Getting Involved Guide

For Current and Prospect Members

April 2022

lfaidata.foundation

LF AI & DATA

Thank you!

Thank you for joining the LF AI & Data Foundation. Your support is vitally important in sustaining the health and innovative momentum of open source AI and Data and the LF AI & Data projects.

We've written this guide to provide you a complete reference to the LF AI & Data community. You will learn how to engage with your communities of interest, all the different ways you can contribute, and how to get help when you need it. If you have suggestions for enhancing this guide, please get in touch with LF AI & Data staff.

We are creating a sustainable open source AI and Data ecosystem that makes it easier to create AI and Data products and services.

Our mission is to build and support an open AI community, and drive open source innovation in the AI, ML and DL domains by enabling collaboration and the creation of new opportunities for all the members of the community

Table of Contents

Organizational Overview

Key Foundation Staff and Resources

Joining LF AI & Data

Hosting Projects Under LF AI & Data

Getting Involved

LF AI & Data Landscape

Organizational Overview

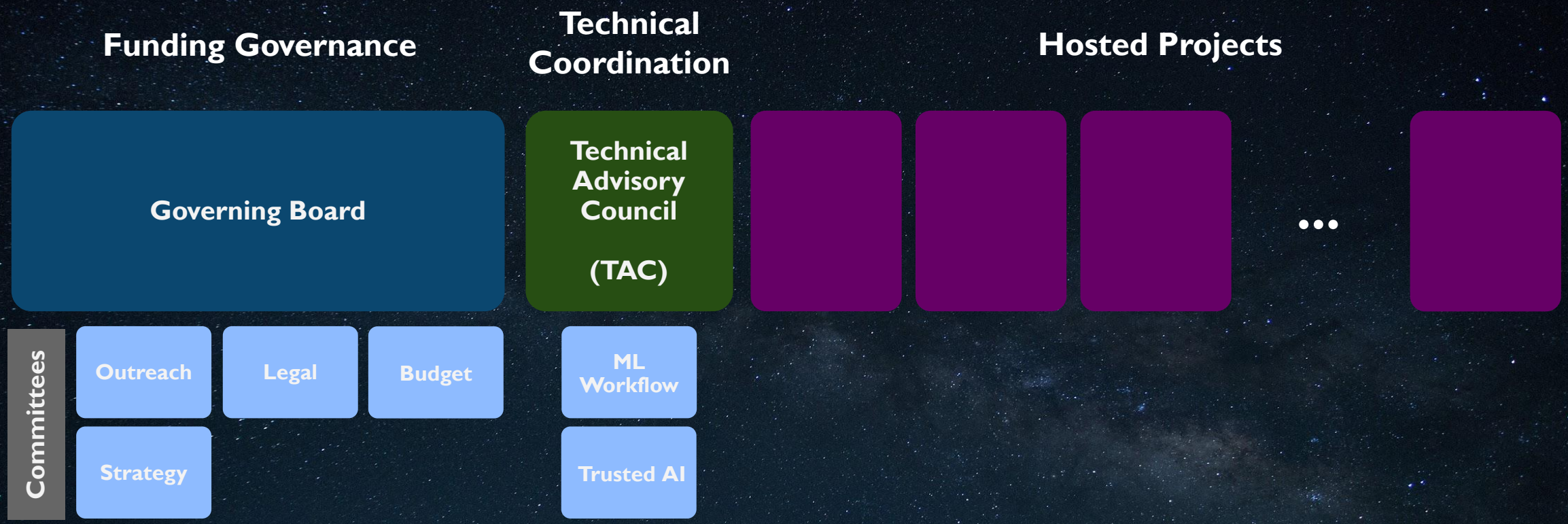
Governance

LF AI & Data is a single funding effort to support technical projects hosted under the Foundation.

Technical projects have their own technical governance following our principle of separation of technical governance from funding governance.

The Foundation's Charter is available from https://lfaidata.foundation.

Structure


Funding Governance
Technical Coordination
Hosted Projects
Governing Board
Technical Advisory Council
(TAC)
...
Committees
Outreach
Legal
Budget
Strategy
ML Workflow
Trusted AI

https://wiki.deeplearningfoundation.org/pages/viewpage.action?pageId=7733341

Governing Board

- Composition:
  - Appointees of Premier Members
  - Representative of General Members (elected annually)
  - TAC Chair (elected annually)
- Responsibilities include:
  - Strategic direction and evolution of the LF AI & Data Foundation
  - Funding the Foundation
  - Approving Graduate projects
- Meets monthly over a conference call

Technical Advisory Council (TAC)

- Membership on the TAC is made up of Premier Member appointees and a representative of "Graduate" projects
- The TAC serves a coordination role:
  - Facilities communication across and among the supported technical projects; and
  - 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 Foundation
  - Vote on approving new projects joining the LF AI as incubation or graduate projects
  - Creates a conceptual architecture for the projects,aligning projects, removing or archiving projects
  - Defines common practices to be implemented across LF AI projects
- Meetings via conference calls take place every 2 weeks, are recorded and open to the general public
  - https://wiki.lfaidata.foundation/pages/viewpage.action?pageId=7733341

Outreach Committee

- Composition:
  - One appointed voting representative from each Premier Member
  - Each General Member may appoint a non-voting representative to observe and participate in the Outreach Committee
- Responsible for community outreach and marketing strategy for LF AI & Data
- Meetings happen monthly over a conference call
  - https://wiki.lfaidata.foundation/display/DL/Outreach+Committee

Budget Committee

- Composition:
  - Representatives of the Governing Board that volunteer to be named participants on the Budget Committee
- The responsibilities of the Budget Committee include:
  - Assisting the Treasurer in preparation of annual budgets that adhere to the principles and guidelines established by the Governing Board;
  - Developing and reporting metrics for the allocation of budget in relation to meeting the priorities of the Governing Board;
  - Reviewing the progress of the Directed Fund against the annual budget;
  - Preparing forecasts for future financial needs of the Directed Fund; and such other matters related to finance and the financial operation of the Directed Fund as may be directed to the Budget Committee by the Governing Board

Legal Committee

- The Legal Committee consists of members of the Governing Board that wish to participate on the Legal Committee together with their company's legal counsel

- Responsible for the creation of recommendations to the Governing Board in response to questions submitted to the Legal Committee by the Governing Board or the TAC

LF AI Elected Leadership

Governing Board (GB)
Chair: Junping Du, Huawei

GB General Member Representative
Rep: Vacant

Treasurer
Vacant

Technical Advisory Council
Chair: Nancy Rausch, SAS

Outreach Committee
Chair: Hu Xiaoman (Charlotte), Huawei

ML Workflow & Interop Committee

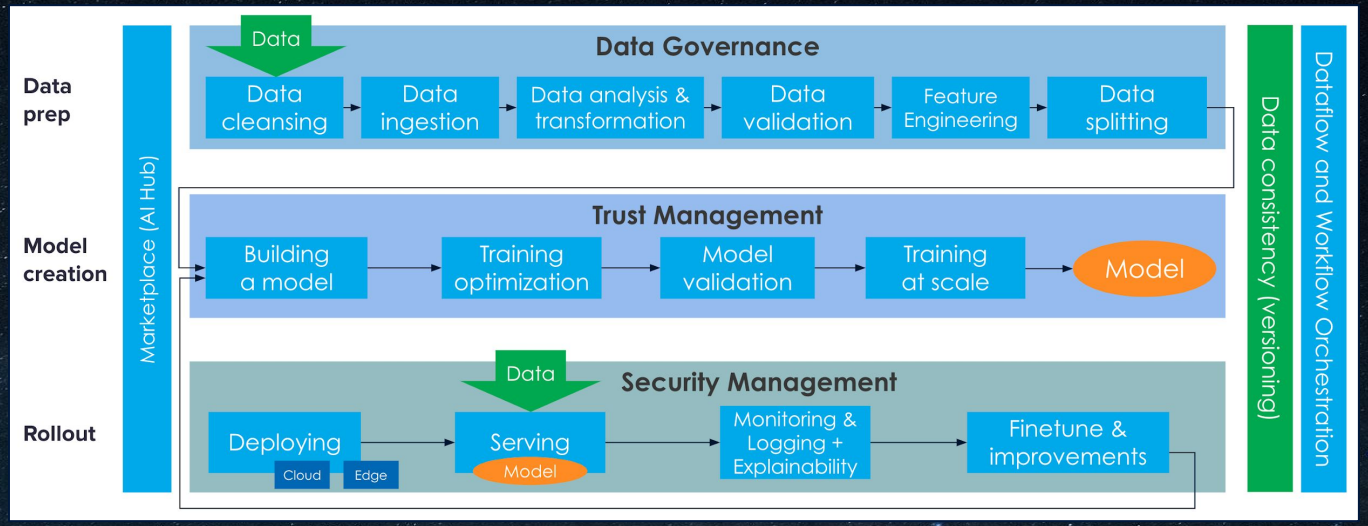
Participants

- Amdocs
- GOC
- IBM
- IFC
- ITU
- Orange
- Seldon
- Siomtics
- SyLabs.io
- TechM
- Tencent
- Verizon
- WBG

Anyone can participate

Scope

1. **Provide a reference ML workflow architecture and implementation**



2. **Increase interoperability and integration across technical projects**

https://wiki.lfaidata.foundation/display/DL/ML+Workflow+Committee

Trusted AI Committee

Participants

- Amdocs
- ATT
- Ericsson
- ETL
- Huawei
- IBM
- Orange
- TechM
- Tencent

Open to members.

Scope

1. Define policies, guidelines, tooling and use cases by industry to create responsible and trusted AI
2. Create a badging process for open source projects that meet the Trusted AI policies/guidelines as defined by LF AI

Two main efforts:

1. Principles: Define ethics, guidelines and principles for Trusted AI
2. Use Cases: Define and implement trusted AI use cases within different AI projects and industry domains

https://wiki.lfaidata.foundation/display/DL/Trusted+AI+Committee

Key Foundation Staff and Resources

LF AI & Data: Key Contacts and Resources

These contacts can answer any questions you may have about how to engage with any aspect of the Foundation, and will provide introductions to other community members and activities as you choose to get more deeply involved.

Staff

Ibrahim Haddad, Executive Director

Jacqueline Z Cardoso, Senior Program Manager

Erin Thacker, Program Coordinator

Contact Us

https://members.lfaidata.foundation/

Resources

Web Site	https://lfaidata.foundation
Landscape	https://landscape.lfai.foundation
GitHub	https://github.com/lfai
Wiki	https://wiki.lfaidata.foundation/
Artwork	https://github.com/lfai/artwork
Email	info@lfaidata.foundation

Mail Lists
https://lists.lfaidata.foundation/g/main/subgroups

Joining LF AI & Data

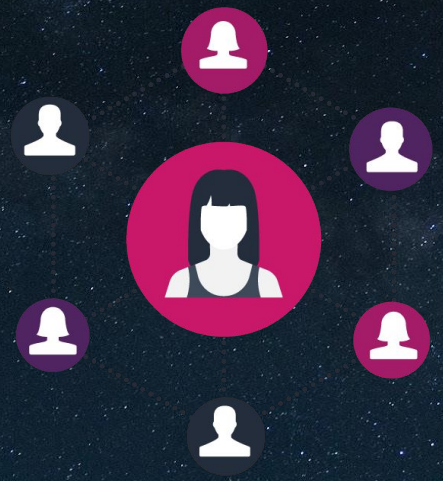
(For prospect members)

Join the LF AI & Data Foundation - Membership Benefits

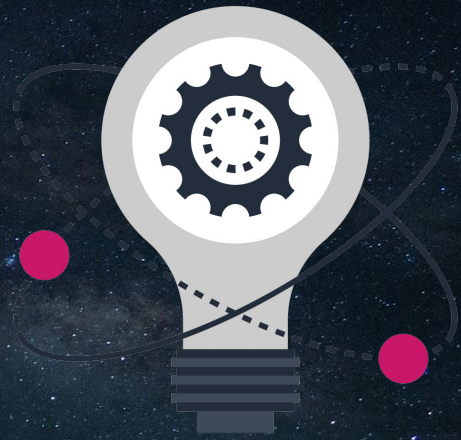
Organizations join LF AI & Data because they want to take an active role in supporting the growth and evolution of the open source AI, ML, DL, and Data ecosystem.

LF AI & Data Membership Provides:

Marketing Amplification and Brand Awareness



Community Engagement



Thought Leadership

LF AI & Data Foundation 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

Highest tier of membership – For organizations who contribute heavily to open source AI, ML, DL, and Data and bring in their own projects to be hosted at the Foundation. They work in concert with LF AI & Data team members. These companies want to take the most active role in enabling open source AI, ML, DL, and Data.

Premier members are eligible to:

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

- Appoint one representative to the LF AI & Data Governing Board
- Appoint one representative as a voting member in any subcommittee or activities of the Governing Board
- Enjoy most prominent placement in displays of membership including website and marketing materials
- Increased access to Linux Foundation's invitation-only Open Source Leadership Summit
- Create an individualized press release upon membership announcement with the LF AI & Data PR team
- Receive ongoing, individual engagement and guidance from LF AI & Data Executive Director and staff

General Membership

Targeted for organizations that want to put their organization in full view in support of LF AI & Data and our mission. Organizations that join at the General level are 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 to appoint one representative to the LF AI & Data Governing Board per every ten General members, up to three maximum General representatives
- 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 individualized press release upon membership announcement with the LF AI & Data PR team
- Participate in all Marketing, Community, Thought Leadership opportunities
- Demonstrate your support for LF AI & Data by displaying your logo on the LF AI & Data website and in marketing materials

Associate Membership

Limited to academic and nonprofit institutions respectively and requires approval by the Governing Board

- Academia
- Not-for-profit
- Government agencies
- Standardization bodies
- etc.

Free membership


LF AI & Data Projects

LF AI & Data Hosted Projects

Please visit https://lfaidata.foundation/projects/ for the most updated list of hosted projects.

Acumos

Brief Description:

Acumos is an Open Source Platform, which supports design, integration and deployment of AI models. Furthermore, Acumos supports an AI marketplace that empowers data scientists to publish adaptive AI models, while shielding them from the need to custom develop fully integrated solutions.

Contributed by:

AT&T and Tech Mahindra in May 2018 as a Graduate Project

Key Links:

Github https://github.com/acumos

Web Site https://www.acumos.org/

Artwork
https://github.com/lfai/artwork/tree/master/projects/acumos

Mail Lists

- acumosai-announce
- acumosai-technical-discuss
- acumosai-tsc

Angel ML

Brief Description:

Angel is a high-performance distributed machine learning platform based on the philosophy of Parameter Server. It is tuned for performance with big data from Tencent and has a wide range of applicability and stability, demonstrating increasing advantage in handling higher dimension model.

Contributed by:

Tencent in August 2018 as an Incubation Project
Angel graduated in December 2019

Key Links:

Github https://github.com/Angel-ML/angel

Artwork
https://github.com/lfai/artwork/tree/master/projects/angel

Mail Lists

- angel-announce
- angel-technical-discuss
- angel-tsc

Elastic Deep Learning

Brief Description:

EDL optimizes the global utilization of the cluster running deep learning job and the waiting time of job submitters. It includes two parts: a Kubernetes controller for the elastic scheduling of distributed deep learning jobs, and a fault-tolerable deep learning framework.

Contributed by:

Baidu in August 2018 as an Incubation Project

Key Links:

Github https://github.com/PaddlePaddle/edl

Web site http://www.paddlepaddle.org/

Artwork
https://github.com/lfai/artwork/tree/master/projects/edl

Mail Lists

- edl-announce
- edl-technical-discuss
- edl-tsc

Horovod

Brief Description:

Horovod, a distributed training framework for TensorFlow, Keras and PyTorch, improves speed, scale and resource allocation in machine learning training activities. Uber uses Horovod for self-driving vehicles, fraud detection, and trip forecasting. It is also being used by Alibaba, Amazon and NVIDIA. Contributors to the project outside Uber include Amazon, IBM, Intel and NVIDIA.

Contributed by:

Uber in December 2018 an Incubation Project

Graduated:

September 2020

Key Links:

Github https://github.com/horovod/

Web site https://horovod.ai

Artwork
https://github.com/lfai/artwork/tree/master/projects/horovod

Mail Lists

- horovod-announce
- horovod-technical-discuss
- horovod-tsc

Pyro

Brief Description:

Pyro is a universal probabilistic programming language (PPL) written in Python and supported by PyTorch on the backend. Pyro enables flexible and expressive deep probabilistic modeling, unifying the best of modern deep learning and Bayesian modeling.

Contributed by:

Uber in January 2019 as an Incubation Project

Graduated:

February 2021

Key Links:

Github https://github.com/pyro-ppl

Web site https://pyro.ai

Artwork
https://github.com/lfai/artwork/tree/master/projects/pyro

Mail Lists

- pyro-announce
- pyro-technical-discuss
- pyro-tsc

ADLIK

Brief Description:

Adlik is a toolkit for accelerating deep learning inference. The goal of Adlik is to accelerate deep learning inference process both on cloud and embedded environments. Adlik consists of two sub projects: model compiler and serving platform. Model compiler supports several optimizing technologies like pruning, quantization and structural compression to optimize models developed in major frameworks like Tensorflow, Keras, and Caffe, so that they can run with lower latency and higher computing efficiency. Serving platform provides deep learning models with optimized runtime based on the deployment environment such as CPU, GPU, and FPGA. Based on a deep learning model, the users of Adlik can optimize it with model compiler and then deploy it to a certain platform with serving platform.

Contributed by:

ZTE in September 2019 as an Incubation Project

Key Links:

Github https://github.com/Adlik/Adlik

Web site https://adlik.ai

Artwork https://github.com/lfai/artwork/tree/master/projects/adlik

Mail Lists

- adlik-announce
- adlik-technical-discuss
- adlik-tsc

ONNX

Brief Description:

ONNX is an open format to represent deep learning models. With ONNX, AI developers can more easily move models between state-of-the-art tools and choose the combination that is best for them. ONNX is developed and supported by a community of partners. It is a community project created by Facebook and Microsoft. Many people are working on great tools, but developers are often locked in to one framework or ecosystem. ONNX is the first step in enabling more of these tools to work together by allowing them to share models. The goal of the project is to make it possible for developers to use the right combinations of tools for their project. We want everyone to be able to take AI from research to reality as quickly as possible without artificial friction from toolchains.

Contributed by:

ONNX Community in October 2019 as a Graduate Project

Key Links:

Github: https://github.com/onnx

Website: https://onnx.ai

Artwork: https://github.com/lfai/artwork/tree/master/projects/onnx

Mailing lists:

- onnx-announce
- onnx-technical-discuss
- onnx-tsc

sparklyr

Brief Description:

sparklyr is an open-source and modern interface to scale data science and machine learning workflows using Apache Spark™, R, and a rich extension ecosystem.
It enables using Apache Spark with ease using R by providing access to core functionality like installing, connecting and managing Spark and using Spark's MLlib, Spark Structured Streaming and Spark Pipelines from R.
Supports well-known R packages like dplyr, DBI and broom to reduce the cognitive overhead from having to re-learn libraries.

Contributed by:

RStudio in December 2019 as an Incubation Project

Key Links:

Github: https://github.com/sparklyr/sparklyr

Website: https://sparklyr.ai/

Artwork: https://github.com/lfai/artwork/tree/master/projects/sparklyr

Mailing lists:

- sparklyr-announce
- sparklyr-technical-discuss
- sparklyr-tsc

Marquez

Brief Description:

Marquez is an open source metadata service for the collection, aggregation, and visualization of a data ecosystem's metadata. It maintains the provenance of how datasets are consumed and produced, provides global visibility into job runtime and frequency of dataset access, centralization of dataset lifecycle management, and much more.

Contributed by:

WeWork in December 2019 as an Incubation Project

Key Links:

Github: https://github.com/MarquezProject/marquez

Website: https://marquezproject.ai

Artwork: https://github.com/lfai/artwork/tree/master/projects/marquez

Mailing lists:

- marquez-announce
- marquez-technical-discuss
- marquez-tsc

Milvus

Brief Description:

Milvus is an open source similarity search engine for massive-scale feature vectors. Built with heterogeneous computing architecture for the best cost efficiency. Searches over billion-scale vectors take only milliseconds with minimum computing resources. Milvus can be used in a wide variety of scenarios to boost AI development.

Contributed by:

ZILLIZ in January 2020 as an Incubation Project and Graduated in June 2021

Key Links:

Github: https://github.com/milvus-io/milvus

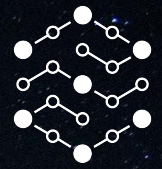
Website: https://milvus.io/

Artwork: https://github.com/lfai/artwork/tree/master/projects/milvus

Mailing lists:

- milvus-announce
- milvus-technical-discuss
- milvus-tsc

NNStreamer



NNStreamer

Brief Description:

NNStreamer is a set of Gstreamer plugins that allow Gstreamer developers to adopt neural network models easily and efficiently and neural network developers to manage neural network pipelines and their filters easily and efficiently.

Contributed by:

Samsung in March 2020 as an Incubation Project

Key Links:

Github: https://github.com/nnstreamer/nnstreamer

Artwork: https://github.com/lfai/artwork/tree/master/projects/nnstreamer

Mailing lists:

- nnstreamer-announce
- nnstreamer-technical-discuss
- nnstreamer-tsc

ForestFlow

Brief Description:

ForestFlow is a scalable policy-based cloud-native machine learning model server. ForestFlow strives to strike a balance between the flexibility it offers data scientists and the adoption of standards while reducing friction between Data Science, Engineering and Operations teams.

Contributed by:

ZILLIZ in March 2020 as an Incubation Project

Key Links:

Github: https://github.com/ForestFlow/ForestFlow

Website: https://forestflow.ai/

Artwork: https://github.com/lfai/artwork/tree/master/projects/forestflow

Mailing lists:

- forestflow-announce
- forestflow-technical-discuss
- forestflow-tsc

Ludwig

Brief Description:

Ludwig is a toolbox built on top of TensorFlow that allows to train and test deep learning models without the need to write code. All you need to provide is your data, a list of fields to use as inputs, and a list of fields to use as outputs, Ludwig will do the rest. Simple commands can be used to train models both locally and in a distributed way, and to use them to predict on new data.

Contributed by:

Uber in May 2020 as an Incubation Project

Key Links:

Github: http://github.com/uber/ludwig

Website: http://ludwig.ai

Artwork: https://github.com/lfai/artwork/tree/master/projects/ludwig

Mailing lists:

- ludwig-announce
- ludwig-technical-discuss
- ludwig-tsc

Adversarial Robustness Toolkit

Brief Description:
Adversarial Robustness Toolbox (ART) provides tools that enable developers and researchers to evaluate, defend, certify and verify Machine Learning models and applications against the adversarial threats.

Contributed by:
IBM in June 2020 as an Incubation Project and graduated in February 2022

Key Links:

Github: https://github.com/Trusted-AI/adversarial-robustness-toolbox

Website: https://artoolbox.lfprojects.linuxfoundation.org/

Artwork: https://github.com/lfai/artwork/tree/master/projects/adversarial-robustness-toolbox

Mailing lists:

- trusted-ai-360-announce
- trusted-ai-360-technical-discuss
- trusted-ai-360-tsc

AI Explainability 360

AI Explainability 360

Brief Description:

AI Explainability 360 is an open source toolkit that can help users better understand the ways that machine learning models predict labels using a wide variety of techniques throughout the AI application lifecycle.

Contributed by:

IBM in June 2020 as an Incubation Project

Key Links:

Github: https://github.com/Trusted-AI/AIX360

Website: https://aiexplainability.lfprojects.linuxfoundation.org/

Artwork: https://github.com/lfai/artwork/tree/master/projects/ai-explainability-360

Mailing lists:

- trusted-ai-360-announce
- trusted-ai-360-technical-discuss
- trusted-ai-360-tsc

AI Fairness 360

Brief Description:

AI Fairness 360 is an extensible open source toolkit that can help users understand and mitigate bias in machine learning models throughout the AI application lifecycle.

Contributed by:

IBM in June 2020 as an Incubation Project

Key Links:

Github: https://github.com/Trusted-AI/AIF360

Website: https://aifairness.lfprojects.linuxfoundation.org/

Artwork: https://github.com/lfai/artwork/tree/master/projects/ai-fairness-360

Mailing lists:

- trusted-ai-360-announce
- trusted-ai-360-technical-discuss
- trusted-ai-360-tsc

Amundsen

Brief Description:

Amundsen is a data discovery and metadata engine for improving the productivity of data analysts, data scientists and engineers when interacting with data. It does that today by indexing data resources (tables, dashboards, streams, etc.) and powering a page-rank style search based on usage patterns (e.g. highly queried tables show up earlier than less queried tables). Think of it as Google search for data.

Contributed by:

Lyft in July 2020 as an Incubation Project

Key Links:

Github: https://github.com/amundsen-io

Website: https://www.amundsen.io/

Artwork: https://github.com/lfai/artwork/tree/master/projects/amundsen

Mailing lists:

- amundsen-announce
- amundsen-technical-discuss
- amundsen-tsc

DELTA

Brief Description:

DELTA is a deep learning based end-to-end natural language and speech processing platform. DELTA aims to provide easy and fast experiences for using, deploying, and developing natural language processing and speech models for both academia and industry use cases. DELTA is mainly implemented using TensorFlow and Python 3.

Contributed by:

Didi in September 2020 as an Incubation Project

Key Links:

Github: https://github.com/didi/delta

Website: To be updated

Artwork: https://github.com/lfai/artwork/tree/master/projects/delta

Mailing lists:

- delta-announce
- delta-technical-discuss
- delta-tsc

SOAJS

Brief Description:

SOAJS is an open source microservices and API management platform, SOAJS eliminates the IT plumbing challenges, so you can deploy microservices significantly earlier and faster. IT initiatives such as digital transformation are simplified, accelerated, cost reduced, and risk mitigated. Our fully integrated, world-class API lifecycle management, multi-cloud orchestration, release management, and IT Ops automation capabilities eliminate your IT organization's modernization pain.

Contributed by:

Herron Tech in September 2020 as an Incubation Project

Key Links:

Github: https://github.com/soajs

Website: https://www.soajs.org/

Artwork: https://github.com/lfai/artwork/tree/master/projects/soajs

Mailing lists:

- soajs-announce
- soajs-technical-discuss
- soajs-tsc

Feast

Brief Description:

Feast is an open source feature store for machine learning. It was developed as a collaboration between Gojek and Google in 2018. Feast aims to:

- Provide scalable and performant access to feature data for ML models during training or serving.
- Provide a consistent view of features for both training and serving.
- Enable re-use of features through discovery, documentation, and metadata tracking.
- Ensures model performance by tracking, validating, and monitoring features in production.

Contributed by:

Gojek in September 2020 as an Incubation Project

Key Links:

Github: https://github.com/feast-dev/feast

Website: https://feast.dev/

Artwork: https://github.com/lfai/artwork/tree/master/projects/feast

Mailing lists:

- feast-announce
- feast-technical-discuss
- feast-tsc

Egeria

Brief Description:

Egeria is the world's first open source metadata standard. It provides open APIs, event formats, types and integration logic so organizations can share data management and governance across the entire enterprise without reformatting or restricting the data to a single format, platform, or vendor product.

Contributed by:

Initially contributed to ODPi by IBM and ING Bank in August 2018, and added as a Graduate Project in November 2020

Key Links:

Github: https://github.com/odpi/egeria

Website: https://egeria-project.org

Artwork: https://artwork.lfaidata.foundation/projects/egeria/

Mailing lists:

- egeria-announce
- egeria-technical-discuss
- egeria-tsc

OpenDS4All

Brief Description:

OpenDS4All is an open source project built to accelerate the creation of data science curricula at academic institutions.

Contributed by:

Initially contributed to ODPi by IBM and the University of Pennsylvania in August 2019, and added as an Incubation Project in November 2020

Key Links:

Github: https://github.com/odpi/opends4all

Artwork: https://artwork.lfaidata.foundation/projects/opends4all/

Mailing lists:

- opends4all-announce
- opends4all-technical-discuss
- opends4all-tsc

JanusGraph

Brief Description:

JanusGraph is a scalable graph database optimized for storing and querying graphs containing hundreds of billions of vertices and edges distributed across a multi-machine cluster.

Key Links:

Github: https://github.com/janusgraph

Website: https://janusgraph.org

Artwork: https://artwork.lfaidata.foundation/projects/janusgraph/

Mailing lists:

- janusgraph-announce
- janusgraph-users
- janusgraph-dev
- janusgraph-tsc

Contributed by:

Initially launched at the Linux Foundation by Expero, Google, GRAKN.AI, Hortonworks, IBM and others in January 2017, and added as an Incubation Project in December 2020

Datapractices

DATAPRACTICES.ORG

Brief Description:

DataPractices is a “Manifesto for Data Practices,” comprised of values and principles to illustrate the most effective, modern, and ethical approach to data teamwork.

Contributed by:

Initially contributed to the Linux Foundation by data.world in March 2019, and added as an Incubation Project in December 2020

Key Links:

Github: https://github.com/datadotworld/data-practices-site

Website: https://datapractices.org/

Artwork: https://artwork.lfaidata.foundation/projects/datapractices/

Mailing lists:

- datapractices-announce
- datapractices-technical-discuss
- datapractices-tsc

Datashim

Brief Description:

Datashim is enabling and accelerating data access for Kubernetes/Openshift workloads in a transparent and declarative way. Opensourced since September of 2019 and is growing to support use-cases related to data access in AI projects. It brings benefits across different entities:

- Data scientists/engineers: Focus on workload/experiments development and not on configuring/tuning data access
- Storage Providers: Increase adoption since the framework is extensible without hindering the User Experience
- Data-oriented Frameworks: Can build capabilities (caching, scheduling) on top of DLF using a declarative way to access/manage data sources

Contributed by:

IBM in January 2021 as an Incubation Project

Key Links:

Github: https://github.com/datashim-io/datashim

Website: https://datashim.io/

Artwork:

https://github.com/lfai/artwork/tree/master/projects/datashim

Mailing lists:

- datashim-announce
- datashim-technical-discuss
- datashim-tsc

Flyte

Brief Description:

Flyte is a production-grade, declarative, structured and highly scalable cloud-native workflow orchestration platform. It allows users to describe their ML/Data pipelines using Python, Java or (in the future other languages) and Flyte manages the data flow, parallelization, scaling and orchestration of these pipelines. Flyte builds on top of Docker containers and kubernetes.

Contributed by:

Lyft in February 2021 as an Incubation Project and graduated in December 2021

Key Links:

Github: https://github.com/flyteorg/flyte

Website: https://flyte.org/

Artwork: https://github.com/lfai/artwork/tree/master/projects/flyte

Mailing lists:

- flyte-announce
- flyte-technical-discuss
- feast-tsc

RosaeNLG

Brief Description:

RosaeNLG is a template-based Natural Language Generation (NLG) automates the production of relatively repetitive texts based on structured input data and textual templates, run by a NLG engine. Production usage is widespread in large corporations, especially in the financial industry.

Typical use cases are:

- describing a product based on its features for SEO purposes
- produce structured reports like risk reports or fund performance in the financial industry
- generate well formed chatbot answers

RosaeNLG is an open source NLG engine. It aims:

- to offer the same NLG features as product NLG solutions
- to be developer and IT friendly for template configuration and integration
- to provide NLG on both server-side and browser-side

Contributed by:

Ludan Stoecklé in March 2021 as a Sandbox Project

Key Links:

Github: https://github.com/RosaeNLG/rosaenlg

Website: https://rosaenlg.org

Artwork:

https://github.com/lfai/artwork/tree/master/projects/rosaenlg

Mailing lists:

- rosaenlg-announce
- rosaenlg-technical-discuss
- rosaenlg-tsc

Substra

Brief Description:

Substra is a framework offering distributed orchestration of machine learning tasks among partners while guaranteeing secure and trustless traceability of all operations. It enables *privacy-preserving federated learning* projects, where multiple parties collaborate on a Machine Learning objective while each one keeps their private datasets behind their own firewall.

Contributed by:

OWKIN in March 2021 as an Incubation Project

Key Links:

Github: https://github.com/Substra

Website: https://www.substra.ai/

Artwork:

https://github.com/lfai/artwork/tree/main/projects/substra

Mailing lists:

- substra-announce
- substra-technical-discuss
- substra-tsc

Kompute

Brief Description:

Kompute is a general purpose GPU compute framework for cross vendor graphics cards (AMD, Qualcomm, NVIDIA & friends). Blazing fast, mobile-enabled, asynchronous and optimized for advanced GPU data processing use cases.

Contributed by:

The Institute for Ethical AI & Machine Learning in May 2021 as a Sandbox Project and graduated to incubation-stage in February 2022.

Key Links:

Github: https://github.com/KomputeProject/kompute

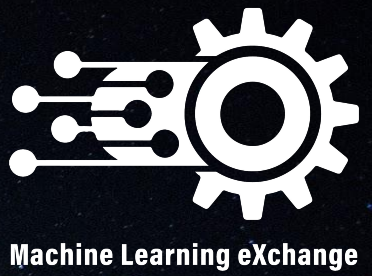
Website: https://kompute.cc/

Artwork: https://github.com/lfai/artwork/tree/main/projects/kompute

Mailing lists:

- kompute-announce
- kompute-technical-discuss
- kompute-tsc

Machine Learning eXchange



Brief Description:

Machine Learning eXchange (MLX) is a Data and AI Assets Catalog and Execution Engine. It allows upload, registration, execution, and deployment of: AI pipelines and pipeline components, models, datasets, and notebooks.

Contributed by:

IBM in May 2021 as a Sandbox Project

Key Links:

Github: https://github.com/machine-learning-exchange/mlx

Website: To be updated

Artwork:

https://github.com/machine-learning-exchange/artwork

Mailing lists:

- mlx-announce
- mlx-technical-discuss
- mlx-tsc

Open Lineage

Brief Description:

The OpenLineage project proposes an open standard and API for lineage collection that data processing engines can implement to publish at run time details of the data sources that it is reading, the types of processing it is performing and the destination of the results.

Contributed by:

Datakin in May 2021 as a Sandbox Project

Key Links:

Github: https://github.com/OpenLineage/OpenLineage

Website: http://openlineage.io/

Artwork:

https://github.com/lfai/artwork/tree/main/projects/openlineage

Mailing lists:

- openlineage-announce
- openlineage-technical-discuss
- openlineage-tsc

TonY

Brief Description:

TonY is a framework to natively run deep learning jobs on Apache Hadoop. It currently supports TensorFlow, PyTorch, MXNet and Horovod. TonY enables running either single node or distributed training as a Hadoop application. This native connector, together with other TonY features, aims to run machine learning jobs reliably and flexibly.

Contributed by:

LinkedIn in July 2021 as an Incubation Project

Key Links:

Github: https://github.com/linkedin/TonY

Website: https://tony-project.ai/

Artwork:

https://github.com/lfai/artwork/tree/main/projects/tony

Mailing lists:

- tony-announce
- tony-technical-discuss
- tony-tsc

Kedro

Kedro

Brief Description:

Kedro is an open-source Python framework for creating reproducible, maintainable and modular data science code. It borrows concepts from software engineering best-practice and applies them to machine-learning code; applied concepts include modularity, separation of concerns and versioning.

Contributed by:

McKinsey and QuantumBlack in August 2021 as an Incubation Project

Key Links:

Github: https://github.com/kedro-org

Website: https://kedro.readthedocs.io/en/stable/index.html

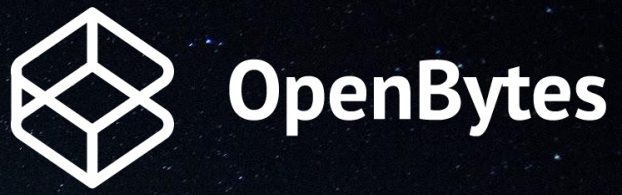
Artwork:

https://github.com/lfai/artwork/tree/main/projects/kedro

Mailing lists:

- kedro-announce
- kedro-technical-discuss
- kedro-tsc

OpenBytes



Brief Description:

OpenBytes aims to facilitate wider sharing of, and collaboration with, data in the AI community through the promotion of data standards and formats and enabling contributions of data. The value of this project lies in its stimulus on academic interest and AI innovation by promoting high-quality datasets and pushing the boundaries of science further.

Contributed by:

Graviti in October 2021 as a Sandbox Project

Key Links:

Github: https://github.com/Project-OpenBytes

Website: https://www.openbytes.io

Artwork:

https://github.com/lfai/artwork/tree/main/projects/openbytes

Mailing lists:

- openbytes-announce
- openbytes-technical-discuss
- openbytes-tsc

KServe

Brief Description: KServe provides a Kubernetes Custom Resource Definition for serving machine learning (ML) models on arbitrary frameworks. It aims to solve production model serving use cases by providing performant, high abstraction interfaces for common ML frameworks like Tensorflow, XGBoost, ScikitLearn, PyTorch, and ONNX. It encapsulates the complexity of autoscaling, networking, health checking, and server configuration to bring cutting edge serving features like GPU Autoscaling, Scale to Zero, and Canary Rollouts to your ML deployments. It enables a simple, pluggable, and complete story for Production ML Serving including prediction, pre-processing, post-processing and explainability.

Contributed by:
KServe Project Community in November 2021 as Incubation Project

Key Links:

Github: https://github.com/kserve

Website: https://kserve.github.io/website

Artwork: https://github.com/lfai/artwork/tree/main/projects/kserve

Mailing lists:

- kserve-announce
- kserve-technical-discuss
- kserve-tsc

Artigraph

Brief Description:

Contributed by:
Replica in January 2022 as Sandbox Project

Key Links:

Github: https://github.com/artigraph/artigraph

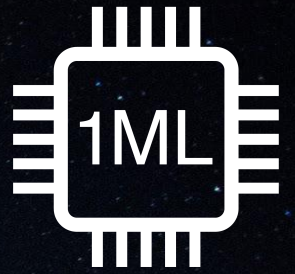
Website:

Artwork:

Mailing lists:

- artigraph-announce
- artigraph-technical-discuss
- artigraph-tsc

1chipML



Brief Description: 1chipML is an open source library for basic numerical crunching and machine learning for microcontrollers. As the Internet of Things and Edge Computing are becoming a ubiquitous reality, we need to a reliable and open framework to use on limited and low power demanding hardware.

Contributed by:
Ericsson in April 2022 a Sandbox Project

Key Links:

Github: https://github.com/1chipML/1chipML

Website: To be updated

Artwork:
https://github.com/lfai/artwork/tree/main/projects/1chipML

Mailing lists:

- 1chipml-announce
- 1chipml-technical-discuss
- 1chipml-tsc

BeyondML

Brief Description:

BeyondML is a framework for developing sparse neural networks that can perform multiple tasks across multiple data domains. This framework provides value to the community by:

- simplifying of the development and deployment of advanced machine learning capabilities for use on low-end devices and in dynamic environments characteristic of the resource-constrained edge
- reducing in the complexity and cost of deploying ML models or systems of models to cloud platforms
- reducing in the carbon footprint of deployed ML models

Key Links:

Github: https://github.com/Beyond-ML-Labs

Website: To be updated

Artwork: https://github.com/lfai/artwork/tree/main/projects/BeyondML

Mailing lists:

- beyondml-announce
- beyondml-technical-discuss
- beyondml-tsc

Contributed by:

AI Squared in June 2022 as a Sandbox Project

FlagAI

Brief Description:

FlagAI (Fast LArge-scale General AI models) is a fast, easy-to-use, and extensible toolkit for large-scale models. The goal is to support training, fine-tuning, and deployment of large-scale models on various downstream tasks with multi-modality. Currently, the project focuses on NLP models and tasks. In the near future, the project will support for other modalities.

Contributed by:

BAAI in June 2022 as a Sandbox Project

Key Links:

Github: https://github.com/BAAI-Open/FlagAI

Website: To be updated

Artwork: https://github.com/lfai/artwork/tree/main/projects/flagai

Mailing lists:

- flagai-announce
- flagai-technical-discuss
- flagail-tsc

Hosting Projects under the LF AI & Data Foundation

LF AI & Data provides the support for open development to occur among a diverse and thriving community, in addition to a number of enabling services that include membership and funding, ecosystem development, legal, PR and marketing, events, and compliance scans.

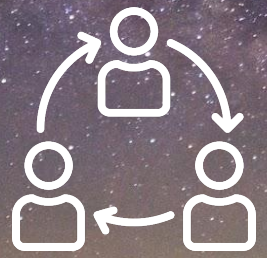
Host Your Project in LF AI & Data

- The process to host projects under LF AI is simple and includes:
  - Presenting the project to the LF AI Technical Advisory Council (TAC)
  - Providing a proposal on hosting the project under LF AI (via Github: https://github.com/lfai/proposing-projects/tree/master/proposals)
  - For “Incubation” project: A vote by the Technical Advisory Council
  - For “Graduation” project: A vote by the Technical Advisory Council and Governing Board
  - Signing the trademark and project account transfer agreement
  - Working with LF AI staff on integrating the project with the LF AI Foundation
- Detailed description of the process: https://lfaidata.foundation/proposal-and-hosting-process/
- Detailed requirements for hosting projects under the LF AI: https://github.com/lfai/proposing-projects

Creating sustainable open source projects requires hard work

Our Value Proposition to Projects

Getting Involved



Ways to Get Involved with LF AI & Data Projects

- ☑ Participate in the development: Review and submit patches, report bugs, request features, test, etc.
- ☑ Attend developer events for LF AI & Data projects
- ☑ Contribute to project documentation
- ☑ Provide your testing and deployment feedback via appropriate project channel
- ☑ Join the projects' mailing lists and participate in the discussions
- ☑ Start a local User Group Meetup

Ways to Get Involved with LF AI & Data Marketing & Outreach

- Promotion of project updates, releases, and news via LF AI & Data social media accounts
- Attend Outreach Committee meetings and participate in ongoing activities
- Publish use cases, case studies, white papers, and deployment insights
- Marketing and PR support for demos at meetups and events
- Host vendor neutral content via LF AI & Data blog site
- Get support for artwork, web site, content creation, etc., for LF AI & Data projects
- Contribute to the LF AI & Data landscape, promote in talks
- Coordination at events, speaking proposals, booth attendance, demos, etc.
- Volunteer to host an LF AI & Data Day and Project developer events
- Identify speaking opportunities and help secure speakers from the LF AI & Data community
- Help secure user stories about LF AI & Data based deployments
- Support LF AI & Data marketing and PR staff

Ways to Get Involved with the TAC

- ☑ Support TAC leadership in inviting speakers
- ☑ Attend TAC Bi-weekly calls, participate in the discussion, volunteer
- ☑ Share success stories, opportunities and challenges with the broader technical community
- ☑ Identify opportunities for collaboration on common interests and initiatives, seek input from peers
- ☑ Support technical leadership for harmonization efforts with other open source projects
- ☑ Support TAC in hosting and sponsoring intra-project and inter-project developer events
- ☑ Support TAC in evaluating new projects for inclusion in LF AI & Data; recommend new projects
- ☑ Support TAC Chair working with the GB to highlight opportunities and needed resource

Ways to Get Involved with the Trusted AI Committee

- Join the committee as a representative of your company
- Join the Trusted AI bi-weekly conference calls and contribute to ongoing efforts
- Promote the work of the Trusted AI committee, invite your collaborators to participate
- Contribute to the LF AI & Data landscape, promote in talks
- Coordination at events, speaking proposals, booth attendance, demos, etc.
- Subscribe to the mailing list and participate in discussions
- Host your Ethics related projects in the LF AI Data Foundation
- Help secure user stories about LF AI & Data based deployments
- Invite prominent researchers and developers in the space to speak to the LF AI & Data community

Members - Display Your LF AI & Data Membership Badge

- As an LF AI & Data member, we encourage you to download your membership badge to display on your website and collateral, as well as in trade show booths

- These badges are available (svg and png) from the LF AI & Data artwork GitHub repo: https://github.com/lfai/artwork/tree/master/lfai-membership-badge

Project - Display Your LF AI & Data Association

- As an LF AI & Data hosted project, we encourage you to download your association badge to display on your website and collateral, as well as in trade show booths

- These badges are available (svg and png) from the LF AI & Data artwork GitHub repo: https://github.com/lfai/artwork/tree/master/lfai-project-badge

LF AI & DATA

Landscape

https://landscape.lfai.foundation/

Linux Foundation AI & Data Landscape

2022-04-12T10:04:39Z de884f2

See the interactive landscape at l.lfaidata.foundation

Greyed logos are not open source

Machine Learning

Framework: Accord.NET, Microsoft CNTK, Microsoft LightGBM, Shogun, MLNET, RAY, ZenML

Platform: Angel (Graduated), ForestFlow (Incubating), Alibaba Alink, Apache SystemML, Ax, cortex, H2O, Kubeflow, Metaflow, mlflow, Orchest, SELDON, Virtual Kubelet

Library: 1ML (Sandbox), Apache Arrow, Caffe2, Flashlight, MediaPipe, mlpack, OpenDR, opencv, PyCaret, scikit-learn, Shogun, Sonnet, Transmogrifai, XGBoost, xLearn

Deep Learning

Framework: SINGA, Chainer, CNTK, dynet, Alibaba euler, [M]^3 Mindspore, nnabla, nnvm, PaddlePaddle, PyTorch, TensorFlow

Platform: TonY (Incubating), DeepDetect, Determined AI, jina, Keras, Polyaxon

Library: BigDL, Catalyst, DL4J, fast.ai, Keras, Ignite, PyTorch Lightning

Tool: Bolt, Intel Distiller, PipelineAI, TensorRT, tvm

Reinforcement Learning

Coach, Dopamine, Horizon, OpenAI, Google Dopamine, Google SEED RL

Programming

Graduated: Pyro

Incubating: Kompute

DASK, Intel daal4py, julia, MARS, Numba, NumPy, NYOKA, pomegranate, PyMC3, python, R, SciPy, SKIP, Stan

Data

Education: DataPractices.org (Incubating), OpenDS4All (Incubating)

Lineage: OpenBytes (Sandbox), OpenLineage (Sandbox)

Relational DB: CouchDB, MySQL, PostgreSQL, TiKV

Store & Format: Milvus (Graduated), JanusGraph (Incubating), Alluxio, Apache Iceberg, Apache ORC, Apache Ozone, Arrow, Avro, ceph, Delta Lake, druid, Hudi, InfluxDB, pandas, Parquet, pilosa, VEARCH, vespa, Vineyard

Versioning: Dolt, DVC, Quilt

Operations: Amundsen (Incubating), datashim (Incubating), MARQUEZ (Incubating), Alluxio, Beta Hub, ckan, WHYLABS whylogs

Feature Engineering: FEAST (Incubating), tsfresh

Stream Processing: NNStreamer (Incubating), Apache Beam, brooklin, Flink, fluentd, kafka, logstash, PREFECT, PULSAR, samza, Uber uReplicator

SQL Engine: Apache Drill, HAWQ, presto, SQLFlow, trino

Visualization: bokeh, IBM Data Science Experience, D3, plotly Dash, Uber deck.gl, Ecco, Google Facets, Grafana, Metabase, Immerse, RCloud, re:dash, NYU, Superset, Google TensorBoard

Pipeline Management: Artigraph (Sandbox), Intel Analytics Zoo, Apache Airflow, DAGSTER, Kedro, PiFlow, TEKTON, Argo

Labeling & Annotation: CVAT, Doccano, Label Studio, Labelbox, LabelImg, HITACHI Visual Intelligence, Microsoft VoTT

Governance: EGERIA (Graduated), Apache Atlas

Model

Inference: ADLIK (Incubating), Apache MXNet, KFServing, MNN, NVIDIA Triton Inference Server, NVIDIA TensorRT, uTensor

Federated Learning: SUBSTRA (Incubating), FATE, Intel OpenFL, PySyft, TensorFlow Federated

Training: Graduated, LUDWIG (Incubating), Microsoft NNI, Petastorm, torchnet

Parameter: Hivemall, Horovod, talos

Format & Interface: ONNX (Graduated), Uber Neuropod

Marketplace: Acumos (Graduated), Machine Learning eXchange (Sandbox), IBM Model Asset eXchange

Workflow: Flyte (Graduated), Kedro (Incubating), Apache Airflow, nifi, argo, Azkaban, BENTOML, Cadence, Couler, Hydrosphere, DataBolt, kestra, Spotify luigi, mleap, PREFECT, TRAINS

Benchmarking: DAWNBench, MLPerf

Tool: Qualcomm AIMET, Facebook Glow, FLAME, Microsoft Infer.NET, aws MMS, MDB, amazon Neo-AI, NETRON, ONNX, PipelineAI, studio.ml, Google TensorFlow Model Analysis, aws TorchServe, turi

Trusted & Responsible AI

Explainability: AI Explainability 360 (Incubating), ALIBI, ELI5, Google Lucid, Bolt, SHAP, SKATER

Adversarial: Adversarial Robustness Toolbox (Graduated), AdvBox, cleverhans, foolbox

Bias & Fairness: AI Fairness 360 (Incubating), Aequitas, Audit AI, Fairlearn

Distributed Computing

Computing & Management: EDL (Incubating), SOAJS (Incubating), Apache Bahir, MESOS, Spark, STORM, GNES, NETFLIX genie, Apache Singa, kubernetes, Intel Nauta, OPENSEF, Singularity

Interface: sparklyr (Incubating), TORCH, LEVY

The LF AI & Data landscape explores open source projects in Artificial Intelligence and Data and their respective sub-domains.

l.lfaidata.foundation

LF AI & DATA Landscape

LF AI & DATA

Security & Privacy

Google Private Join and Compute, IBM HElib, Microsoft SEAL, Google Tink, TF Encrypted

Natural Language Processing

DELTA (Incubating), RosaeNLG (Sandbox), Google ALBERT, AllenNLP, Google Bert, Facebook LASER, Mozilla DeepSpeech, fastText, flair, LUCIN, Kashgari, Facebook Pytext, RASA, spaCy, Hugging Face Transformers, Facebook XLM, Yoav Goldberg's bist-parser

Notebook Environment

BeakerX, colab, Elyra, IPython, jupyter, IBM Watson Studio Desktop, Polyaxon, Streamlit

Getting Involved Guide

For Current and Prospect Members

Web site:	https://lfaidata.foundation
Email:	info@lfaidata.foundation
Twitter:	@LFAIDataFdn
Landscape:	http://l.lfai.foundation
GitHub:	https://github.com/lfai
Mail Lists:	https://lists.lfaidata.foundation/g/main
Contact Us:	https://members.lfaidata.foundation/