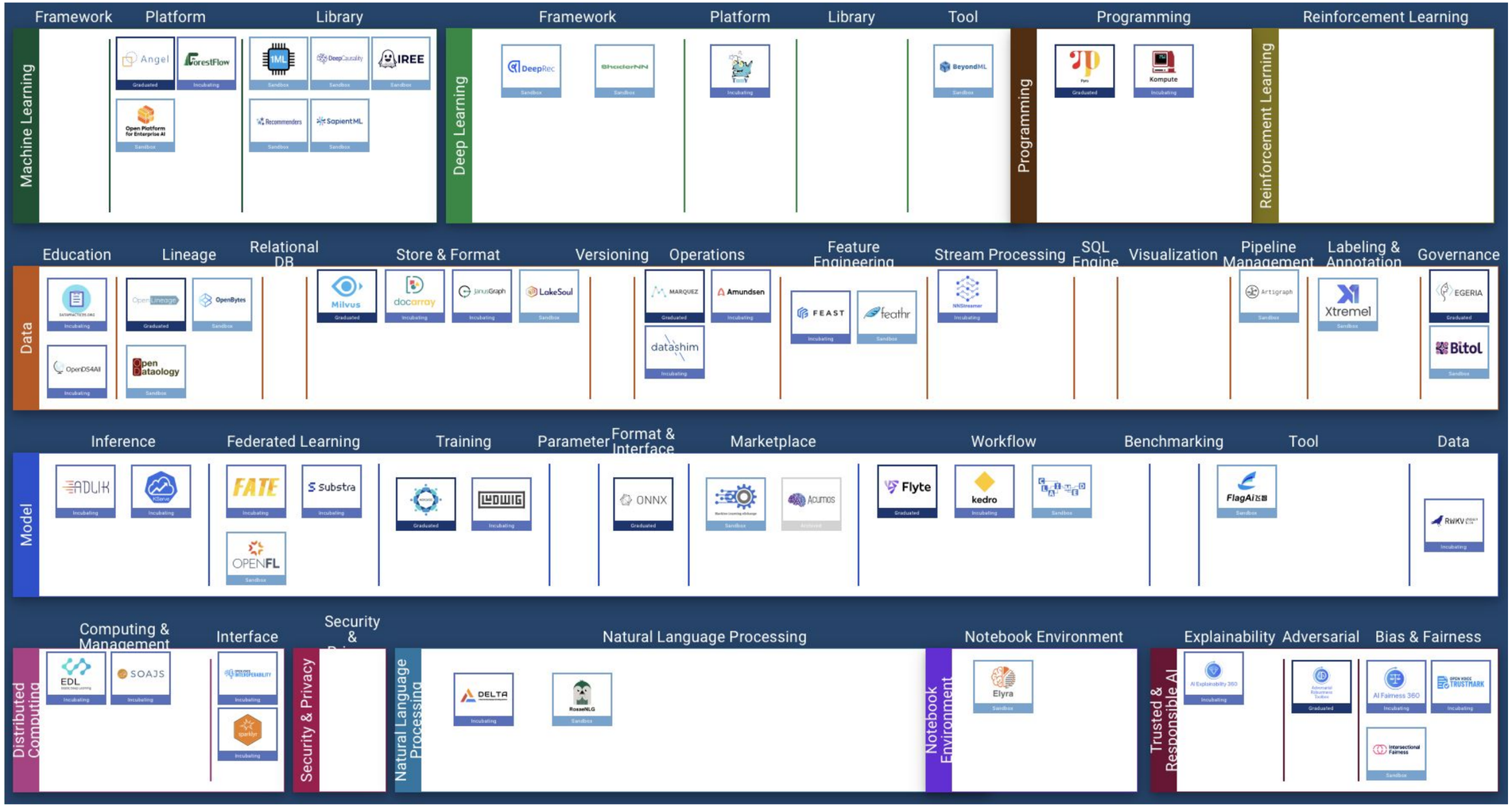
A total of **658 organizations** participated in the code commits during the past 5 years.



434 organizations participated in the code commits in 2023

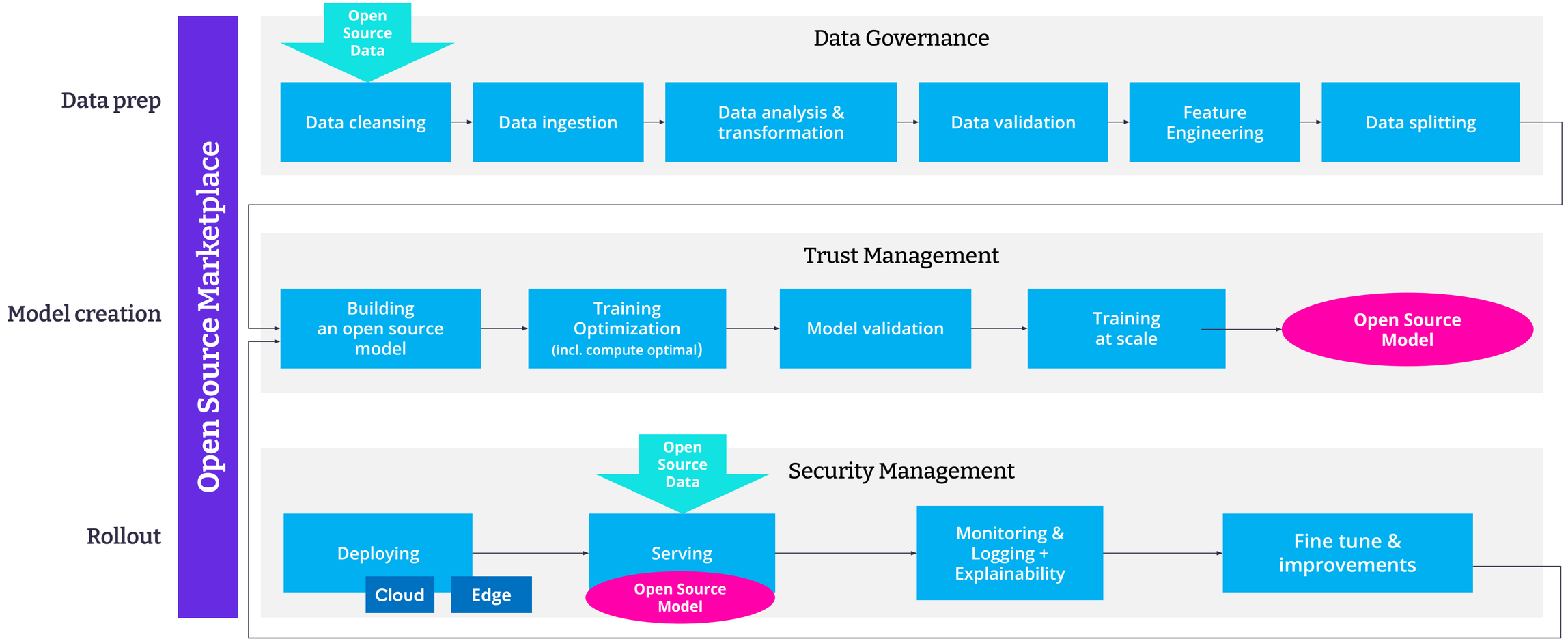
Over 14% of the ecosystem's key projects depend on LF AI & Data

for their open governance, safe haven for their assets, infrastructure, and enabling marketing, legal and event services, and staff that is eager to help and support the communities of these various projects



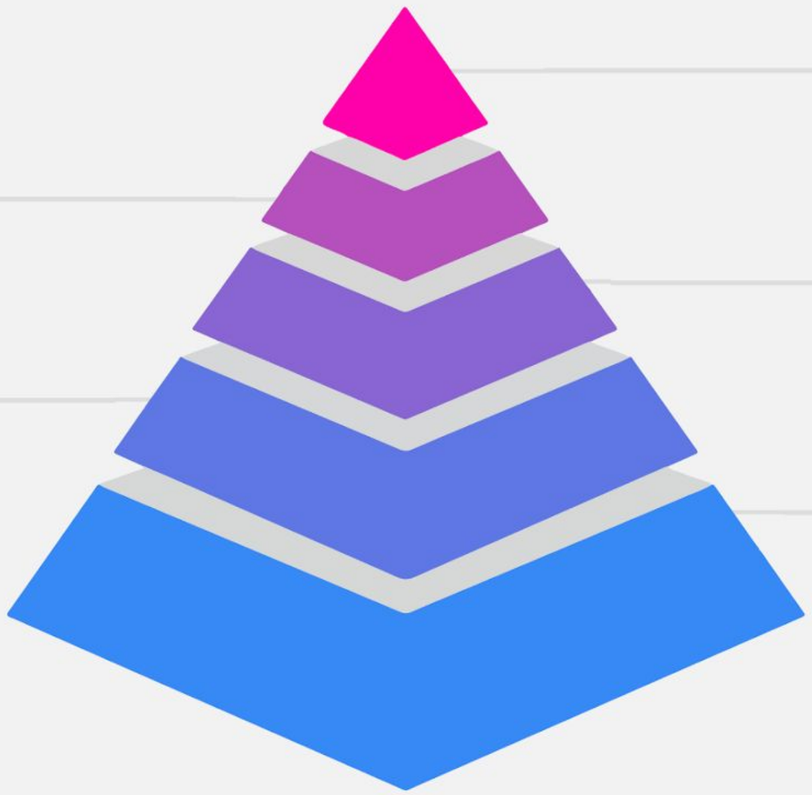
Industry Leaders Host Projects in LF AI & Data

With 50 Hosted Technical Projects, LF AI & Data offers all required open source software and elements required to build and manage an end-to-end ML Workflow



LF AI & DATA

LF AI & Data is addressing key ecosystem challenges



Supporting functional growth of hosted projects

Fostering technical integration and collaboration across projects

Providing a safe haven and managing technical and non technical project assets

Not-for-profit host & technical open governance

Minimizing fragmentation, rallying industries around a set of leading projects

Support Programs Available to Hosted Projects

NEUTRAL HOSTING

A neutral home for an open source project increases the willingness of developers from software companies, startups, academia, and elsewhere to collaborate, contribute, and become committers.

DEDICATED STAFF

Projects have access to full-time staff (executive director, program manager, project coordinator) who cultivate the maturity and adoption of open source AI and data projects

TRAINING AND CERTIFICATION

We develop training classes and, through the Linux Foundation, can execute and launch certification programs in support of hosted projects.

EVENTS MANAGEMENT

Events are part of LF AI & Data's core strategy to help projects build a community and accelerate knowledge-sharing and integration. Many LF AI & Data projects have their own events.

DEV-FOCUSED OPERATION

Services include IT infrastructure, release management, IT ops, support, security audits, and a host of tools (FOSSA, LastPass, Slack, Synk, Zoom, etc.).

MENTORSHIP

Members of the LF AI & Data technical advisory committee and leaders of graduated projects are available to support and mentor new projects.

LEGAL SERVICES

We help projects navigate licensing requirements, IP regimes, trademark management, compliance scans, export control filings, and developer certificate of origin or contributor license agreement integration with GitHub, etc.

MARKET SERVICES

We offer a wide range of marketing services to increase project awareness, project adoption, and the number of contributors.

DESIGN AND AESTHETICS

Our in-house team provides graphic design resources for new logos, websites, and website refreshes or enhancements.

PROGRAM MANAGEMENT

We have decades of experience in program management of open source projects. We bring best practices to all LF AI & Data hosted projects.

LFX PLATFORM EXPERIENCE

This Linux Foundation product offers a set of integrated tools for project insights, security, easy contributor license agreements, crowdfunding, member engagement, and more.

Strong Project Participation from China

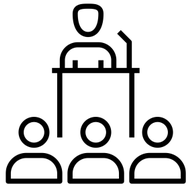
Hosted Projects

Organizations Hosting Projects

Focused Developer & Community Marketing Effort

MarComm services to support community and ecosystem engagement

- Events and promotion
- Project graduation
- Project incubation
- Project releases
- Project cross collaboration



Events and community ecosystem building services

- OSS Summits
- KubeCon
- AICON + GOTC
- AI.Dev
- LF AI & Data Days
- Kubernetes & AI

Why Launch Generative AI Commons in the LF?

Not-for-profit

A non-profit that hosts and promotes the collaborative development of open source projects (sw, hw, standards)

Neutral

Trusted neutral foundation hosting projects and their IP, increasing the willingness of organizations to collaborate & contribute

Credible

3000+ members, 1000+ open source projects, launching cross industry efforts, leading oss technologies in all industry sectors

Open Governance

Open, transparent and fair governance model across all efforts and projects in the Linux Foundation and its umbrella foundations

Available Support Programs

Growth

Build momentum within the LF community of thousands members

Financial Support

Raise and manage funds for the projects or launched efforts

Experience

A team of professional dedicated staff who share deep open source experience

Asset Manager

Manage project/community assets with deep expertise at large scale

Ecosystem Builder

Building ecosystems across oss, open hw, standards, and data sharing

Credible Entity

In launching cross industry efforts, leading oss technologies in all sectors

Resources

Events, PM support, community, LFX Platform, marketing, training & cert, IT, Creative, market services

Legal Services

Trademark management, license compliance, export control filings, DCOs, CLAs, etc.

Generative AI Commons - Initial Workstreams

Emerging LLM Architectures

- Curation of oss projects
- Security
- Privacy
- Trusted & responsible AI

Model-Centric

- Hosting models
- Collaborating on building foundational models
- Hosting Gen AI tools
- Application framework tools

Data-Centric

- Hosting and curating open source datasets
- Hosting Data supporting tools (db, annotation, labeling, workflow, etc.)

Education & Outreach

- Share knowledge and expertise
- Provide thought leadership
- Training and certification
- Share best practices

Openness Framework

A framework to evaluate the openness of ML models by measure of how many of its elements are made available under an open source license.

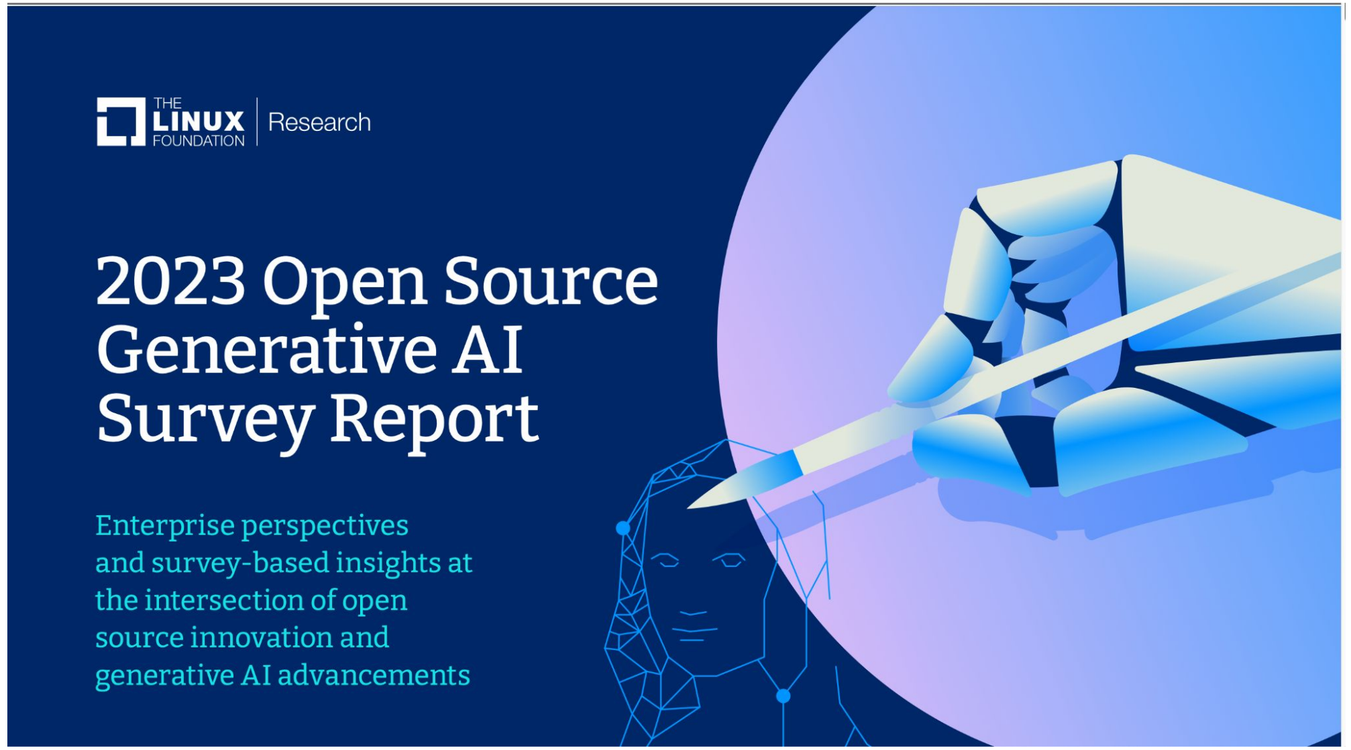
Don’t believe calls to “pause” or “restrict” open source AI

- It didn’t work with cryptography it won’t work here
- Security through obscurity doesn’t work
- Regulatory capture of open source is anti-innovation and anti-competitive
- The market is already addressing AI concerns this with responsible AI market solutions

Gen AI Survey Report - December 2023 - Overview

The State of Generative AI and Open Source Innovation in Enterprises

- A global survey conducted by LF AI & Data in collaboration with LF Research
- Delves into GenAI's integration into businesses and its openness.
- Explores findings highlighting sustainable and ethical development.
- Focuses on AI systems like large language models (LLMs) and usage of open-source databases, applications and frameworks in enterprise and industry.



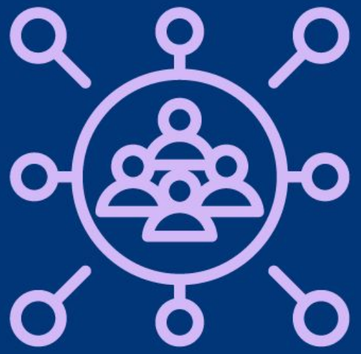
Read Report

Download report, images and infographics.

LF AI & DATA

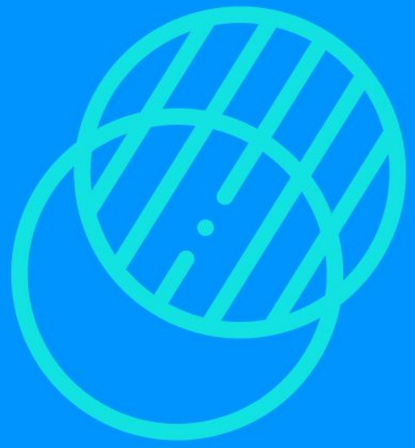
2023 GENAI REPORT

Open source GenAI is considered better at supporting collaboration, innovation, and ease of integration over proprietary solutions, according to our respondents.



2023 GENAI REPORT

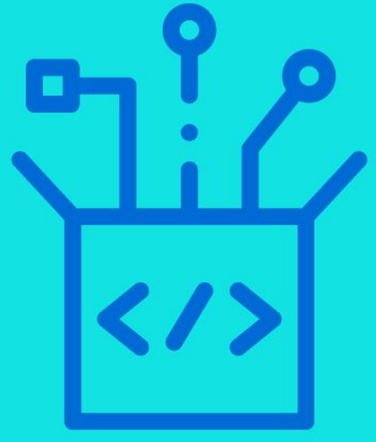
Open source GenAI leads to increased data control and transparency, according to 69% of respondents.



2023 GENAI REPORT

Neutrality is a key aspect of GenAI governance, according to almost all of our respondents (95%).

2023 GENAI REPORT



Openness is important. 63% of respondents are extremely or moderately concerned by the openness of GenAI systems their companies are using or developing.

A Year in Review

+11
Projects

+22
Members

+7500
New Contributors

Gen AI Commons

40+ companies
100+ participants

AI.Dev

Dec 12-13
San Jose, CA, USA

Hosting Process

Technical Advisory Committee

- **The TAC serves a coordination role:**
  - Votes on new projects joining LF AI & Data, as well as on promoting projects across incubation stages
  - Facilitates communication and fosters collaboration across hosted technical projects
  - Communicates needs and requirements of the projects to the Governing Board
  - Onboards new projects, assists in progression of existing projects, and reviews projects annually
  - Defines and maintains the technical vision for the LF AI & Data Foundation
  - Creates a conceptual architecture for the projects, aligning projects, promoting, removing or archiving projects
  - Defines common practices to be implemented across LF AI & Data projects
- **Meetings via conference calls take place every 2 weeks, are recorded and open to the general public**
  - https://wiki.lfai.foundation/pages/viewpage.action?pageId=7733341

Project Incubation levels

Sandbox

- Any project that intends to join LF AI & Data Incubation in the future and wishes to lay the foundations for that.
- New projects that are designed to extend one or more LF AI & Data projects with functionality or interoperability libraries.
- Independent projects that fit the LF AI & Data mission and provide the potential for a novel approach to existing functional areas (or are an attempt to meet an unfulfilled need).

Incubation

Sandbox requirements plus:

- Have 2+ organizations actively contributing to the project.
- Have a defined Technical Steering Committee (TSC) with a chairperson identified, with open and transparent communication.
- Have a sponsor who is an existing LF AI & Data member.
- Have at least 300 stars on GitHub.
- Have achieved and maintained a Core Infrastructure Initiative Best Practices Silver Badge.
- Have the affirmative vote of the TAC.

Graduate

Incubation requirements plus:

- Have a healthy number of code contributions coming from at least five organizations.
- Have reached a minimum of 1000 stars on GH.
- Have achieved and maintained a Core Infrastructure Initiative Best Practices Gold Badge.
- Have demonstrated a substantial ongoing flow of commits and merged contributions for the past 12 months.
- Receive the affirmative vote of two-thirds of the TAC and the affirmative vote of the Governing Board.
- Have completed at least one collaboration with another LF AI & Data hosted project
- Have a technical lead appointed for representation of the project on the LF AI & Data TAC

Incubation Requirements

Sandbox

- Fit the scope and mission of LF AI & Data
- Have an OSI-approved license.
- Have a sponsor who is an existing LF AI & Data member. Alternatively, a new organization would join LF AI & Data and sponsor the project's incubation application.
- Have an open and documented technical governance. The LF team can help set this up.
- The project's founders adopt an open governance model documented in a Technical Charter for the project, and execute the Project Contribution Agreement transferring the project's assets to the LF.

Incubation

- Have at least three organizations actively contributing to the project.
- Have a defined Technical Steering Committee (TSC) with a chairperson identified, with open and transparent communication.
- Have reached a minimum of 500 stars on GitHub.
- Have achieved and maintained an OpenSSF Best Practices Badge Program (Silver).

Graduate

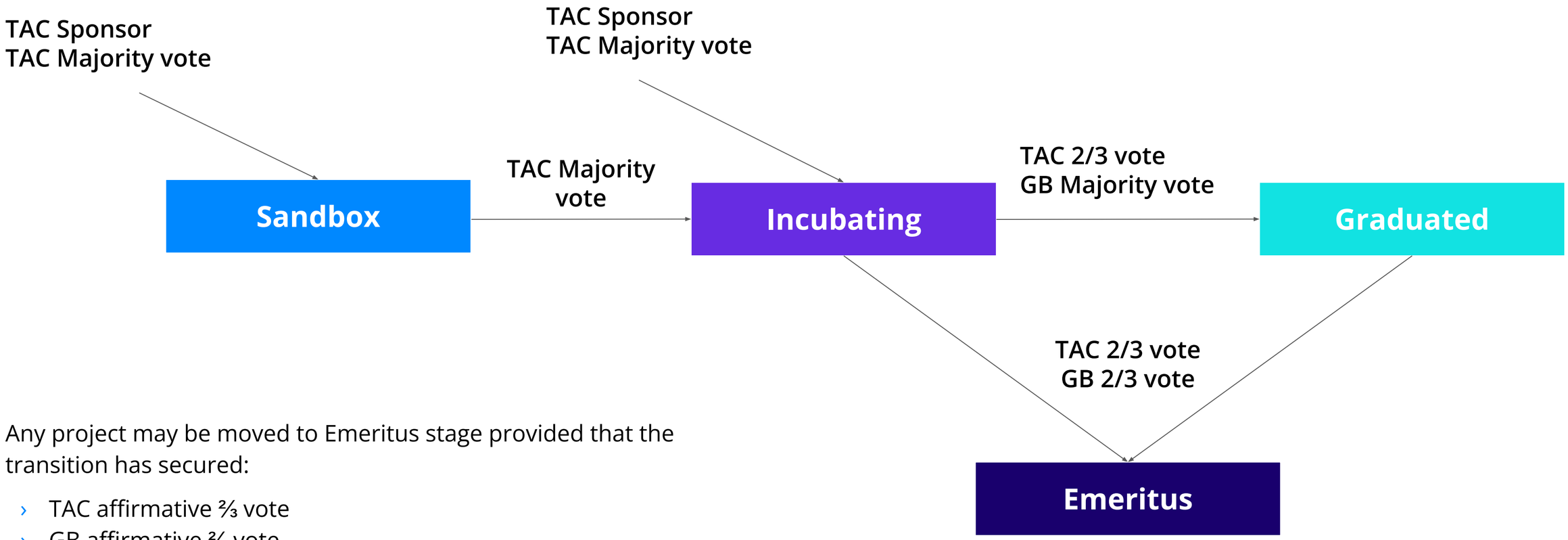
- Have a healthy number of code contributions from at least five organizations.
- Have reached a minimum of 1000 stars on GitHub.
- Have achieved and maintained an OpenSSF Best Practices Badge Program (Gold).
- Have demonstrated a substantial ongoing flow of commits and merged contributions for the past 12 months*.
- Have completed at least one collaboration with another LF AI & Data hosted project

Transitioning from Incubation to Graduation

- The TAC will undertake an annual review of all projects
- Projects in the Sandbox Stage are generally expected to move to Incubation within 12 months from joining the Foundation following an evaluation by the TAC committee
- Projects in the Incubation Stage are generally expected to move to Graduation within 12-18 months from joining the Foundation following an evaluation by the TAC committee

Projects can be provided with an extension of time in their stage
(up to the discretion of the TAC)

Project Lifecycle



Any project may be moved to Emeritus stage provided that the transition has secured:

- TAC affirmative ⅔ vote
- GB affirmative ⅔ vote

Process for proposing a project for hosting in LF AI & Data

1. Decide on a date to present to the TAC and request incubation
2. Ensure that your project implements these recommendations
3. Submit a formal request to incubate the project via a GH PR
4. Prepare deck and share with ED about 10 days prior to the presentation
5. Present to the TAC and get approval
6. Onboard the project with the LF AI & Data team and integrate the project with our service
7. Announce the project becoming hosted in LF AI & Data

Incubation Benefits by Levels

Sandbox

- Neutral hosting of the project's trademark and assets by LF AI & Data.
- Appointment of a TAC member as a project sponsor
- LF AI & Data blog post or similar announcing the project's hosting in the Foundation
- Right to refer to the project as an "LF AI & Data Sandbox Project," and to display the LF AI & Data logo on the project's code repository and web properties
- An initial and regularly scheduled license scan of the project's codebase
- Ongoing source code security scans and reports
- Infrastructure support includes mailing lists, wiki space, slack channel, etc.
- Marketing, communication, and PR support are limited to significant announcements.
- Access to the LFX platform.
- Support of the Foundation staff who are eager to help with the project.

Incubation

Sandbox benefits plus:

- Right to refer to the project as an "LF AI & Data Incubation Project," and to display the LF AI & Data logo on the project's code repository.
- Creative and artwork support covering website, logo, and other required creative work.
- Marketing, communication, and PR support, including project promotion via blog posts, social media, and LF AI & Data website.
- Access to the Bevy platform for community-hosted events.

Graduate

Incubation benefits plus:

- LF AI & Data blog announcement or similar announcing the project graduation, including promotion activities.
- Graduation stage projects may receive support as determined by the Governing Board.
- Right to refer to the project as an "LF AI & Data Graduation Project," and to display the LF AI & Data logo on the project's code repository.
- Voting seat on the TAC.
- Advanced IT infrastructure support (pending board approval).
- Additional ecosystem development opportunities include training courses, certification development, and conformance programs (pending board approval).

LF AI & Data Membership Benefits

LF AI & Data is positioned to bolster cross-company collaboration and interoperability with our neutral IP zone, rich contributor programs, and most importantly, the trust end users place in us.

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 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 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

LF AI & Data Membership Benefits

Marketing Amplification & Brand Awareness

Community Engagement

Thought & Tech Leadership Across Key Technologies

Marketing amplification and brand awareness

Broaden your reach and awareness in the community with LF AI & Data marketing programs.

As a member you can participate in:

LF AI & Data Outreach Committee
Participate in the marketing committee monthly to engage with your peers in the cloud native space.

LF AI & Data Online Programs
Showcase your organization's open source technology by educating new and existing community members about best practices, trends, and new technologies.

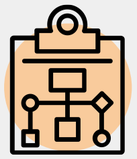
LF AI & Data Blog
Showcase your thought-leadership and industry commentary, as well as share technical walkthroughs for LF AI & Data projects here.

LF AI & Data Summit
LF AI & Data organizes its own community's open source developer summits focused on open source AI and Data technologies.

Public Relations & Analyst Relations Support
LF AI & Data supports members with analyst reports and research highlights.

Community Engagement

LF AI & Data is a constellation of open source projects. Our members leverage many efforts to engage with our project's ecosystems and share their stories. As a member you can participate in:



ML Workflow and Interoperability
Design a reference architecture for end-to-end ML Workflow, provide a reference implementation using hosted projects, and facilitate integration and interop across projects.

Trusted AI
A global group working on policies, guidelines, tools and use cases by industry to ensure the development of trustworthy AI systems and processes to develop them continue to improve over time.

DataOps
Share best practices, and bring technology awareness around DataOps across industries and enable collaboration across LF AI & Data member companies, projects, and external participants.

BI & AI
Integrate the power of AI and BI to make it CI (Cognitive Intelligence) by combing the speed machines accelerate (AI) with the direction intuited by human insight (BI).

LF AI & Data Day
A 1-day event organized multiple times per year focused on specific technical projects and collaboration ongoing in LF AI & Data.

Thought leadership

Members of the LF AI & Data can network and help shape the open source AI and Data market.

As a member you can participate in:

Governing Board Participation
Participate in elections to serve on the Governing Board to oversee the vision of LF AI & Data and work with the TAC.

Engagement with other Foundations and Industry Consortia
Participate in establishing collaborations with other umbrella foundations under the LF and broadly with other industry consortia organizations.

Technical Advisory Council Leadership
LF AI & Data TAC provides technical leadership to the open source AI and Data community, accepts projects into incubation, graduates projects and launches new efforts and collaborations.

Interactive Landscape Placement
Our landscape is a comprehensive view of all open source AI and Data critical projects in the ecosystem. Anyone looking to adopt open source AI and Data projects comes here to review what technologies they should assess and adopt.

LF AI & Data Annual Dues

	Not Yet a Linux Foundation Member	Existing Linux Foundation Member
Premier Member	$120,000	$100,000
General Member	5,000 employees +: $45,000 2,000 - 4,999: $30,000 500 - 1,999: $25,000 Up to 499 employees: $10,000	5,000 employees +: $25,000 2,000 - 4,999: $15,000 500 - 1,999: $10,000 Up to 499 employees: $5,000
Associate Member	Free Limited to academic and nonprofit institutions respectively and requires approval by the Governing Board	

Premier Membership

LF AI & DATA
PREMIER MEMBER

Highest tier of membership – for organizations contributing heavily to open source AI and Data, and bringing their own projects to be hosted at the Foundation. They work in concert with the Foundation team members. These companies want to take the most active role in enabling the open source AI and Data ecosystem.

Premier members are eligible to:

(Enjoy all the benefits of General level, plus;)

- **Hold one (1) guaranteed seat** on the LF AI & Data Governing Board + one alternate (1) representative
- **Appoint one (1) voting representative in any subcommittees** or activities of the LF AI & Data Governing Board
- Receive greater **insight into LF AI & Data strategy and projects** through engagement with the LF AI & Data leadership team. Premier level members have the unique opportunity to **customize their experience** with LF AI & Data. The team will make themselves available to help achieve your strategic goals. We can help with guidance in open source contributions, new market creation, and/or open source project donation. **Have ideas? Just ask!**
- Enjoy most prominent placement in displays of membership including website, landscape and marketing materials.
- Create an individualized press release upon membership announcement with the LF AI PR team.

General Membership

LF AI & DATA
GENERAL MEMBER

Targeted for organizations that want to support of LF AI & Data and our mission. Organizations that join at the General level are deeply committed to using open source technology, helping LF AI & Data grow, voicing the opinions of their customers, and giving back to the community.

General members are eligible to:

- **Participate in elections** between **other General members** to appoint one (1) representative to the **LF AI & Data Governing Board**. Three (3) total General representatives will be elected to represent all General members. Voice your opinions amongst the leaders in the industry and help determine the strategic direction of LF AI & Data.
- Receive **greater insight into LF AI & Data strategy and project roadmaps** through increased engagement with the LF AI & Data Executive Director and staff.
- Create an announcement upon membership announcement with the LF AI & Data PR team.
- Participate in all Marketing, Community, Thought Leadership opportunities.
- Opportunity to **host "LF AI & Data Day"**, including on-demand webinars and livestream.
- Receive **discounts** on LF AI & Data event sponsorships.
- Demonstrate your support for LF AI & Data by displaying your logo on the LF AI & Data website, landscape and in marketing materials.

Associate Membership

LF AI & DATA
ASSOCIATE MEMBER

The Associate membership is a free complimentary membership limited to academic and non-profit institutions.

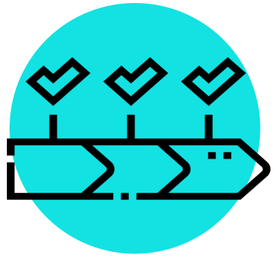
Associate members are eligible to:

- Participate in all Marketing, Community, and Thought Leadership opportunities.
- Identify your organization as a member and display your logo on the LF AI & Data website, landscape and in marketing materials.
- Feature your organization in the quarterly new members announcement.
- Receive discounts on LF AI & Data events' sponsorship.

Challenges and Trends

Global Open Source Trends

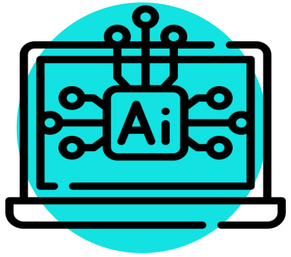
Training & Cert, Talent Shortage



Emphasis on SW Supply Chain

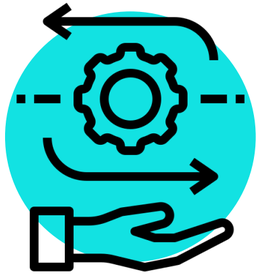
Rise of Adoption of OSPOs

Importance of Software Security



Infusing AI and ML in products and services

Focused Efforts on Ethical AI Practices



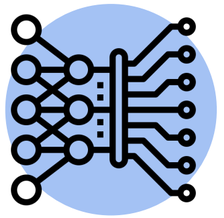
Criticality of OSS to Digital Transformation

Accelerated OSS Adoption in Governments

AI Challenges and Opportunities

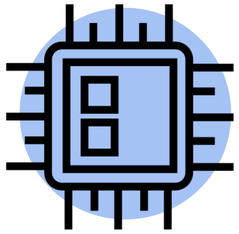
Trusted and Responsible AI

AI on edge devices



Federated learning

Social Impact



Specialized Hardware

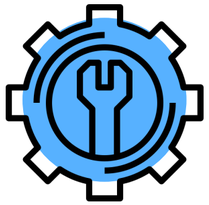
Availability of Skilled Talent

Scalability, Efficiency

Policies, Regulations & Standards

User Acceptance, Building Trust

Data Challenges and Opportunities



Security

Integration

Labeling

Privacy

Storage

Quantity

Quality

Ethical

Diversity

Accessibility

Bias

Lineage

Licensing

Governance

Regulations

Standards

Legislations

Compliance

Deriving value

LLM Trends

1. **Model Size**
   LLMs have been steadily increasing in size and complexity. The shift from GPT-2 to GPT-3 saw a 100x increase in model parameters, and 6x increase from GPT-3 to GPT-4, enabling more nuanced language processing and higher-quality outputs.
2. **Pre-training and Fine-Tuning**
   Pre-training on large-scale datasets followed by fine-tuning on specific tasks has become a common approach in LLM development. This two-step process enhances model performance and allows for transfer learning across domains.
3. **Few-shot and Zero-shot Learning**
   LLMs have shown promising capabilities in few-shot and zero-shot learning, where they can perform tasks with minimal or no task-specific training examples. This trend has implications for rapid prototyping and reducing data requirements.
4. **Multimodal LLMs**
   There is an increasing focus on developing multimodal LLMs that can process and generate text in conjunction with other media types, such as images, audio and video. These models enable more comprehensive and interactive AI experiences.

LLM Challenges

1. **Ethical Concerns**
   LLMs raise ethical concerns related to biases, misinformation amplification, and malicious use. Ensuring fairness, transparency, and responsible deployment of these models is crucial.
2. **Data Privacy**
   LLMs require large amounts of data for training, raising concerns about user privacy and data protection. Striking a balance between data access and privacy is an ongoing challenge.
3. **Dataset Limitations**
   Open source models rely on publicly available datasets, which may be limited in terms of diversity, quality, and potential biases. The availability of diverse and representative datasets is essential for training robust and unbiased models.
4. **Resource Intensiveness**
   Training and deploying large-scale LLMs demand significant computational resources and energy consumption, hindering accessibility and sustainability.
5. **Reproducibility and Benchmarking**
   Comparing and reproducing results across different open source LLMs can be challenging due to differences in architectures, training data, and evaluation metrics. Establishing standardized benchmarks and reproducibility guidelines are essential for transparent evaluation and advancement in the field.

Closed LLMs

- **Lack of Accessibility**
  Closed source models are developed and owned by specific company, and access to these models may be restricted or available only through licensing agreements. This limited accessibility can restrict the ability of researchers and developers to utilize and modify the models.
- **Lack of Transparency**
  Closed source models often lack transparency in terms of their architecture, training data, and fine-tuning techniques. The inner workings and details of these models are not openly shared, making it challenging to understand how they make predictions or generate outputs. This lack of transparency can limit researchers' ability to analyze and improve the models.
- **Lack of Customizability**
  Closed source models may offer limited customization options. The ability to fine-tune or modify the models for specific tasks or domains may be restricted or entirely unavailable. This can limit the flexibility and adaptability of the models to specific use cases.
- **Intellectual Property Rights**
  Closed source models are typically protected by intellectual property rights, such as copyrights or patents. This means that using or distributing these models without proper authorization may infringe on these rights. This can limit the freedom to use and distribute the models for various purposes.

Opportunities – Benefits Open Source LLMs

- **Research and Innovation**
  Open source LLMs provide researchers with valuable resources for exploring new techniques, developing novel applications, and advancing the field of natural language processing.
- **Collaboration and knowledge sharing**
  Open source LLMs promote collaboration by allowing developers to build on top of existing models. Researchers and developers can build upon existing models, share improvements, and collectively advance the field.
- **Transparency**
  Open source LLMs encourage transparency and accountability by allowing researchers to inspect the models and data used for training.
- **Democratize access**
  Open source LLMs help to democratize access to state-of-the-art natural language processing technology enabling developers and organizations with limited resources to leverage state-of-the-art models for their applications.
- **Federated LLMs**
  Open source LLMs support the emergence of decentralized and federated LLMs, allowing for increased privacy and data control.

Open Source & AI - Core Principles to Embrace

We can’t hire all the AI Experts in the world

We need to collaborate with all of them working to address the same challenges as us.

Open Source AI R&D creates significant value

Internal R&D claims portions of that external open source AI R&D and builds innovation on top of it.

We don’t need to create an AI project to benefit from it

We can join an existing open source AI project and excel in it and drive to a leadership position.

What does the remainder of 2024 look like?

- Hosting of new projects providing additional enabling AI technologies
- New collaborations with other LF umbrellas
- Implementing new technical integrations across projects
- Growing the Generative AI Commons
- Driving the Model Openness Framework to production
- Providing new training courses

Join us, become a member!

Incubate your project

info@lfaidata.foundation