Meeting of the Technical Advisory Council (TAC)

November 19, 2020

LF AI & DATA

Anti-Trust Policy

- Linux Foundation meetings involve participation by industry competitors, and it is the intention of the Linux Foundation to conduct all of its activities in accordance with applicable antitrust and competition laws. It is therefore extremely important that attendees adhere to meeting agendas, and be aware of, and not participate in, any activities that are prohibited under applicable US state, federal or foreign antitrust and competition laws.

- Examples of types of actions that are prohibited at Linux Foundation meetings and in connection with Linux Foundation activities are described in the Linux Foundation Antitrust Policy available at http://www.linuxfoundation.org/antitrust-policy. If you have questions about these matters, please contact your company counsel, or if you are a member of the Linux Foundation, feel free to contact Andrew Updegrove of the firm of Gesmer Undergone LLP, which provides legal counsel to the Linux Foundation.

Recording of Calls

- **Reminder:**
- TAC calls are recorded and available for viewing on the TAC Wiki

Reminder: LF AI & Data Useful Links

- Web site: lfai.foundation
- Wiki: wiki.lfai.foundation
- GitHub: github.com/lfai
- Landscape: landscape.lfai.foundation or l.lfai.foundation
- Mail Lists: https://lists.lfai.foundation
-
- LF AI Logos: https://github.com/lfai/artwork/tree/master/lfai
- LF AI Presentation Template: https://drive.google.com/file/d/1eiDNJvXCqSZHT4Zk_-czASIz2GTBRZk2/view?usp=sharing
-
- Events Page on LF AI Website: https://lfai.foundation/events/
- Events Calendar on LF AI Wiki (subscribe available): https://wiki.lfai.foundation/pages/viewpage.action?pageId=12091544
- Event Wiki Pages: https://wiki.lfai.foundation/display/DL/LF+AI+Foundation+Events

Agenda

- Roll Call (2 mins)
- Approval of Minutes (2 mins)
- Invited Presentations (15 minutes each)
  - OpenDS4All
  - DataPractices.org
  - edX project - AI Ethics
- LF AI General Updates (3 minutes)
- Open Discussion (3 minutes)

TAC Voting Members

Board Member	Contact Person	Email
AT&T	Anwar Atfab	anwar@research.att.com
Baidu	Ti Zhou	zhouti@baidu.com
Ericsson	Rani Yadav-Ranjan	rani.yadav-ranjan@ericsson.com
Huawei	Huang Zhipeng	huangzhipeng@huawei.com
IBM	Susan Malaika	malaika@us.ibm.com
Nokia	Jonne Soininen	jonne.soininen@nokia.com
SAS	Craig Rubendall	craig.rubendall@sas.com
Tech Mahindra	Nikunj Nirmal	nn006444@techmahindra.com
Tencent	Bruce Tao	brucetao@tencent.com
Zilliz	Jun Gu	jun.gu@zilliz.com
ZTE	Wei Meng	meng.wei2@zte.com.cn
Graduate Project	**Contact Person**	**Email**
Acumos	Nat Subramanian	natarajan.subramanian@techmahindra.com
Angel	Bruce Tao	brucetao@tencent.com
Egeria	Mandy Chessell	mandy_chessell@uk.ibm.com
Horovod	Travis Addair	taddair@uber.com
ONNX	Jim Spohrer (Chair of TAC)	spohrer@us.ibm.com

Approval of November 5, 2020 Minutes

Draft minutes from the Novermber 5th TAC call were previously distributed to the TAC members via the mailing list

Proposed Resolution:

- That the minutes of the November 5th meeting of the Technical Advisory Council of the LF AI & Data Foundation are hereby approved.

Invited Presentation

OpenDS4All -

- Andre de Waal (IBM)

LF AI & DATA

OpenDS4All

An Open Source Approach to Data Science Education

Andre de Waal

October 2020

An Open Source Approach to Data Science Education

- A Data Science Curriculum Kit built on Open Source (Python)
- Built by professors at the University of Pennsylvania
- Launched as an Open Source Project under the Linux Foundation Governance to ensure updates and growth

GOAL

Enable Organizations to build their own Data Science Educational Programs

https://github.com/odpi/OpenDS4All

Lessons (topics) covered in the initial modules

Course Description

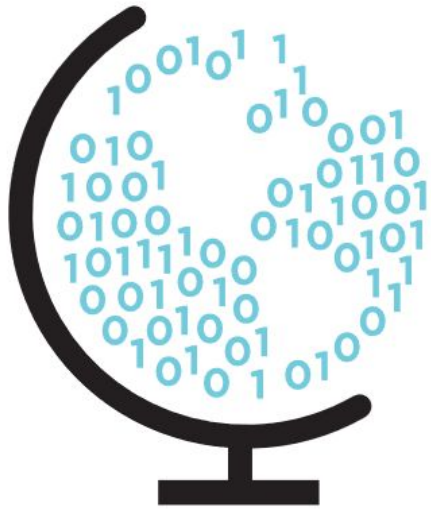
- Introduction: What is Data Science?
- How Your Computer Works
- Getting Started in Data Science: Data Acquisition and Wrangling
- Information Visualization, aka Visual Analytics
- Data representation and Modeling
- Scaling to Big Data and the Cloud
- Cluster-Based Processing of Graph Data
- Efficient Data Processing
- Network Centrality and PageRank
- PageRank and Graph Analysis Using Linear Algebra
- Unsupervised Machine Learning
- Supervised Machine Learning: Overview, Decision Trees, Random Forests
- Supervised Machine Learning: Linear and Logistic Regression
- Supervised Machine Learning: Artificial Neural Networks
- Training Robust Models
- Data Science Ethics

Governance

Technical Steering Committee

- David Mongeau (Executive Director, Berkeley Institute for Data Science, University of California, Berkeley)
- Susan Davidson (Professor, Computer and Information Science, UPenn)
- Zack Ives (Professor, Computer and Information Science, UPenn)
- Jennifer Priestley (Professor, Analytics and Data Science Institute, Kennesaw State University)
- Aric LaBarr (Professor, Institute for Advanced Analytics, NC State)
- Goutam Chakraborty (Professor, Spears School of Business, OSU)
- Ana Echeverri (IBM)
- Andre de Waal (IBM) – chairperson
- James Harroun (SAS Institute)

https://github.com/odpi/OpenDS4All



ODPi

OpenDS4All

GitH

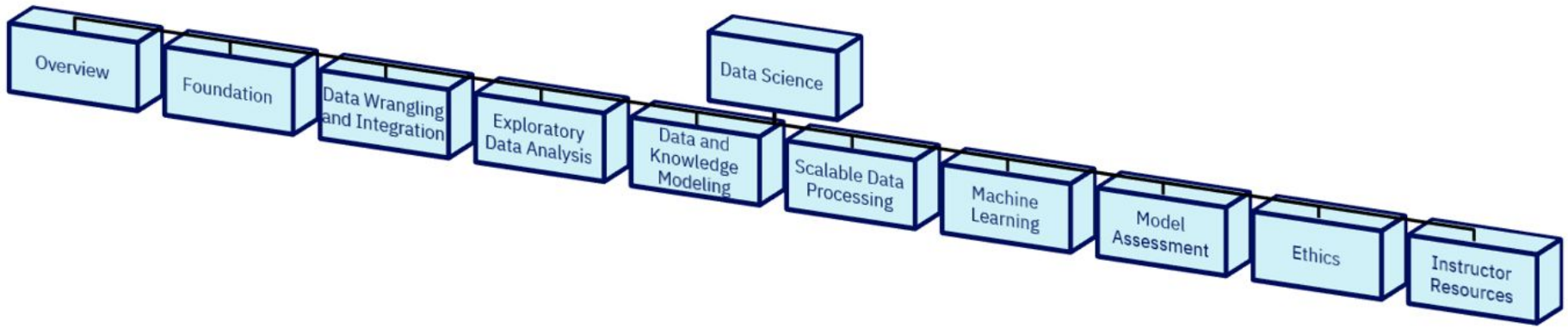
Description

OpenDS4All is a project created to accelerate the creation of data science curricula at academic institutions. While a great deal of online material is available for data science, **including** online courses, we recognize that the **best** way for many students to learn (and for many institutions to deliver) content is through a combination of lectures, recitation or flipped classroom activities, and hands-on assignments.

OpenDS4All attempts to fill this important niche. Our goal is to provide recommendations, slide sets, sample Jupyter

How to use

The following topology shows how content is **currently** organized around categories. This is a living/dynamic taxonomy that is updated as new content is added to the project.



Each category contains modules and each module consists of one or more of the following components:

- instructor notes (`Instructor_Notes.md`) and guide to files
- a set of PowerPoint slides (with presenter notes) ending in `.pptx`
- companion Jupyter notebooks, for students to see the lecture materials "in context" and to be able to experiment
- sample quiz materials (where applicable)
- sample homework assignments (where applicable)
- additional documentation (where applicable)

Branch: master ▾ OpenDS4All / opends4all-resources / opends4all-overview / Create new file | Upload files | Find file | History

SusanBDavidson and zackives test ... Latest commit 82bec3a 25 days ago

..		
INTRODUCTION-Data-Science-basic.pptx	Susan's changes based on Ana's feedback	last month
Instructor_Notes.md	Updates for spellcheck and MarkDown lint	last month
Quiz.md	Fixed line breaks.	28 days ago
README.md	test	25 days ago

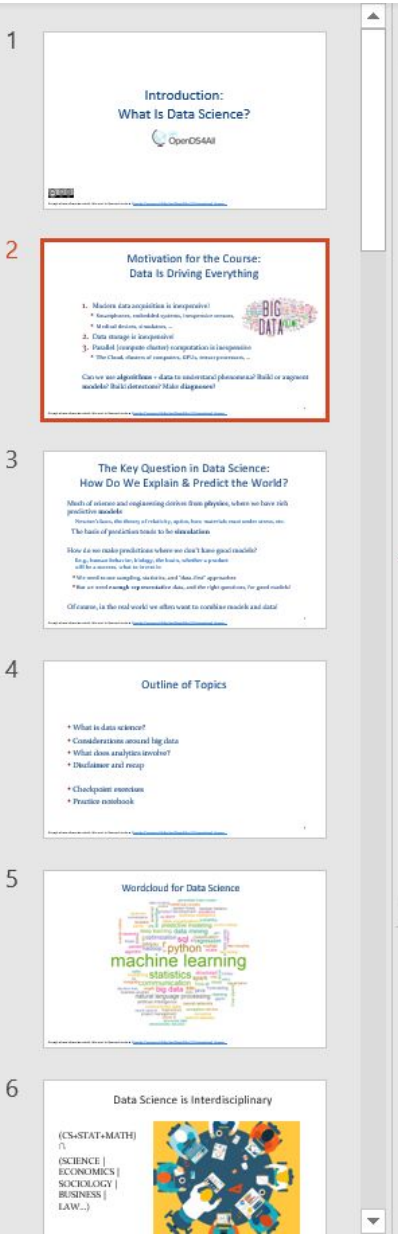
README.md

Overview: What is Data Science?

This module provides a broad overview of modern data science, data analytics / data engineering, and big data. It sets the context for the OpenDS4All curriculum.

Directory Contents

https://github.com/odpi/OpenDS4All/tree/master/opends4all-resources



Motivation for the Course: Data Is Driving Everything

1. Modern data acquisition is inexpensive!
   - Smartphones, embedded systems, inexpensive sensors,
   - Medical devices, simulators, …
2. Data storage is inexpensive!
3. Parallel (compute cluster) computation is inexpensive
   - The Cloud, clusters of computers, GPUs, tensor processors, …

Can we use **algorithms + data** to understand phenomena? Build or augment **models**? Build **detectors**? Make **diagnoses**?

2

Except where otherwise noted, this work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

There is data everywhere. It is constantly being collected (e.g. through smartphones, IoT sensors, Alexa) and contributed (e.g. through Facebook and Twitter). Storage is cheap, and compute power readily available and rapidly increasing. Computer clusters are available to enable parallelism, and special architectures developed for ML operations. All this is available not just for those who can afford to own them but through the commodity of the cloud..

This increased ability to collect, store and process data has led to a revolution in how we do things in almost every field, from biological sciences to social sciences, humanities, political science, policy, marketing and so on. But just because we have the data doesn't imply we understand it. It must be processed to derive information, make models, enable predictions, and make diagnoses. We want to use the data to develop (or augment) models using machine learning and other algorithms.

That is the essence of this course.

Branch: **master** ▾

Create new file | Upload files | Find file | History

OpenDS4All / opends4all-resources / **opends4all-exploratory-data-analysis** /

zackives More detail on presentation, added preview for ML, removed comments ✓ Latest commit `7803f46` 16 days ago

..

README.md	Exploratory data analysis - visual analytics	21 days ago
VISUAL-ANALYTICS-visualization.ip...	added crawled dataset for module 1 wikipedia etc	20 days ago
VISUAL-ANALYTICS-visualization.p...	More detail on presentation, added preview for ML, removed comments	16 days ago

README.md

Exploratory Data Analysis

The topic of exploratory data analysis is extremely important in discovering and formulating hypotheses. A key aspect of that process is *visual analytics* or *information visualization*, which allows one to see certain aspects of the data holistically. This module is focused on such topics.

Directory Contents

Lecture 1: Wrangling

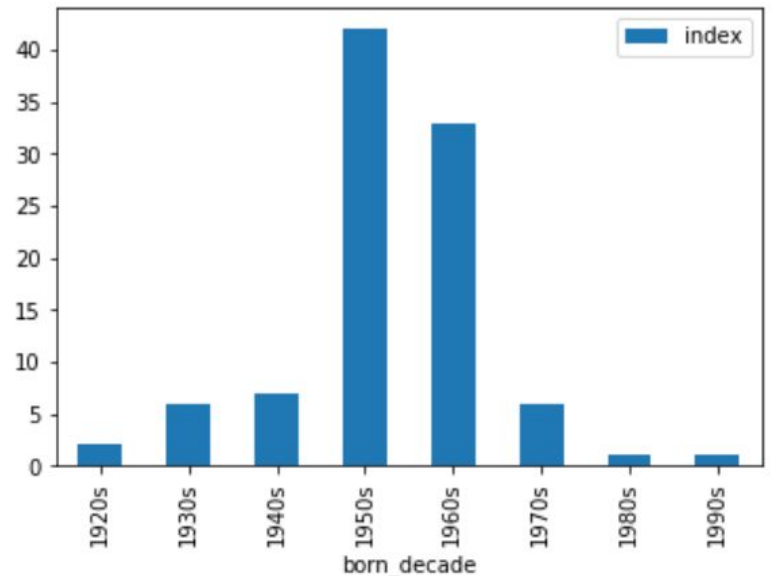
In this lecture notebook, we will test the hypothesis that the CEOs of major companies are typically in their 40s or older, i.e. that they were born after 1980. To do this, we will identify top companies, find their CEOs, extract their ages, and look at the distribution of ages.

In [1]:
```
# Let's start by installing some libraries that are useful for processing web data.

# For crawling pages
!pip3 install scrapy

# # Optional, for parallel execution
!pip3 install swifter

# # For string similarity
!pip3 install py_stringsimjoin

# # lxml to parse xml tree
!pip3 install lxml
```

```
Collecting scrapy
  Downloading https://files.pythonhosted.org/packages/3b/e4/69b87d7827abf03dea2ea984230d50f347b00a7a3897bc
93f6ec3dafa494/Scrapy-1.8.0-py2.py3-none-any.whl (238kB)
     |████████████████████████████████| 245kB 2.7MB/s
Collecting w3lib>=1.17.0
  Downloading https://files.pythonhosted.org/packages/6a/45/1ba17c50a0bb16bd950c9c2b92ec60d40c8ebda9f3371a
e4230c437120b6/w3lib-1.21.0-py2.py3-none-any.whl
Collecting protego>=0.1.15
```

In [35]:
```
# Plot the result.

exec_df[['born_decade']].reset_index().groupby('born_decade').count().plot(kind='bar')
```

Out[35]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x7fa0045d84e0>
```



So are most CEO's born before 1980s?

7. An Exercise

Take this one step further, and link the companies table with another dataset to determine the market for each company.

Is there a correlation between the kind of company and the age of the CEO?

Adopters of OpenDS4All content

5 lines (5 sloc) | 510 Bytes

Raw Blame

Search this file...

	Academic Institution	Country	Program	Course	Faculty Member	Number of Students
2	Kennesaw State University	USA	Interdisciplinary	DS7900	Jennifer Priestley	25
3	Oklahoma State University	USA	MS in Business Analytics and Data Science	BAN 5743 5753 5763	Goutam Chakraborty	150
4	University of North Carolina at Charlotte	USA	Computer Science	ITCS6100 Introduction to Data Science	Wlodek Zadrozny	25
5	Imperial College London	UK	Design Engineering	DESE50001 Data Science	Michel-Alexandre Cardin	50

andre.dewaal@ibm.com

https://github.com/odpi/OpenDS4All

Collaborating with the project – Help us Find…..

- Contributors
  - Faculty to contribute curricula
- Committers
  - Faculty to ensure curricula integrity

- Adopters
  - Organizations building Data Science academic programs

andre.dewaal@ibm.com

https://github.com/odpi/OpenDS4All

Invited Presentation

DataPractices.Org

- Patrick McGarry (Data.World)

LF AI & DATA

data.world

DataPractices.org

The Agile movement for data

Patrick McGarry

data.world

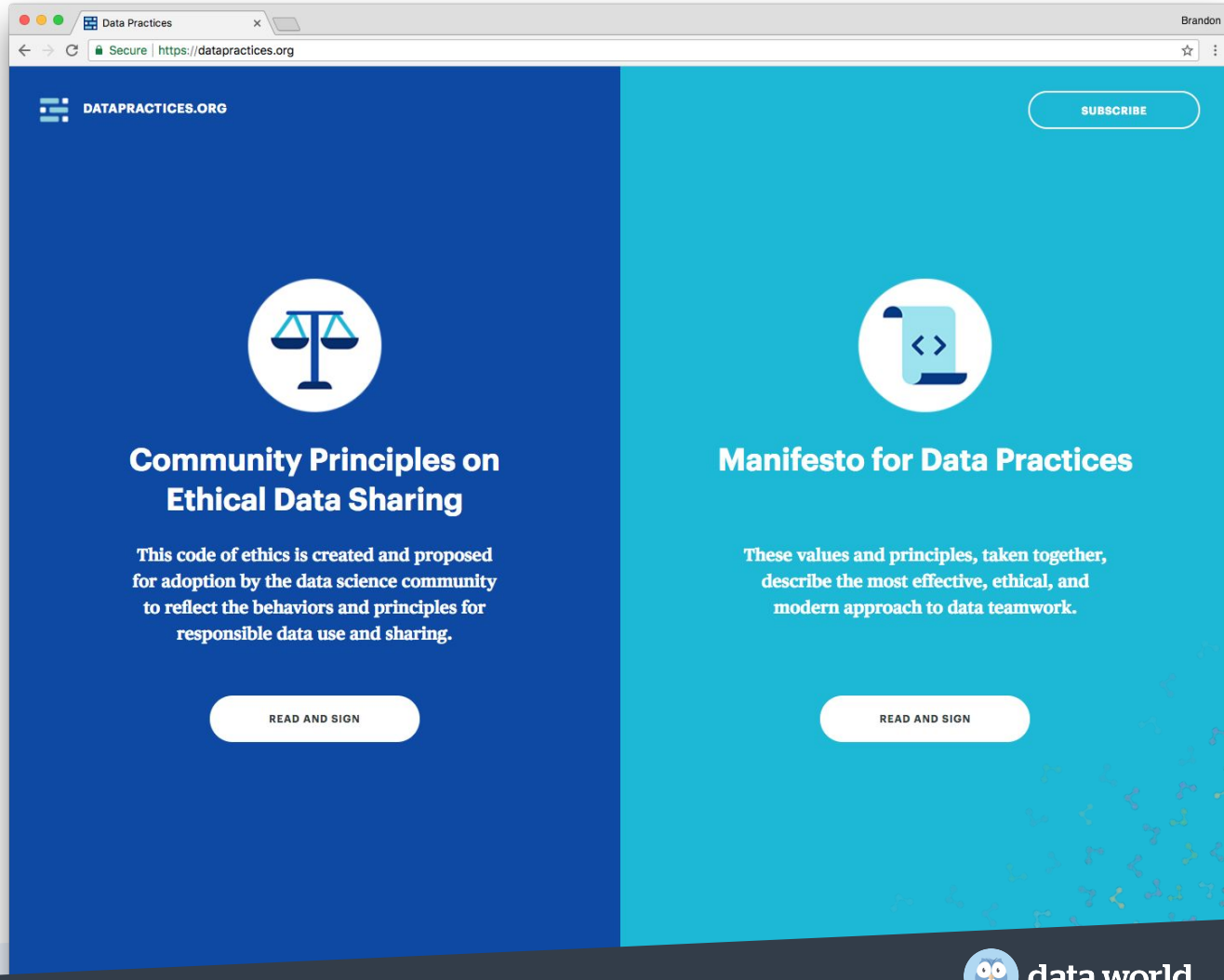
https://datapractices.org

Manifesto for Data Practices

1. First release successful
2. Community-driven
3. Still gathering signatures
4. **Origination:** ODSLS
5. **Focus:** Data work

Global Data Ethics Pledge

1. Was: Community Principles on Ethical Data Sharing
2. Work in progress
3. Community-driven (via D4D)
4. Community draft released
5. **Origination**: Bloomberg, D4D, Brighthive
6. **Focus**: Data Ethics



Recent Evolution

- **Two collections of content**
  - Project Lifecycle
  - Culture / Practice

- **Community Reviewed / Maintained**

- Community
  - Carpentries
  - Zuar
  - Brian Granger / Fernando Perez
  - Data Society
  - UCLA
  - Oracle
  - Comcast
  - Microsoft

- Commercial
  - Praxis Analytics
  - The Data Incubator
  - Data Literacy

Supported by leaders of the data community

39 authors, including:

- **Eric Colson**, Chief Algorithms Officer, StitchFix
- **Amy Gershkoff**, former Chief Data Scientist, Ancestry.com
- **Fernando Perez**, creator of iPython, Assistant Professor, Statistics, UC Berkeley
- **Andrew Therriault**, Chief Data Officer, City of Boston
- **Therese Couture**, Human Trafficking Data Analyst, Polaris
- **Wes McKinney**, BDFL, Pandas

2,200+ signatories, including:

- **DJ Patil**, former Chief Data Scientist of the United States
- **Monica Rogati**, former VP of Data Science, Jawbone
- **Kirk Borne**, Principal Data Scientist, Booz Allen Hamilton
- **Tricia Wang**, Fellow, Harvard Berkman Center
- **Jonathan Albright**, Research Director, Tow Center for Digital Journalism
- **Gregory Piatetsky**, founder, KDnuggets.com

Courseware

Project Lifecycle

- Kicking off a data project
- Sourcing Data
- Profile and Prepare
- Data Exploration
- Analyze and Report
- Data Project Post-Mortem

Culture / Practice

- How to Build a Data-Driven Culture
- A Survey of Data Visualization
- Moving from Computer Science to Data Science
- Data Science for People Who Don’t Code
- Data Visualization Best Practices
- Data Ethics 101

Live Look

What's Next?

- Courseware
  - Engage existing volunteers
  - Expand content collaboration (OpenDS4All)
  - Deploy and use content

- Principles
  - Continue to build on momentum of signatories
  - Use as basis of CoC and Ethics development

Questions?

Appendix

The world of modern data teamwork isn't one that can be created by software and business process alone. Individuals will need to alter their behavior, which is the hardest part about change.

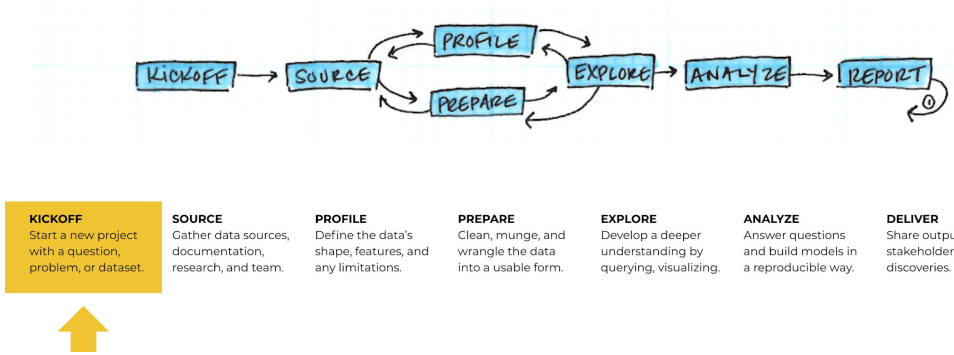
This workshop will show you not only how to help your team evolve, but the reasons that will make it clear as to why they should.

- **Topics covered:**
  - **Play bigger / Winning With Data**
  - **Pillars of a data-driven company**
    - Data infrastructure
    - Data governance
    - Data literacy
  - **Data driven leadership**
  - **Data Literacy (ex: Facebook data bootcamp)**
  - **Data access / dictionary / single source of truth**
  - **Decision making (answer real questions, not frivolous dashboards)**
  - **Data isn't everything, don't get paralyzed**
  - **Be patient, change takes time**
  - **Hold employees accountable, but create social conduits for them to get help**
  - **Newsletter**
  - **Treat data like an asset! (operationalized!)**
  - **Create quantifiable goals that are aligned with business objective**
  - **Data governance isn't a dirty word**
  - **Break down silos (80% of data is never analyzed)**
  - **Ask questions:**
    - "Do you have a data portal/platform?"
    - "What %age of your employees use it each week? (AirBnB/Warby Parker)"

Effective data collaboration takes more than giving everyone access. Inclusion, flexibility, setting, and prioritization are also keys to fruitful data work.

This workshop will help you to understand Data Practices values and principles, which describes the most effective, ethical, and modern approach to data teamwork, and how to best kick off a modern data project. Signed by some of the brightest minds in data, these values can create positive change wherever data collaboration occurs.

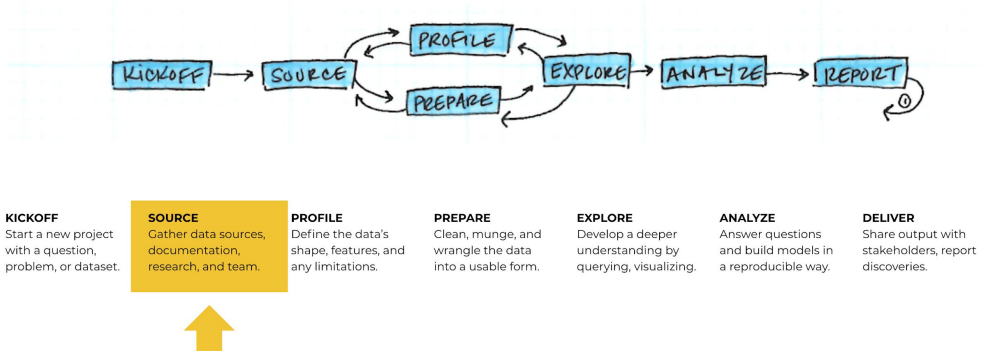
- Topics covered:
  - History of data practices movement
  - Break down objectives
  - Consider stakeholders
  - Refining questions
  - Planning for output / solution



Once you have planned out a data project it's time to gather your resources. There are many things to consider, including where data, documentation, and additional team members can be sourced.

This workshop will help you to understand how to evaluate data project resources including how to aggregate your existing data, the benefits of open data (both using and generating), the pros and cons of purchased data, and how to ensure that your data project can avoid costly pitfalls caused by problems around data provenance, data prep, or anything else.

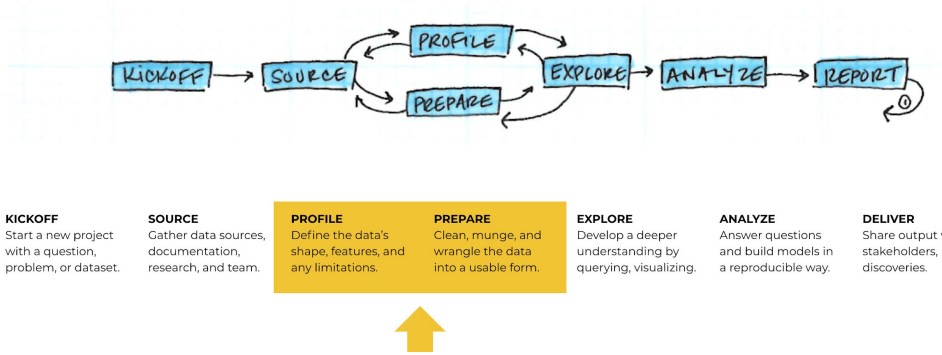
- **Topics covered:**
  - **Defining the question you're asking**
    - Identify the right metrics
    - Qualitative vs quantitative
  - **Importance of the right data**
    - "Garbage in, garbage out"
  - **Open vs closed**
  - **How to find:**
    - Data sources
    - Documentation
    - Research
    - Team members
  - **Purchasing data + / -**
  - **Things to consider**
    - Provenance
    - Source
    - Creation methodology
    - Distribution / size of data
  - **Joining / Normalizing data**



Once you have all of your data, team, documentation, and other resources in place, it's time to dig in and go to work.

This workshop will help users to understand how to effectively evaluate and define their data. Understanding the shape of your data, what features and limitations may be inherent in the data, what data might be missing or improperly formatted, and many other things can help a data team to get the most out of their data.

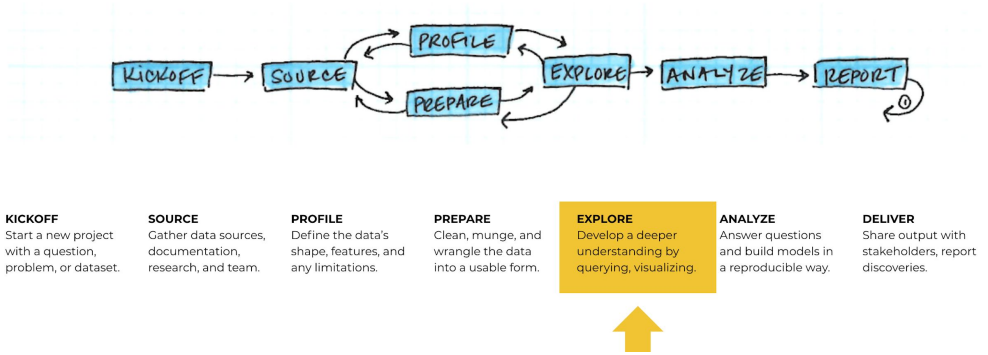
After profiling your data you'll need to clean and prepare your data for actual use. This workshop will help you understand some of the common methods of cleaning data and how both technical and non-technical team members can assist with this process.



- **Topics covered:**
  - **Profile the data**
    - **Define data's:**
      - Shape
      - Features
      - Limitations
    - **Content discovery**
      - what's in the data, what's missing
      - Null or ambiguous content
      - Incorrectly formatted data (ex: emails)
      - Range (and other numerical/quant) analysis
    - **Relationship discovery**
      - matchable / relatable datasets
      - vectors for linkage
      - Where is the data in use
  - **Prepare**
    - **Clean the data**
    - **Wrangle/munge data into usable form**
    - **Queries**

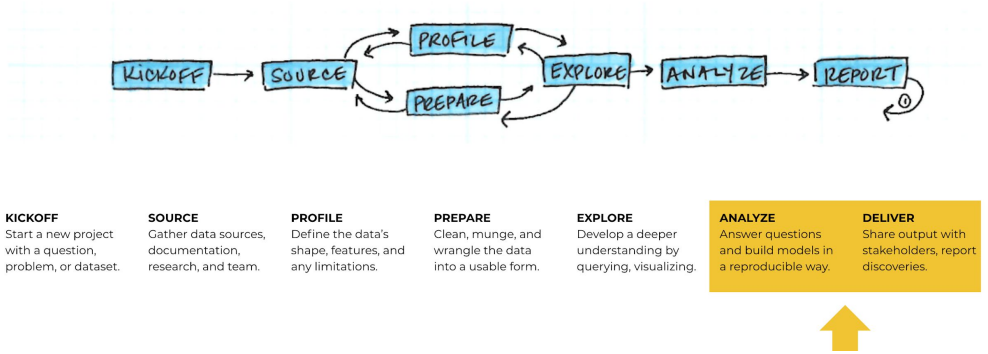
This workshop will help you develop a deeper understanding of your data through querying, visualizing, or other initial exploration techniques.

- Topics covered:
  - Query building
  - Data slicing / sampling
  - Exploratory viz (including the difference between exploratory and reporting viz)
  - Beyond statistics (i.e. why summary stats shouldn't be the only metrics you look at)



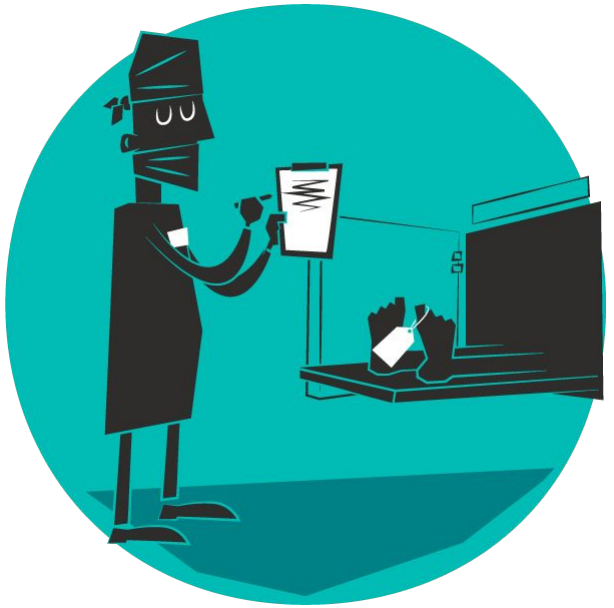
This workshop will help you to understand how to build models and answer questions in a reproducible way and then deliver the results of that work to your target audience.

- Topics covered:
  - Ask
    - What question do you want to answer?
    - What model will help you answer that question?
  - Analyze
    - An introduction to modeling
    - Building dashboards and consumer-ready viz
    - Notebooks and you!
  - Report
    - How to make sure your viz matches your audience
    - Building diversity of outputs into your workflow
    - Data viz best practices (sample - full course teaser)



This workshop will help to guide you through the process of evaluating your team performance and project success.

- **Topics covered:**
  - **Pre-post-mortem questionnaire**
  - **Gathering and using both qualitative and quantitative information**
  - **Include space for subjective input, otherwise it'll be interspersed throughout**
  - **Evaluate:**
    - Team effectiveness
    - Project success
    - Goals met?
    - Stretch goals
    - What worked/didn't
    - Quality of output
    - Auditability



When it comes to data visualization there are many options available. This workshop will compare and contrast many of the popular data visualization tools and why different tools might make sense for different use cases.

- **Topics covered:**
  - **Defining data visualization**
  - **In Depth Exploration**
    - Tableau
    - Power BI
    - Google Data Studio
    - R Studio
    - D3 / Vega / Vega-Lite (ChartBuilder)
  - **Other Options**
    - Excel
    - Infogram
    - ChartBlocks
    - Datawrapper
    - Plotly
    - RAW
    - Visual.ly
    - Ember Charts
    - NVD3
    - Fusion Charts
    - Highcharts
    - Chart.js
    - ...ad infinitum

Data Scientist is the hottest new job title, and people with many different backgrounds are choosing to pivot their careers to become data practitioners. This workshop will help those coming from a developer background to understand the similarities and differences between computer science / software development and data science.

- Topics covered:
  - Statistics
  - Data processing using code and notebooks
  - Reproducibility
  - Provenance

Too often, many data practitioners dismiss their role in the data ecosystem because they aren't writing code. Talented spreadsheet users, database admins, database admins, and a host of others all have the ability to contribute heavily to the data work within an organization.

This workshop will cover how to be as impactful as possible, even if you aren't a technical user.

- Topics covered:
  - Introduction to queries
  - Formulas and calculations
  - How to deliver/share your data and insights with a team
  - Tools / Techniques
    - Spreadsheets
    - ETL
    - (dw)SQL

Too often, when we discuss data visualization it is to cover the tactical nature of how to perform a task within a specific piece of software (ex: “how to make a bar chart”). Rarely do we discuss the correct time to use different types of viz, and how to build them using established best practices for the greatest impact.

This workshop will cover the technology-agnostic best practices for visualization design.

- Topics covered:
  - Types of visualizations
  - When to use different families of viz
  - Data as art
  - How to avoid visual clutter
  - Data viz pitfalls

The pervasiveness of data, especially in business decisions, is exploding. Unfortunately many of those who are new to the data ecosystem don't always stop to think about whether just because they can do something doesn't necessarily mean they should. This course will examine the FORTS (Fairness, Openness, Reliability, Trust, and Social Behavior) framework along with establishing the principles for good ethical data work. This work is based on the Global Data Ethics Pledge from Data for Democracy.

- **Topics covered:**
  - **FORTS Framework**
    - Fairness
    - Openness
    - Reliability
    - Trust
    - Social Behavior
  - **Ethical Principles**
    - Consider informed and purposeful intent
    - Make best effort to guarantee security of data
    - Make best effort to protect anonymous data subjects
    - Practice responsible transparency
    - Foster diversity
    - Acknowledge and mitigate unfair bias
    - Hold up datasets with provenance
    - Respect relevant tensions of stakeholders
    - Communicate responsibly
    - Exercise ethical imagination; increase benefit, prevent harm
  - **Ethical best practices**

Invited Presentation

edX Project

- Marta Ziosi (AI for People)
- Draft Script: https://github.com/lfai/ai-ethics-training/blob/master/ai-ethics-outline.md
- Slack Channel: https://lfaifoundation.slack.com/archives/C015NJUMR3M
- https://www.edx.org/course/ethics-in-ai-and-big-data

LF AI & DATA

LF AI & Data - General Updates

LF AI & DATA

Welcoming New Members

Premier
- SAS Institute

General
- Index Analytics
- ING Bank
- Cloudera
- Precisely
- AlphaBravo
- Broadcom
- Databricks

Associate
- Aivancity School for Technology, Business & Society
- OpenI (OpenIntelligence)
- The Shanghai Open Source Information Technology Association
- Peng Cheng Laboratory

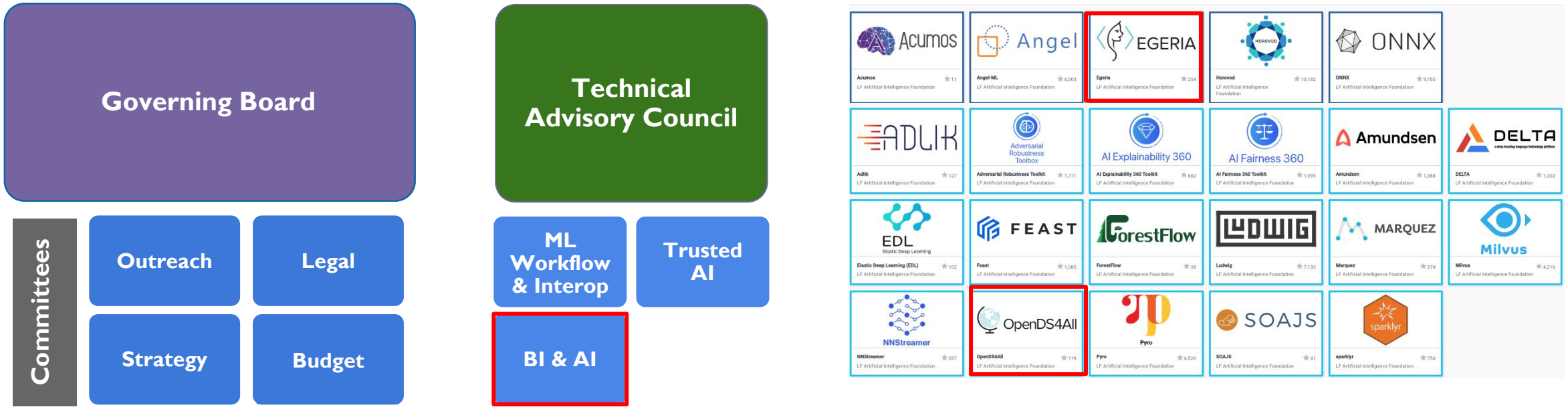
Structure and Governance

Foundation Governance
Funding effort to support hosted technical projects

Technical Coordination, New Initiatives

Hosted Projects

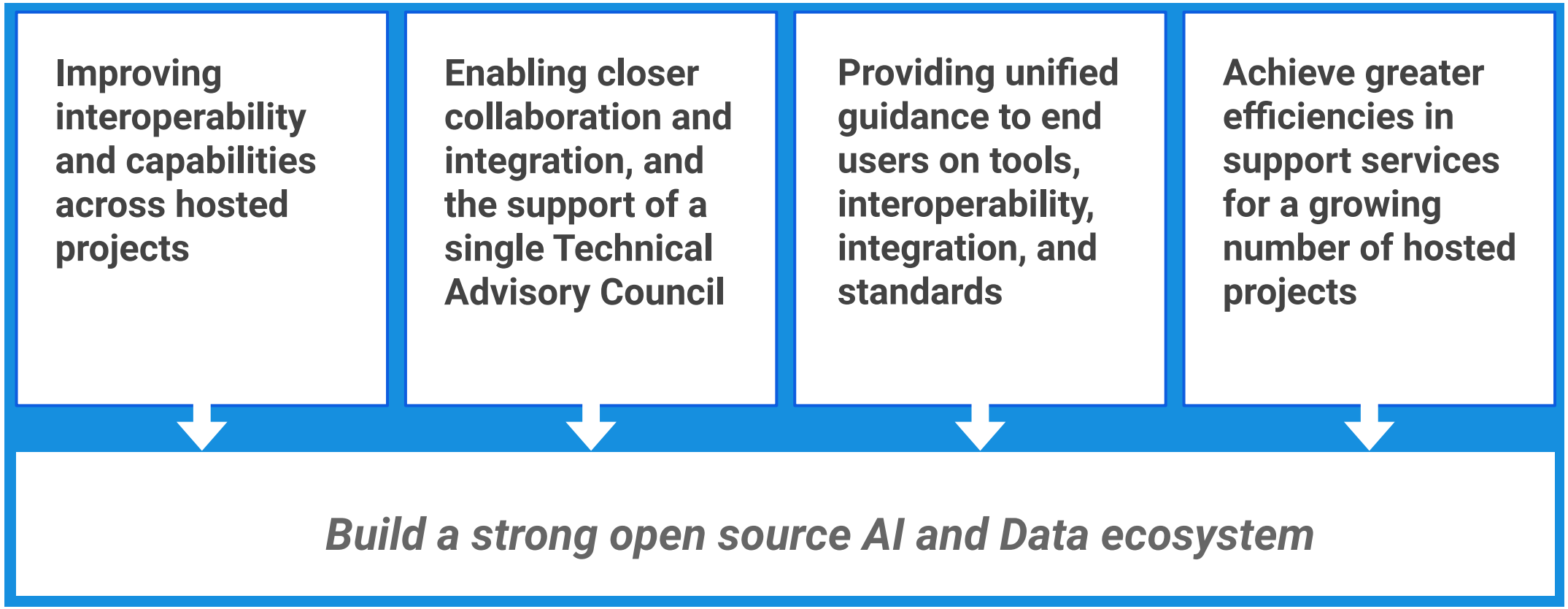
Projects have their own independent technical governance



Mission

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

Motivations for Harmonization



LF AI & DATA TIMELINE 2020

JANUARY 2020

- Milvus

MARCH 2020

- NNStreamer
- ForestFlow

APRIL 2020

- ONNX COMMUNITY VIRTUAL MEETUP

MAY 2020

- Ludwig

JUNE 2020

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

JULY 2020

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

SEPTEMBER 2020

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

OCTOBER 2020

- LF AI DAY ONNX COMMUNITY VIRTUAL MEETUP
- RELAUNCHING AS LF AI & DATA WITH INCREASED DATA FOCUS

INFO@LFAI.FOUNDATION

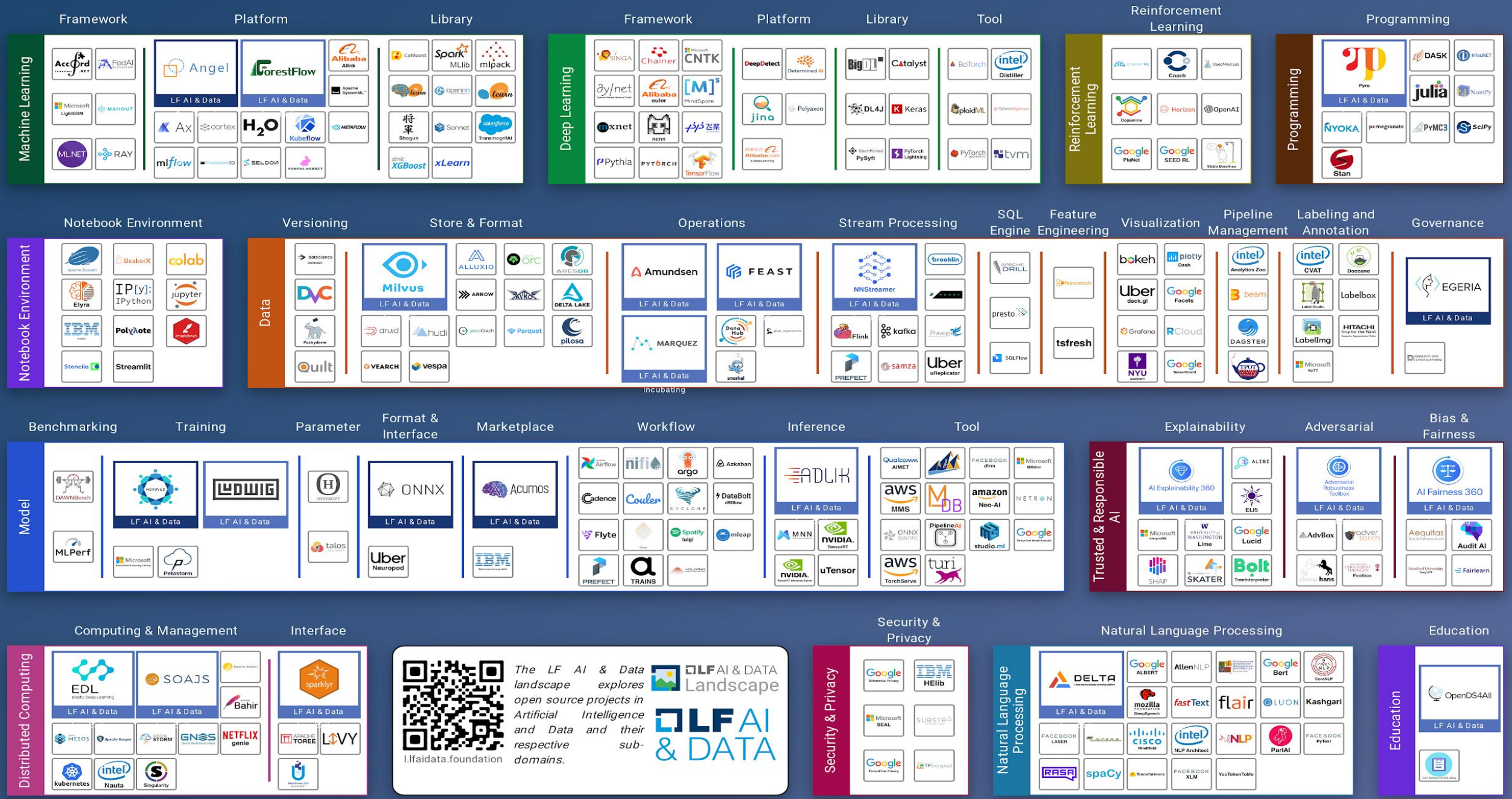
TECHNICAL PROJECTS | FOUNDATION | COMMUNITY

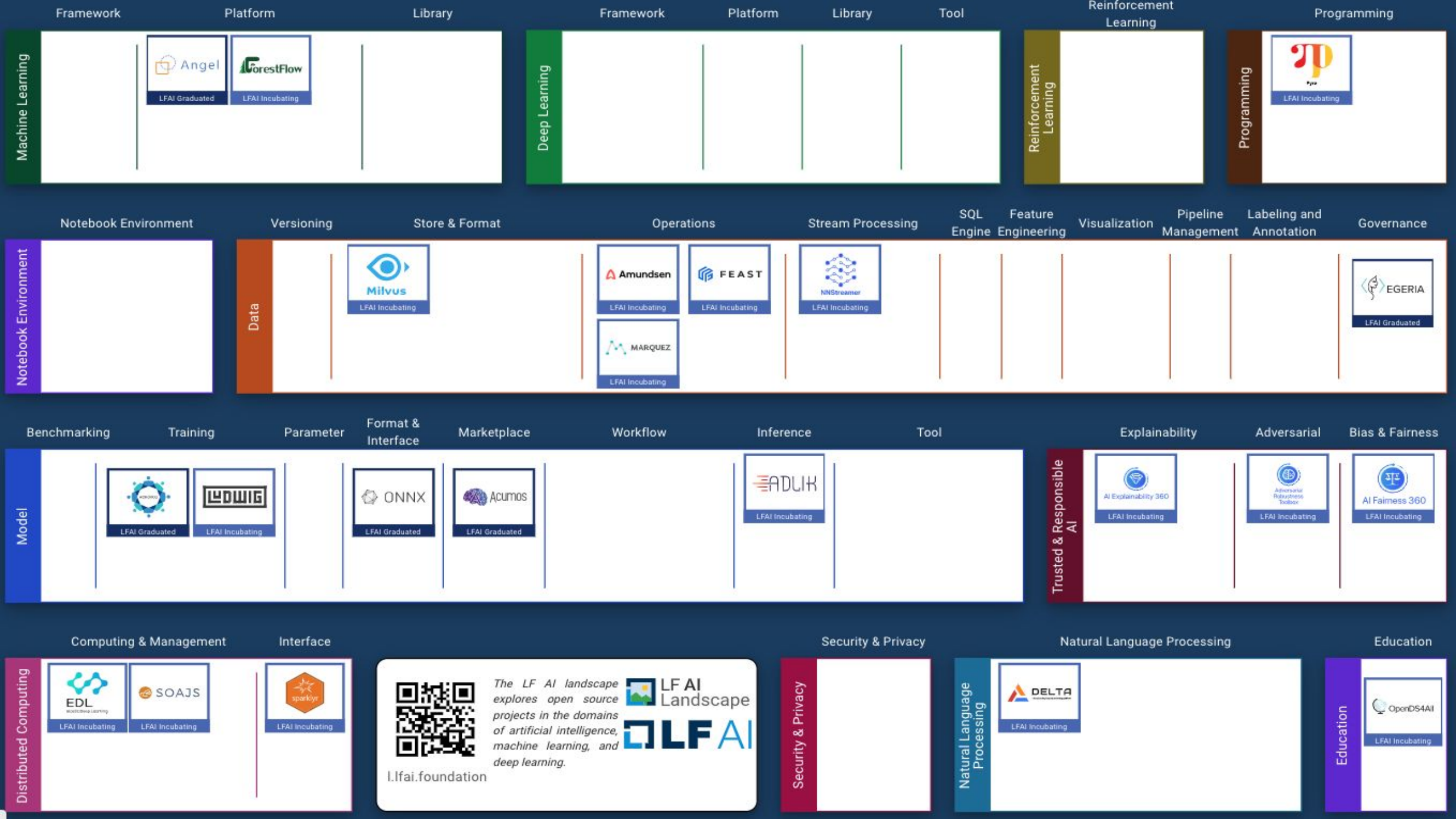
Linux Foundation AI Landscape

2020-10-27T19:08:54Z 4a5775b

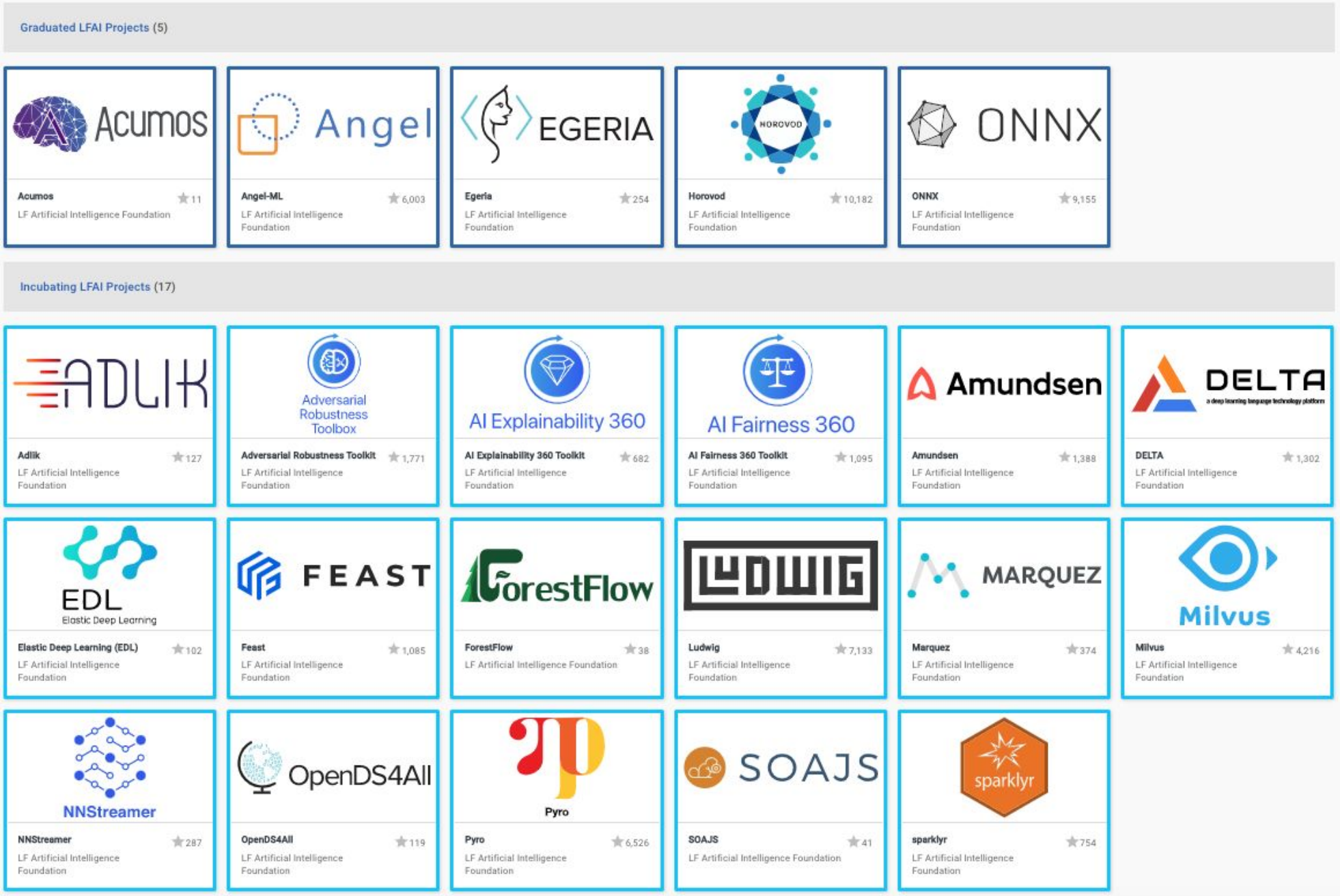
See the interactive landscape at l.lfai.foundation

Greyed logos are not open source




Framework
Platform
Library
Machine Learning
Angel
LFAI Graduated
ForestFlow
LFAI Incubating
Framework
Platform
Library
Tool
Deep Learning
Reinforcement Learning
Reinforcement Learning
Programming
Programming
Pyro
LFAI Incubating
Notebook Environment
Notebook Environment
Versioning
Store & Format
Operations
Stream Processing
SQL Engine
Feature Engineering
Visualization
Pipeline Management
Labeling and Annotation
Governance
Data
Milvus
LFAI Incubating
Amundsen
LFAI Incubating
FEAST
LFAI Incubating
MARQUEZ
LFAI Incubating
NNStreamer
LFAI Incubating
EGERIA
LFAI Graduated
Benchmarking
Training
Parameter
Format & Interface
Marketplace
Workflow
Inference
Tool
Model
LFAI Graduated
LUDWIG
LFAI Incubating
ONNX
LFAI Graduated
Acumos
LFAI Graduated
ADLIK
LFAI Incubating
Explainability
Adversarial
Bias & Fairness
Trusted & Responsible AI
AI Explainability 360
LFAI Incubating
Adversarial Robustness Toolbox
LFAI Incubating
AI Fairness 360
LFAI Incubating
Computing & Management
Interface
Distributed Computing
EDL
LFAI Incubating
SOAJS
LFAI Incubating
sparklyr
LFAI Incubating
l.lfai.foundation
The LF AI landscape explores open source projects in the domains of artificial intelligence, machine learning, and deep learning.
LF AI Landscape
LF AI
Security & Privacy
Security & Privacy
Natural Language Processing
Natural Language Processing
DELTA
LFAI Incubating
Education
Education
OpenDS4All
LFAI Incubating

A Growing LF AI & Data Project Portfolio and Community



LF AI & DATA

Companies hosting projects in LF AI

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

Looking to host a project with LF AI & Data

- Hosted project stages and life cycle:
  https://lfai.foundation/project-stages-and-lifecycle/
- Offered services for hosted projects:
  https://lfai.foundation/services-for-projects/
- Contact:
  Jim Spohrer (TAC Chair) and Ibrahim Haddad (ED, LF AI & Data)

Promoting Upcoming Project Releases

We promote project releases via a blog post and on LF AI & Data Twitter and/or LinkedIn social channels

For links to details on upcoming releases for LF AI & Data hosted projects visit the Technical Project Releases wiki

If you are an LF AI & Data hosted project and would like LF AI & Data to promote your release, reach out to pr@lfai.foundation to coordinate in advance (min 2 wks) of your expected release date.

Note on quorum

As LF AI & Data is growing, we now have 16 voting members on the TAC.

TAC representative - please ensure you attend the bi-weekly calls or email Jacqueline/Ibrahim to designate an alternate representative when you can not make it.

We need to ensure quorum on the calls especially when we have items to vote on.

Updates from Outreach Committee

LF AI & DATA

Upcoming Events

- Upcoming Events
  - Visit the LF AI & Data Events Calendar or the LF AI & Data 2020 Events wiki for a list of all events
  - To participate visit the LF AI & Data 2020 Events wiki page or email info@lfai.foundation

- Please consider holding virtual events

To discuss participation, please email events@lfai.foundation

Upcoming Events

End of the year wrap-up

Nov18 - Open Forum Europe - Turning Ethical AI into Technical Reality
https://www.openforumeurope.org/event/turning-ethical-ai-into-technical-reality/

Just completed

China Open Source Conference - Oct 25

LF AI Mini Summit at OSS EU – Virtual – Oct 29
October 29 @ 12:00 am
LF AI Foundation will hold a Mini Summit at Open Source Summit EU (OSS EU)

LF AI PR/Comms

- Please follow LF AI & Data on Twitter & LinkedIn and help amplify news via your social networks - Please retweet and share!
  - Also watch for news updates via the tac-general mail list
  - View recent announcement on the LF AI & Data Blog

- Open call to publish project/committee updates or other relevant content on the LF AI & Data Blog

- To discuss more details on participation or upcoming announcements, please email pr@lfai.foundation

Call to Participate in Ongoing Efforts

LF AI & DATA

Trusted AI

- **Leadership:**
  Animesh Singh (IBM), Souad Ouali (Orange), and Jeff Cao (Tencent)
- **Goal:** Create policies, guidelines, tooling and use cases by industry
- **Github:**
  https://github.com/lfai/trusted-ai
- **Wiki:**
  https://wiki.lfai.foundation/display/DL/Trusted+AI+Committee
- **To participate:**
  https://lists.lfai.foundation/g/trustedai-committee/
- **Next call:** Bi-weekly on Thursdays at 7am PT, subscribe to group calendar on wiki
  https://wiki.lfai.foundation/pages/viewpage.action?pageId=12091895
- **Slack: #trusted-ai-committee**
  - https://lfaifoundation.slack.com/archives/CPS6Q1E8G

ML Workflow & Interop

- **Leadership:**
  Huang "Howard" Zhipeng (Huawei)
- **Goal:**
  Define an ML Workflow and promote cross project integration
- **Github:**
  TBD
- **Wiki:**
  https://wiki.lfai.foundation/display/DL/ML+Workflow+Committee
- **To participate:**
  https://lists.lfai.foundation/g/mlworkflow-committee
- **Next call:** Every 4 weeks on Thursdays at 7:00 am PT, subscribe to group calendar on wiki
  https://wiki.lfai.foundation/pages/viewpage.action?pageId=18481242
- **Slack: #ml-workflow**
  - https://lfaifoundation.slack.com/archives/C011V9VSMQR

BI & AI

- **Leadership:**
  Cupid Chan (Index Analytics)
- **Goal:** Identify and share industry best practices that combine the speed of machine learning with human insights to create a new kind of business intelligence and better strategic directions for organizations.

- **Github:** https://github.com/odpi/bi-ai
- **Wiki:** TBD
- **To participate:** (check to see if set up)
  - https://lists.lfai.foundation/g/bi-ai-committee/
- **Next call:** Monthly community call TBD
- **Slack: #bi-ai-committee**
  - https://lfaifoundation.slack.com/archives/C01EK5ND073

Launching an effort to create AI Ethics Training

Initial developed course by the LF: Ethics in AI and Big Data - published on edX platform: https://www.edx.org/course/ethics-in-ai-and-big-data

The goal is to build 2 more modules and package all 3 as a professional certificate - a requirement for edX

- **To participate: https://lists.lfai.foundation/g/aiethics-training**

Upcoming TAC Meetings

LF AI & DATA

Upcoming TAC Meetings

- **November 19th:** Tentative new incubation project
- **December 3rd :** RosaNLG & tentative JanusGraph incubation project
- **December 17th:** Tentative Pyro graduated project proposal and additional new incubation project
- **December 31st :** No meeting - New Years Eve Holiday
- **January 14th:** TBD

Please send agenda topic requests to tac-general@lists.lfai.foundation

TAC Meeting Details

- To subscribe to the TAC Group Calendar, visit the wiki: https://wiki.lfai.foundation/x/XQB2
- Join from PC, Mac, Linux, iOS or Android: https://zoom.us/j/430697670
- Or iPhone one-tap:
  - US: +16465588656,,430697670# or +16699006833,,430697670#
- Or Telephone:
  - Dial(for higher quality, dial a number based on your current location):
  - US: +1 646 558 8656 or +1 669 900 6833 or +1 855 880 1246 (Toll Free) or +1 877 369 0926 (Toll Free)
- Meeting ID: 430 697 670
- International numbers available: https://zoom.us/u/achYtcw7uN

Open Discussion

LF AI & DATA

Mission

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

Legal Notice

- he Linux Foundation, The Linux Foundation logos, and other marks that may be used herein are owned by The Linux Foundation or its affiliated entities, and are subject to The Linux Foundation's Trademark Usage Policy at https://www.linuxfoundation.org/trademark-usage, as may be modified from time to time.
- Linux is a registered trademark of Linus Torvalds. Please see the Linux Mark Institute's trademark usage page at https://lmi.linuxfoundation.org for details regarding use of this trademark.
- Some marks that may be used herein are owned by projects operating as separately incorporated entities managed by The Linux Foundation, and have their own trademarks, policies and usage guidelines.
- TWITTER, TWEET, RETWEET and the Twitter logo are trademarks of Twitter, Inc. or its affiliates.
- Facebook and the "f" logo are trademarks of Facebook or its affiliates.
- LinkedIn, the LinkedIn logo, the IN logo and InMail are registered trademarks or trademarks of LinkedIn Corporation and its affiliates in the United States and/or other countries.
- YouTube and the YouTube icon are trademarks of YouTube or its affiliates.
- All other trademarks are the property of their respective owners. Use of such marks herein does not represent affiliation with or authorization, sponsorship or approval by such owners unless otherwise expressly specified.
- The Linux Foundation is subject to other policies, including without limitation its Privacy Policy at https://www.linuxfoundation.org/privacy and its Antitrust Policy at https://www.linuxfoundation.org/antitrust-policy. each as may be modified from time to time. More information about The Linux Foundation's policies is available at https://www.linuxfoundation.org.
- Please email legal@linuxfoundation.org with any questions about The Linux Foundation's policies or the notices set forth on this slide.

LF AI & DATA

LF AI & DATA

LF AI & DATA

Project Contribution Proposal - Name

LF AI & DATA

Project Contribution Proposal Review & Discussion: Project_Name

Project_Description.

Presenter: Name

Resources:

Github: ABC
Project Level: Incubation
Proposal: XYZ

TAC Vote on Project Proposal: Project_Name

Proposed Resolution:

The TAC approves the Project_Name as an Incubation project of the LF AI & Data Foundation

Next Steps

Onboard & Announcement:

LF AI & Data staff will work with Project_Name to onboard the project (e.g., Website, Logo, Foundation, Mailing List, Calendar, Zoom, Wiki, Slack, Repo, Presentations, etc.) leading to the announcement of the project joining LF AI & Data

Collaboration:

Explore potential integrations between the project and other LF AI & Data projects

Operations:

Integrate the project with LF AI & Data operations