Meeting of the Technical Advisory Council (TAC)

May 6, 2021

LF AI & DATA

Antitrust 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: lfaidata.foundation
- Wiki: wiki.lfaidata.foundation
- GitHub: github.com/lfaidata
- Landscape: https://landscape.lfaidata.foundation or https://l.lfaidata.foundation
- Mail Lists: https://lists.lfaidata.foundation
- Slack: https://slack.lfaidata.foundation
-
- LF AI Logos: https://github.com/lfaidata/artwork/tree/master/lfaidata
- LF AI Presentation Template: https://drive.google.com/file/d/1eiDNJvXCqSZHT4Zk_-czASlz2GTBRZk2/view?usp=sharing
-
- Events Page on LF AI Website: https://lfaidata.foundation/events/
- Events Calendar on LF AI Wiki (subscribe available): https://wiki.lfaidata.foundation/pages/viewpage.action?pageId=12091544
- Event Wiki Pages: https://wiki.lfaidata.foundation/display/DL/LF+AI+Data+Foundation+Events

Agenda

- Roll Call (2 mins)
- Approval of Minutes from April 22nd (3 mins)
- Welcome any new Members

- Two Sandbox Proposals (50 minutes = 2 x 25 minutes)
  - ML eXchange (MLX) (Animesh Singh)
  - Vulcan Kompute (Alejandro Saucedo)

- LF AI General Updates (2 minutes)
- Open Discussion (3 minutes)

TAC Voting Members

* = still need backup specified on wiki

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
OPPO	Yifan Xi*	jieyifan@oppo.com
SAS	Nancy Rausch	nancy.rausch@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
Pyro	Fritz Obermeyer*	fritz.obermeyer@gmail.com

Approval of April 22th, 2021 Minutes

Draft minutes from the April 22th TAC call were previously distributed to the TAC members via the mailing list

Proposed Resolution:

- That the minutes of the April 22th meeting of the Technical Advisory Council of the LF AI & Data Foundation are hereby approved.

Members (45)

https://landscape.lfai.foundation/card-mode?project=company

- **New Members**
  - **Premier:** Guangdong OPPO Mobile Telecommunications Corp (China)
  - **Associate:** Chaitanya Bharathi Institute Of Technology (India)

Sandbox Proposal - ML eXchange (MLX)

Animesh Singh (IBM) <asingh@us.ibm.com>

LF AI & DATA

Project Contribution Proposal Review & Discussion: Machine Learning eXchange (MLX)

We are open sourcing **Machine Learning Exchange (MLX)** , a Data and AI Assets Catalog and Execution Engine. It allows upload, register, execute, and deploy: AI pipelines and Components, Models, Datasets, Notebooks. Additionally it provides: Automated sample pipeline code generation to execute egistered models, datasets and notebooks, Pipelines Engine powered by Kubeflow Pipelines on Tekton, core of Watson Pipelines, Serving engine by KFServing (Next gen base for WML), Datasets Management by Datashim, Preregistered Datasets from Data Asset Exchange (DAX) and Models from Model Asset Exchange(MAX), Model Metadata schema aligned with MLSpec.

We have also been exploring integration with Acumos projects.

Presenter: Animesh Singh (IBM) <asingh@us.ibm.com)

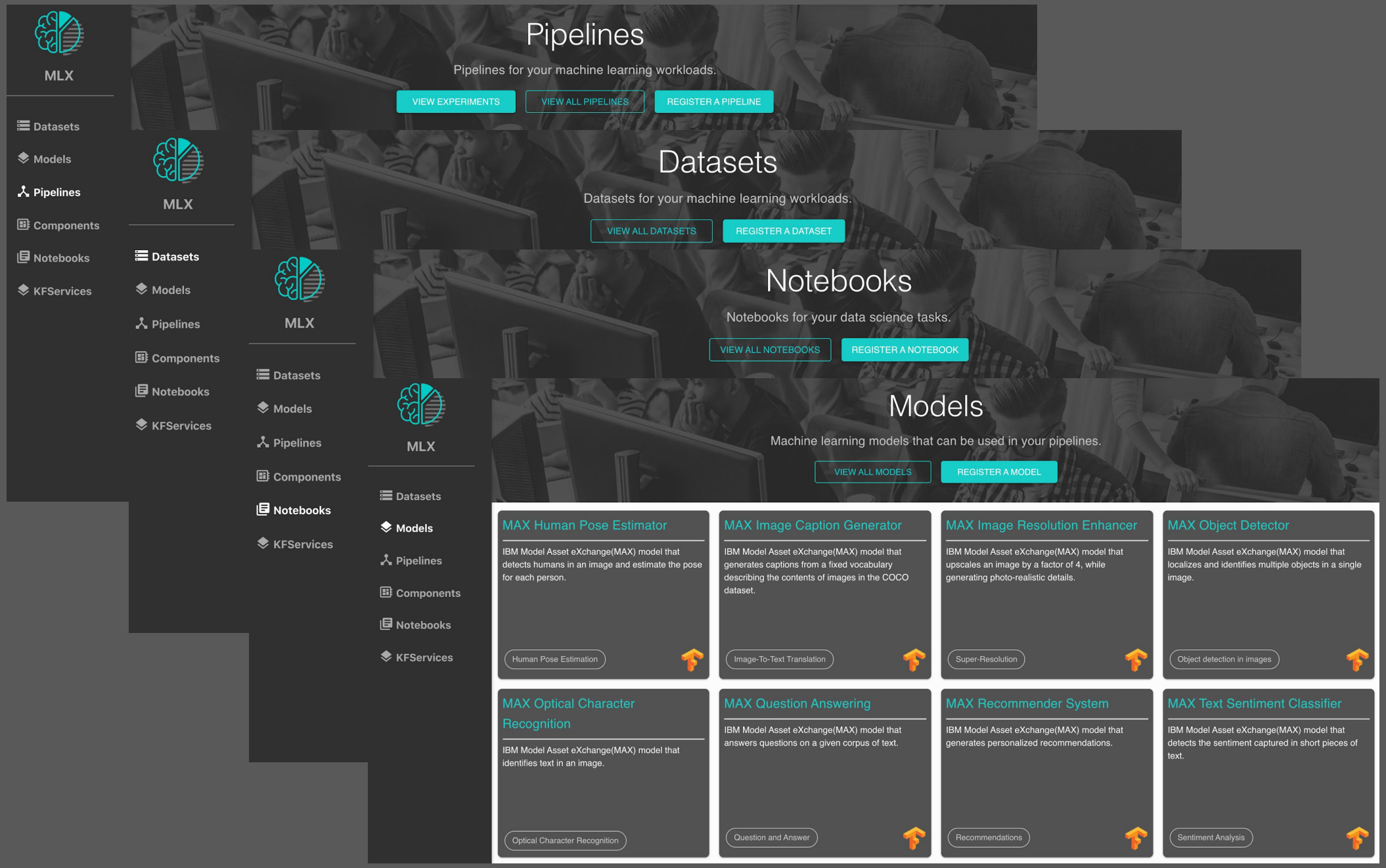
Resources:

Github: https://github.com/machine-learning-exchange
Project Level: Sandbox
Proposal: https://github.com/lfai/proposing-projects/pull/42/files

Machine Learning Exchange (MLX)

Animesh Singh, Christian Kadner, Tommy Li

Machine Learning Exchange (MLX) : Data and AI Assets Catalog and Execution Engine



Machine Learning Exchange (MLX)

- **Data and AI Assets Catalog and Execution Engine**

- Upload, register, execute, and deploy

  -AI pipelines and Components

  -Models

  -Datasets

  -Notebooks

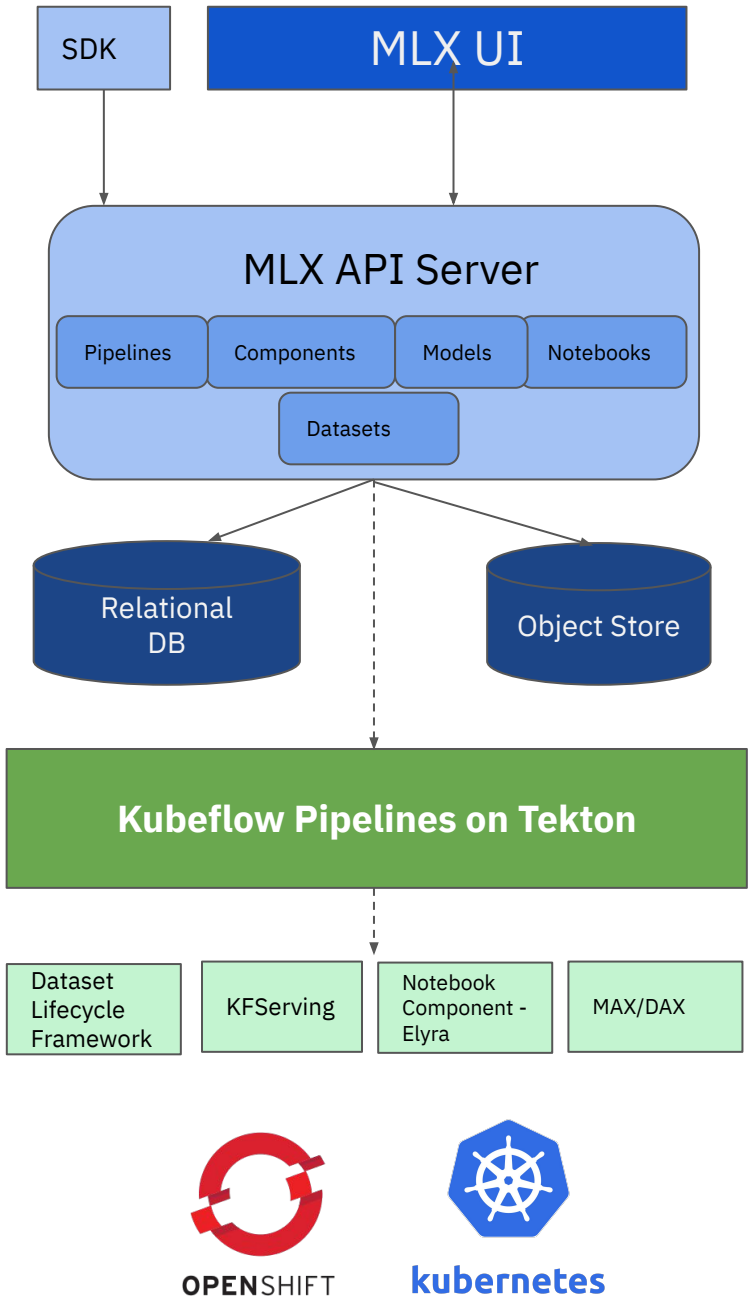
- Automated sample pipeline code generation to train, validate, serve your registered models, datasets and notebooks

- Pipelines Engine powered by Kubeflow Pipelines on Tekton, core of Watson Pipelines

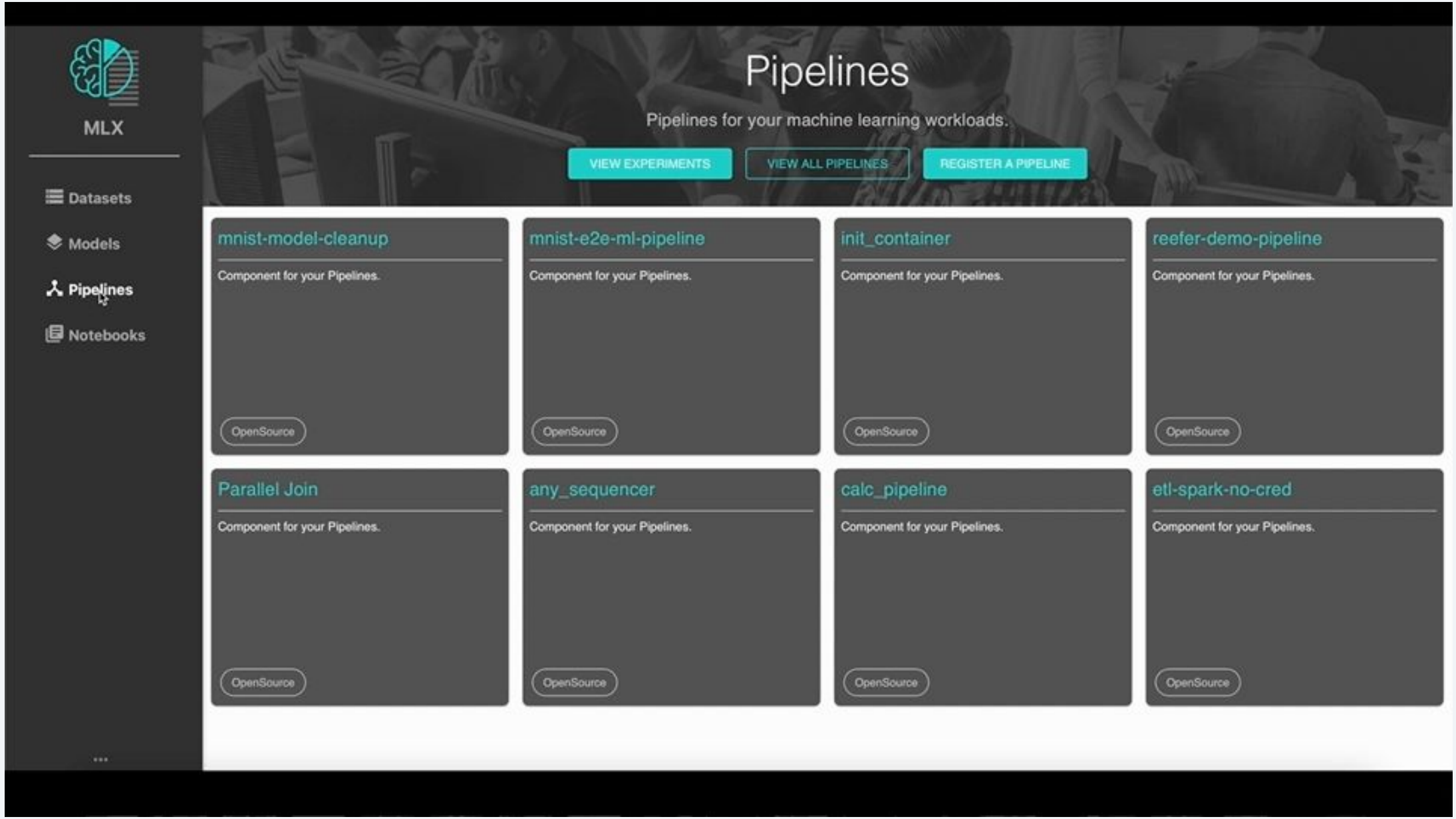
- Serving engine by KFServing (Next gen base for WML) , Datasets Management by Dataset Lifecycle Framework

- Preregistered Datasets from Data Asset Exchange (DAX) and Models from Model Asset Exchange (MAX)

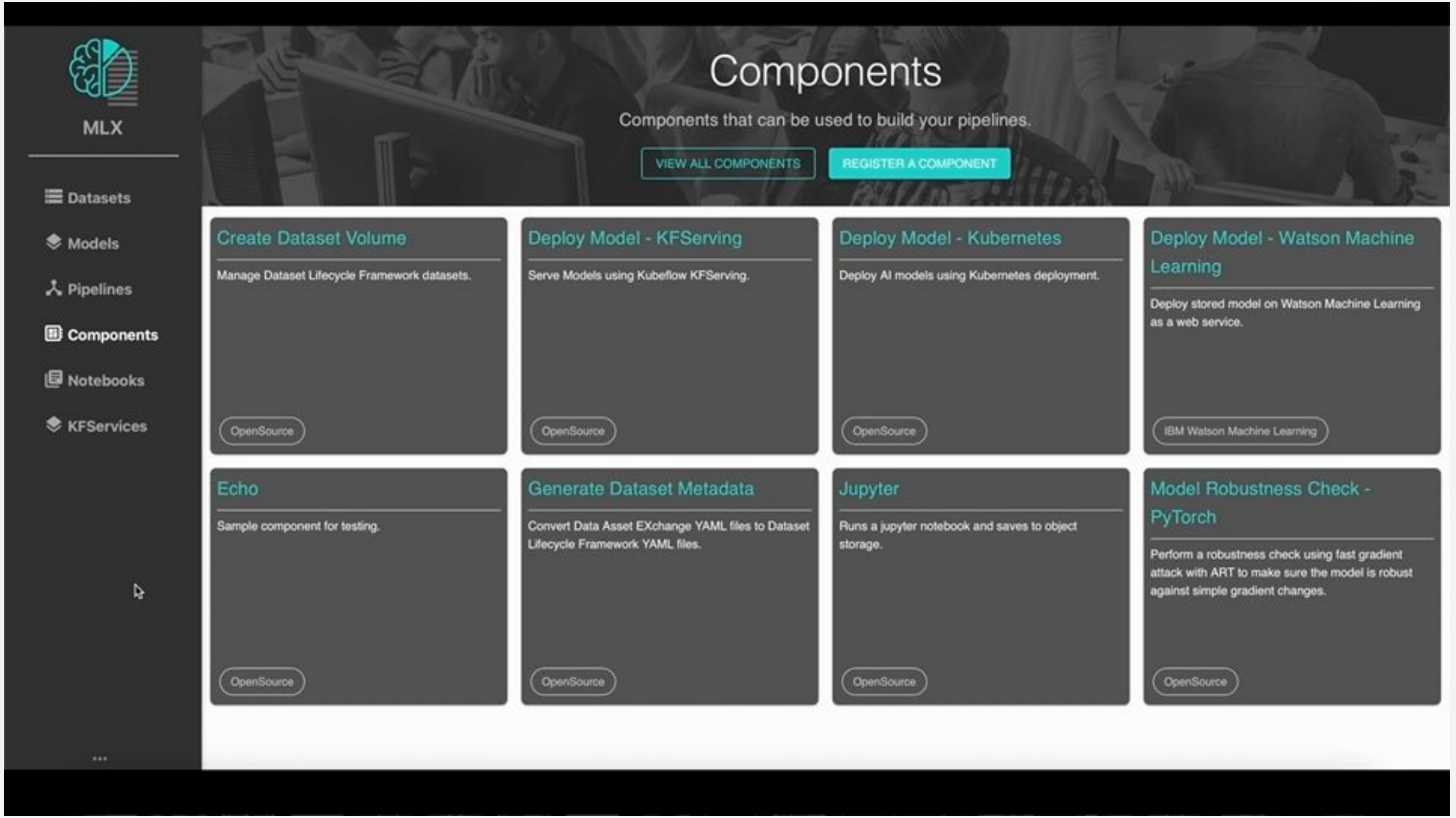
- Model Metadata schema aligned with MLSpec



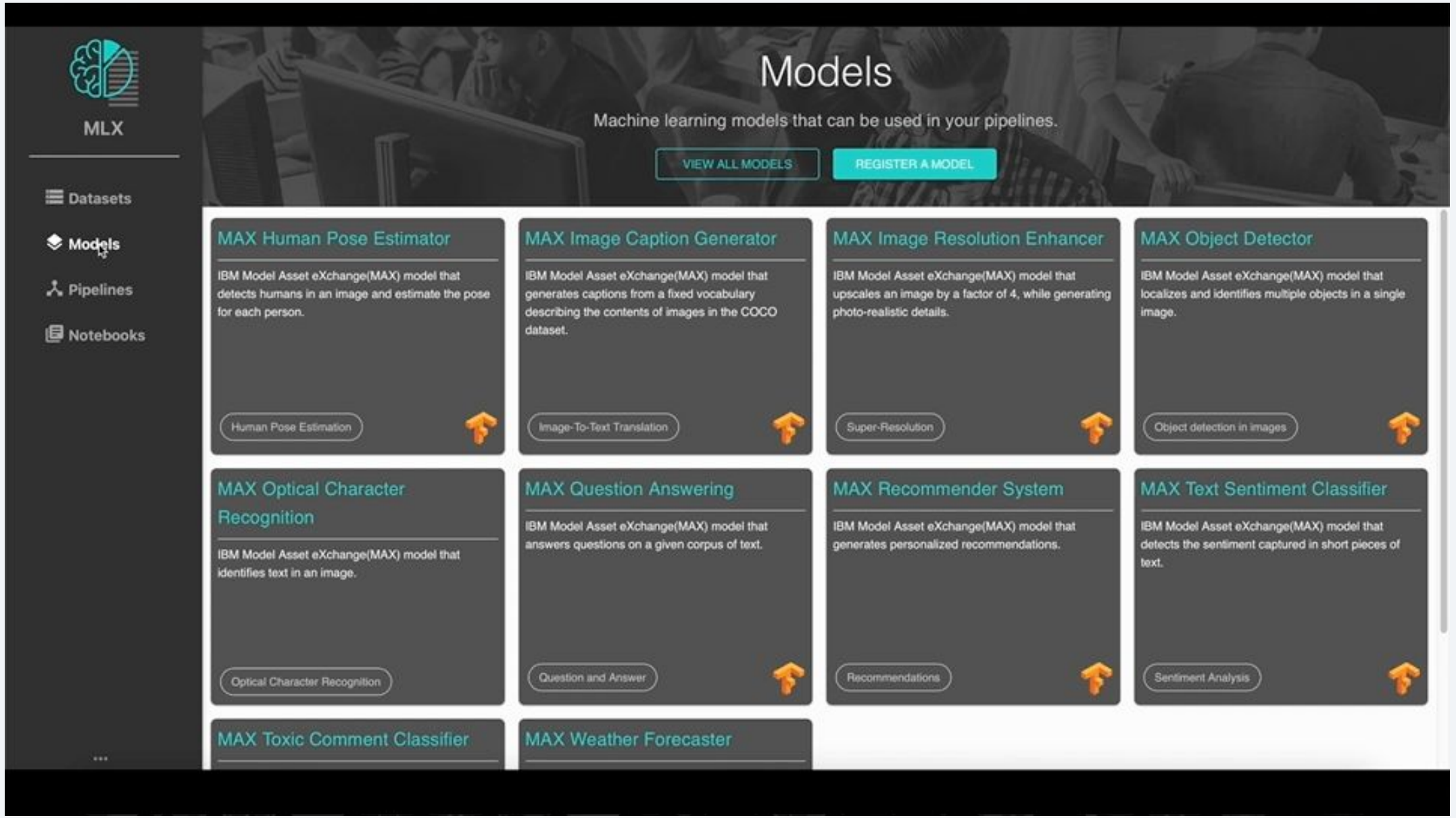
View, download, and execute Pipelines



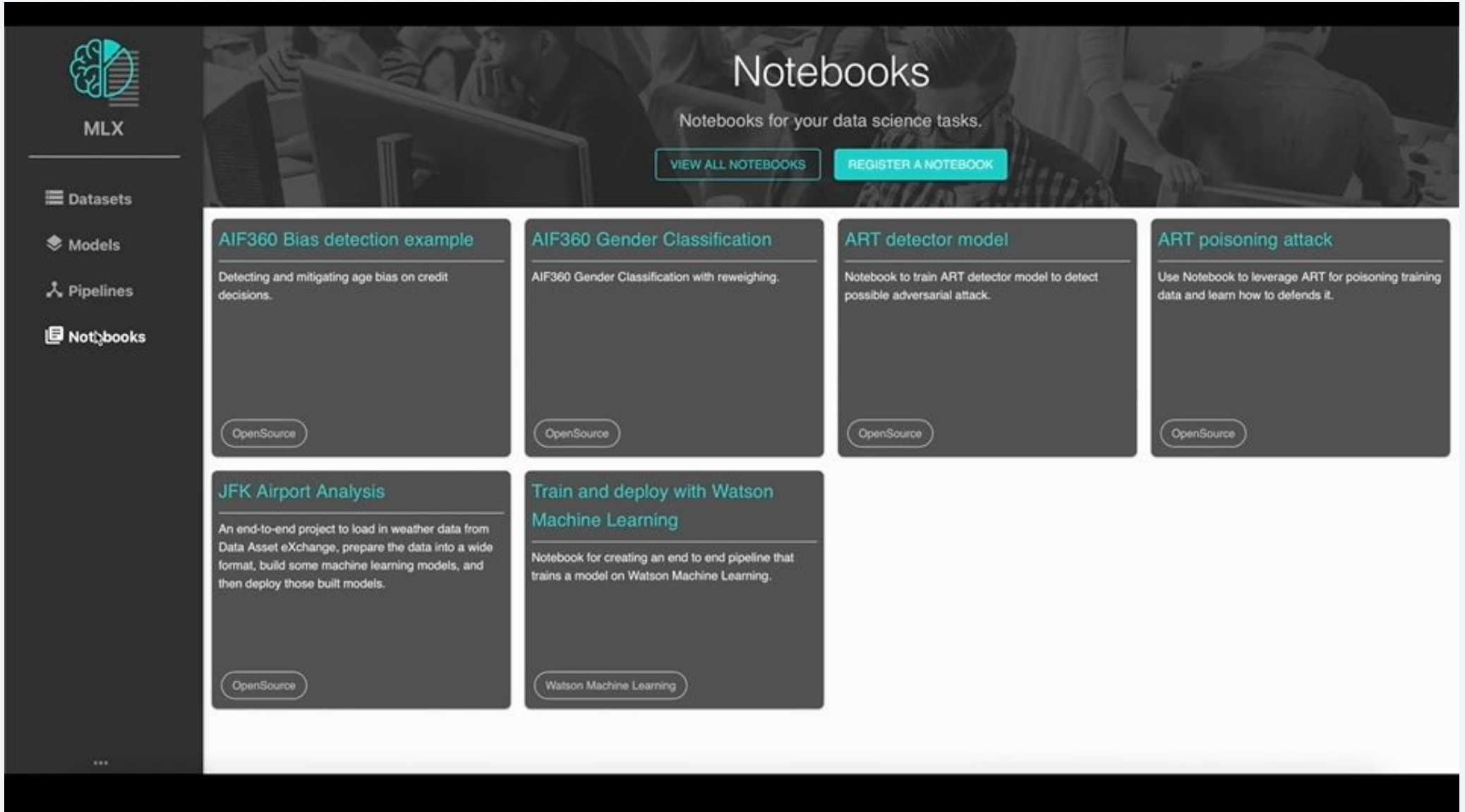
View, download, and execute Pipeline Components



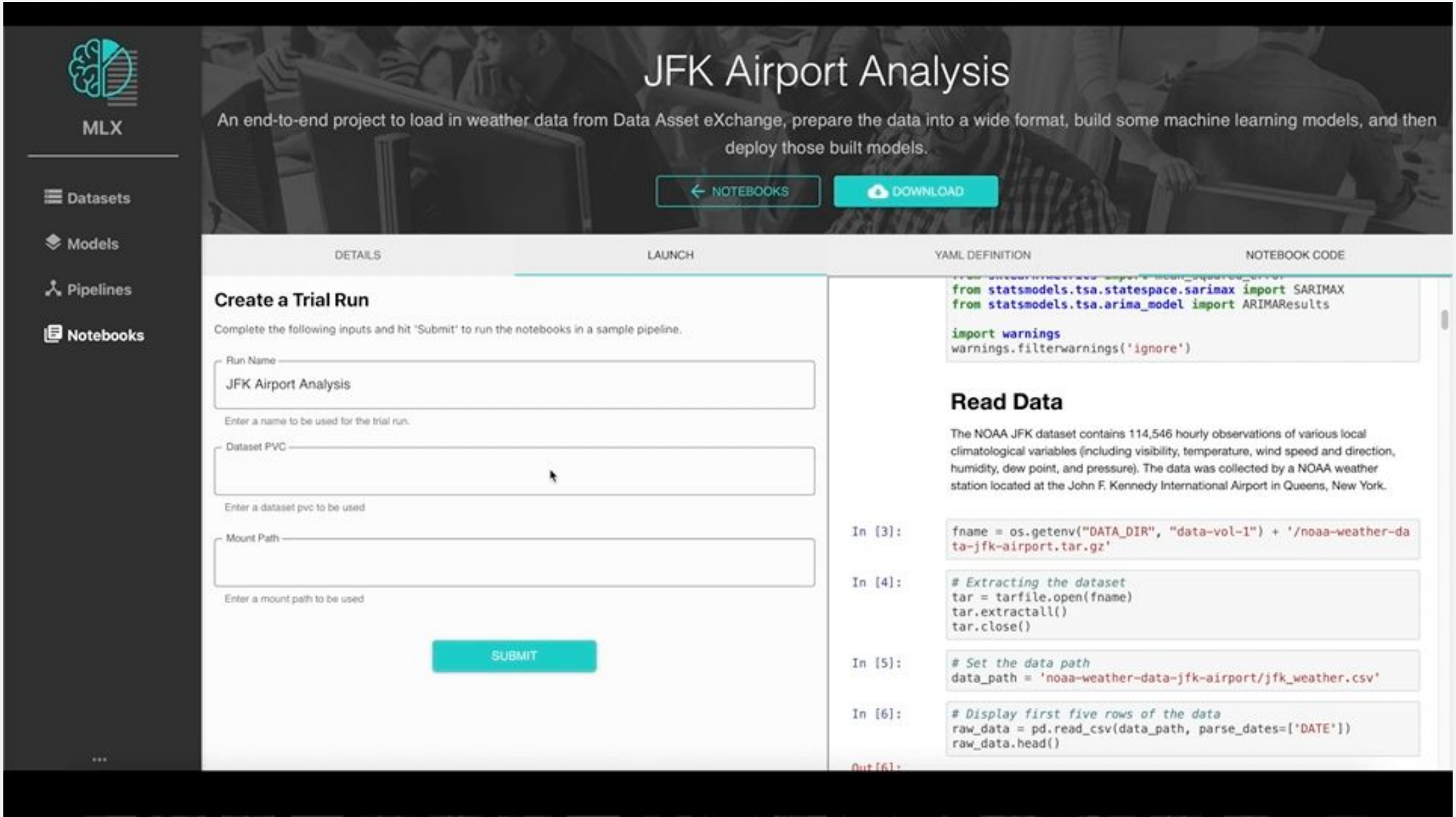
Library of prepackaged models. Register your own models, run with Pipelines



Library of prepackaged notebooks. Register your own notebooks

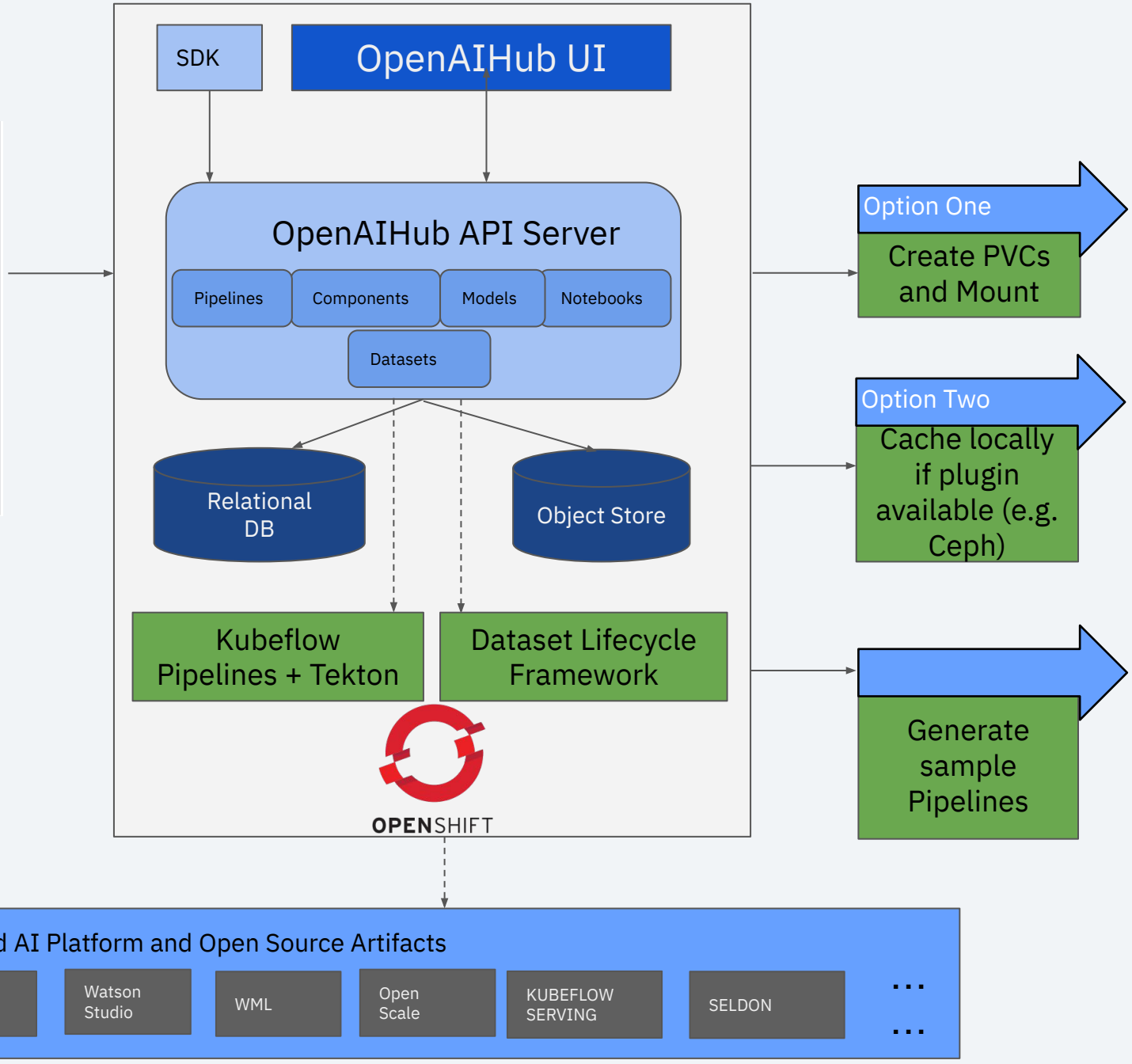


Run Notebooks using Pipelines

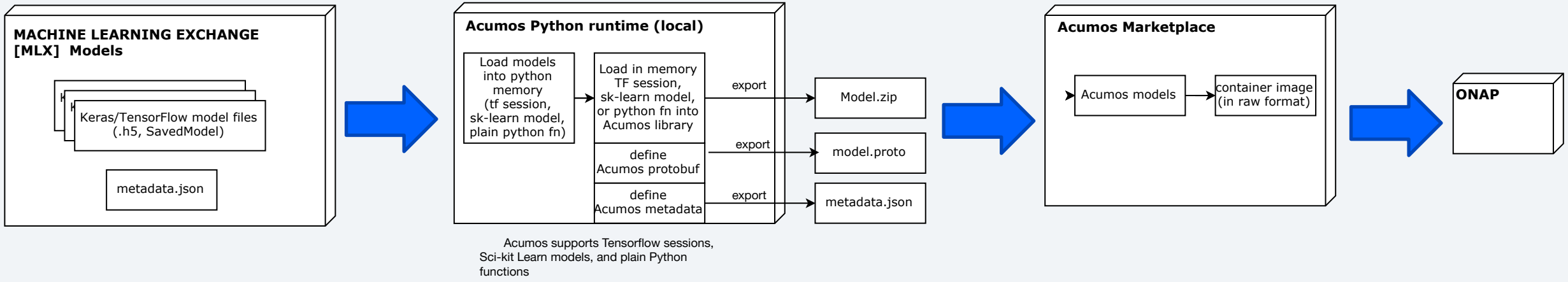


Dataset Integration

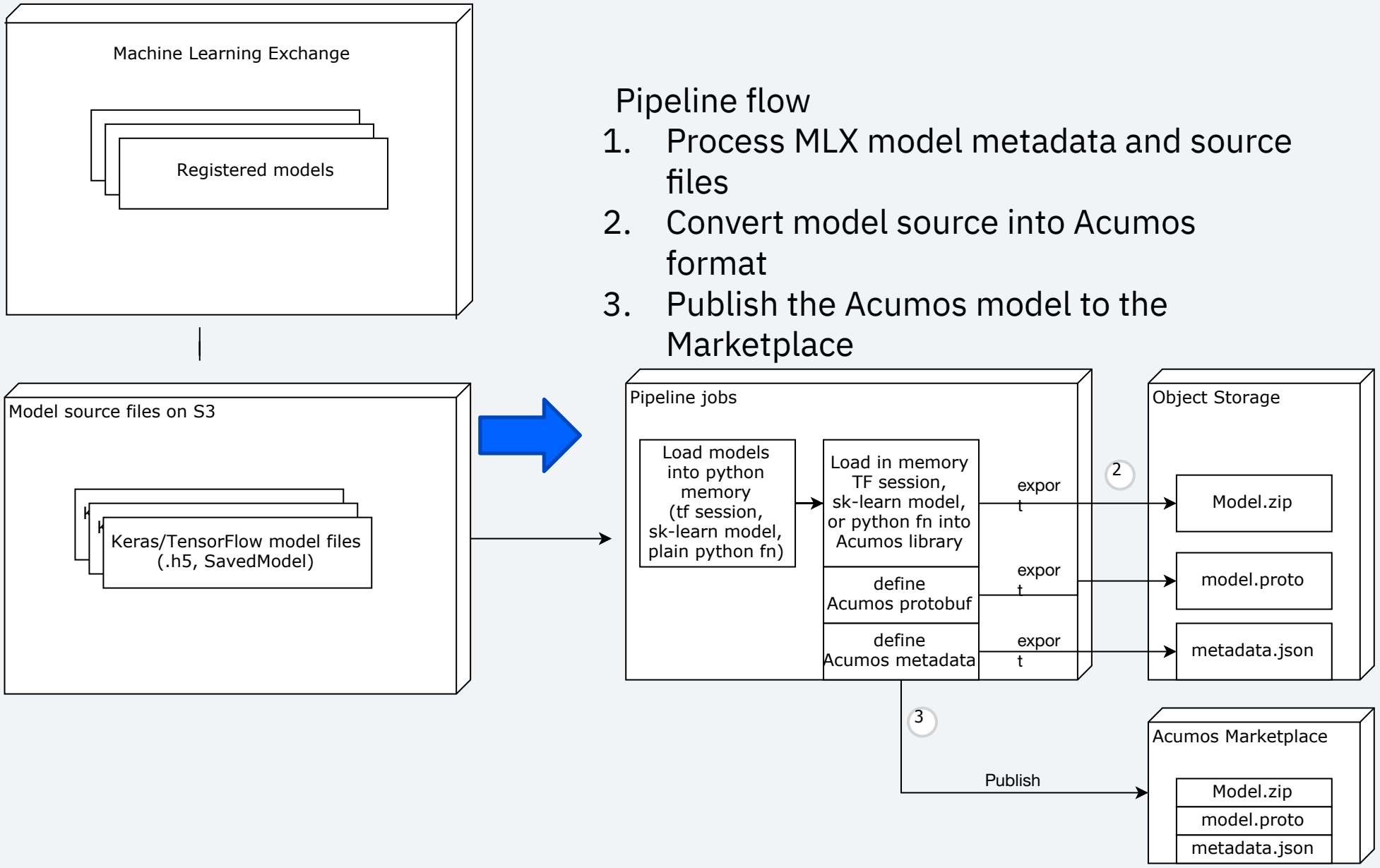
```yaml
basic:
  dataset_name: Finance Proposition Bank
  short_description: Text from approximately 1000 English sentences obtained from IBM's public annual financial reports, annotated with a l
  source: Data Asset eXchange
  source_link: https://developer.ibm.com/exchanges/data/all/finance-proposition-bank/
  download_link: https://dax-cdn.cdn.appdomain.cloud/dax-finance-proposition-bank/1.0.2/finance_proposition_bank.tar.gz
  format: CoNLL-U
# Let's day user prefers to use this data. Next step is to check the license details.
usage:
  license: CDLA-Sharing
  license_link: https://cdla.io/sharing-1-0/
  domain: Natural Language Processing
# How big is the data? What areas do they cover?
statistics:
  number_of_records: ~1,000 annotated sentences corresponding to ~50,000 words
  size: 2.9 MB
  coverage: This dataset contains labeled sentences from IBM's publicly available annual financial reports.
# Let's say data coverage is ok - next step is to preview the data and know about it
explore_data:
  data_preview_glossary: https://dax-cdn.cdn.appdomain.cloud/dax-finance-proposition-bank/1.0.2/data-preview/index.html
# Where do they come from? These are additional details about the data in case user prefers to learn more about.
origin:
  source: IBM Research
# starter code to play with the data
assets:
  ws_project: https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/0e615c46-5e4c-496f-9374-25dde48b46d0/view?access_token=aa16e0d5e344
updates:
  latest: September 12, 2019
keywords: Artificial Intelligence, Natural Language Processing, Text
```



MLX and Acumos – High Level Flow

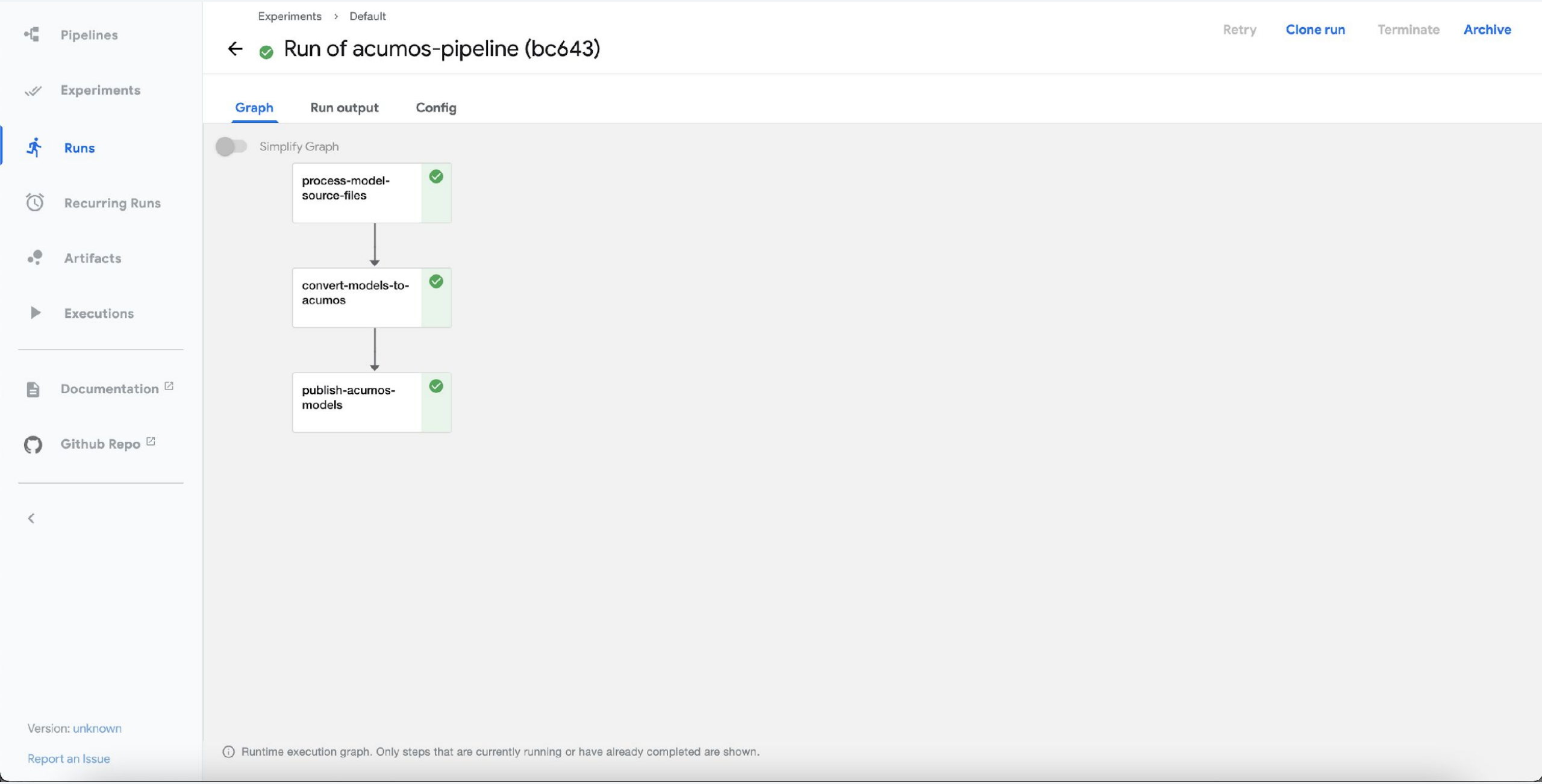


MLX to Acumos Marketplace Pipeline



Pipeline flow

1. Process MLX model metadata and source files
2. Convert model source into Acumos format
3. Publish the Acumos model to the Marketplace


Pipelines
Experiments
Runs
Recurring Runs
Artifacts
Executions
Documentation
Github Repo
Version: unknown
Report an Issue
Experiments › Default
Run of acumos-pipeline (bc643)
Retry
Clone run
Terminate
Archive
Graph
Run output
Config
Simplify Graph
process-model-source-files
convert-models-to-acumos
publish-acumos-models
Runtime execution graph. Only steps that are currently running or have already completed are shown.

Sandbox Proposal - Vulcan Kompute

Alejandro Saucedo <axsauze@gmail.com>

LF AI & DATA

Project Contribution Proposal Review & Discussion: Vulcan Kompyte

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 usecases. Typical usecases are: General Purpose GPU Computing, Develop GPU accelerated kernels for advanced data processing use-cases, Extend scientific applications to enable for mobile and cross vendor GPU support.

Presenter: Alejandro Saucedo <axsauze@gmail.com>

Resources:

Github: https://github.com/EthicalML/vulkan-kompute
Project Level: Sandbox
Proposal: https://github.com/lfai/proposing-projects/pull/40/commits/cf502f741f045ee3ffe080165ef5c3b550a2fc40

a@ethical.institute

Vulkan Kompute

Linux Foundation Proposal for Vulkan Kompute as Sandbox project focused on advancement of cross-vendor GPGPU

Alejandro Saucedo

@AxSaucedo

High level Objectives

Motivations & Background

Objectives of Kompute Initiative

Project Features

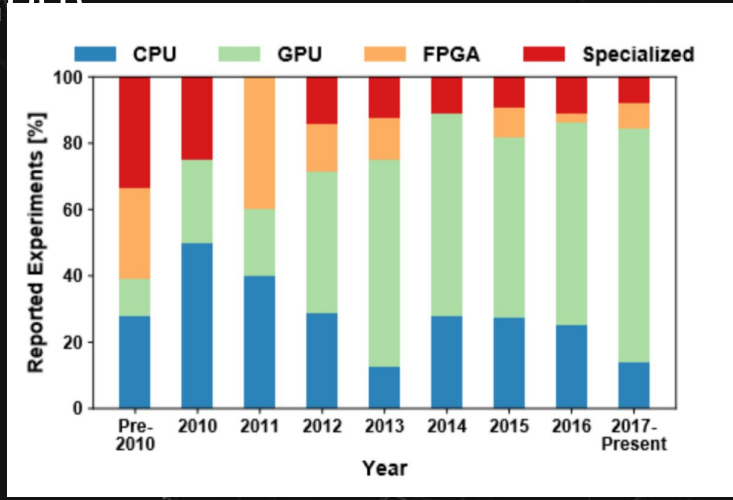
Linux Foundation Proposal

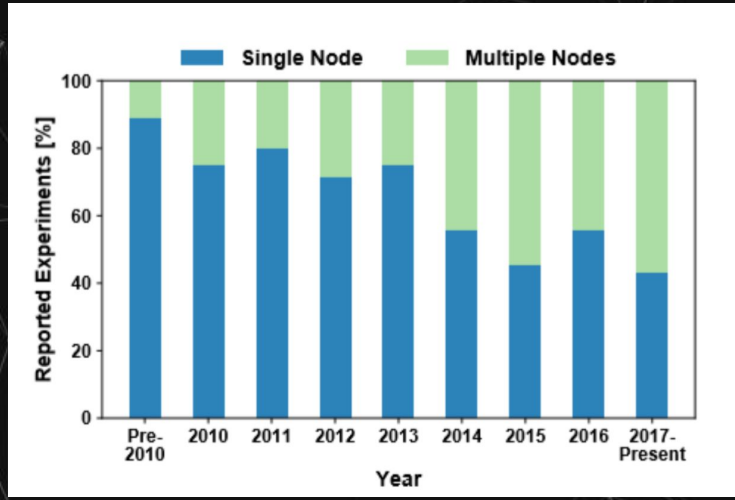
Collaboration with LF projects

Next Steps

Parallel Compute Motivations

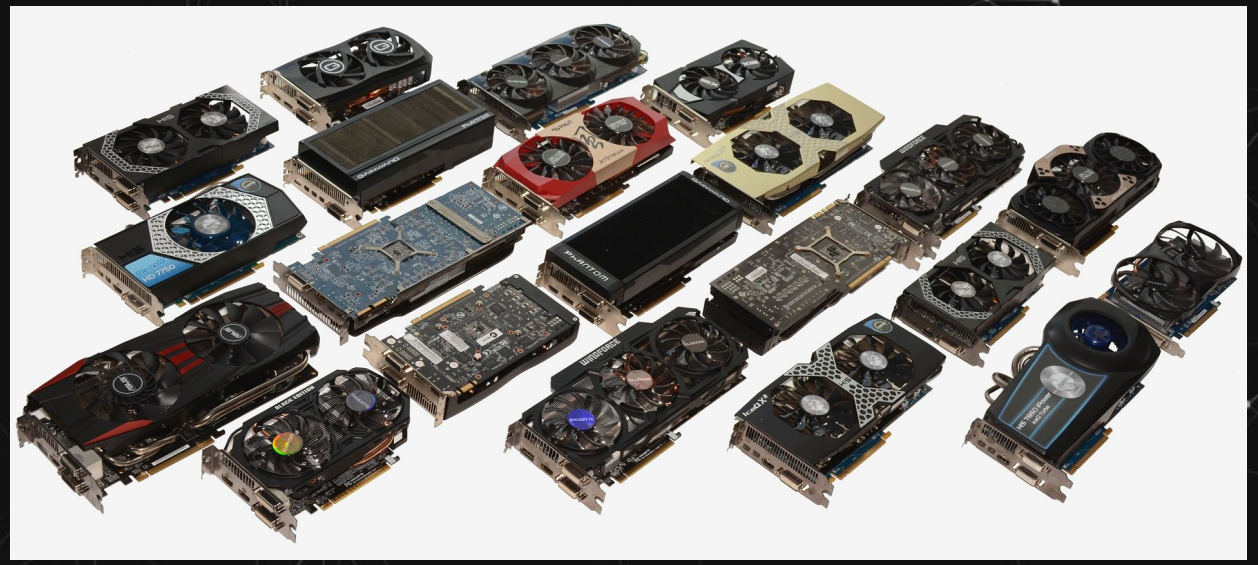
- Fast-growing increase in adoption of standard and specialised GPUs for general compute
- Functions can often be reduced to highly parallelizable stages (Matrix Mult, ML Layers, etc)
- Micro-batching allows for further parallelization of multiple inputs (eg. cost instead of loss)
- Breaking up fractions of each ensemble comp. across tightly coupled hardware (eg. multi-GPU)





Ben-Nun, Tal, and Torsten Hoefler. "Demystifying parallel and distributed deep learning: An in-depth concurrency analysis." ACM Computing Surveys (CSUR) 52.4 (2019): 1-43.

Parallel Processing: Options



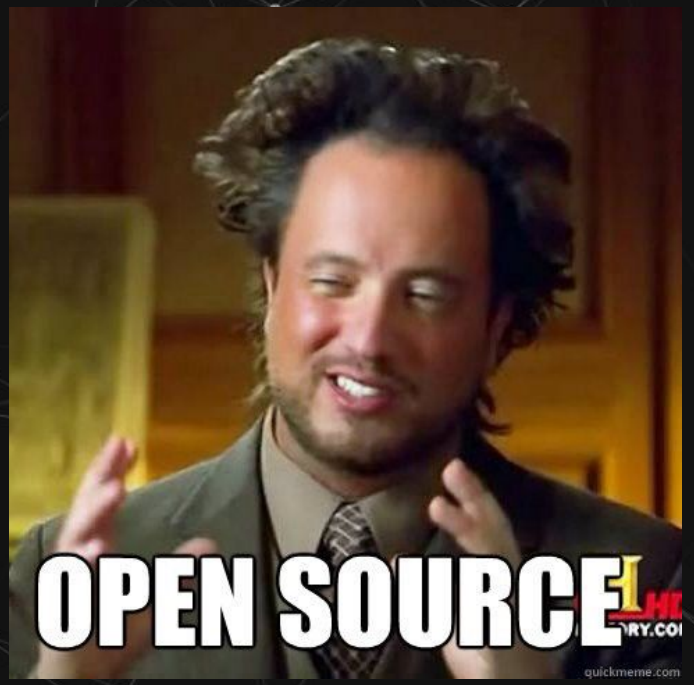
Introducing Vulkan

Created by the Khronos group

The Khronos Group, Inc. is a non-profit member-funded industry consortium, focused on the creation of open standard, royalty free APIs for authoring and accelerated playback of dynamic media on a wide variety of platforms and devices.

Top Vulkan Priorities

1. Performance
2. Interoperability
3. Performance



Oregon State University, SIGGRAPH 2020 Lecture Slides
http://web.engr.oregonstate.edu/~mjb/vulkan/

Khronos Members

GPU users, vendors & suppliers all support, contribute, and further theses initiatives

Oregon State University, SIGGRAPH 2020 Lecture Slides http://web.engr.oregonstate.edu/~mjb/vulkan/

Led & Supported by Key Players

https://developer.nvidia.com/vulkan

Vulkan C++ SDK

Advantages

- Low level with rich access to components
- Explicit and verbose on what is being achieved, with C-style API as core
- A broad range of top players leading the development of the framework
- Highly compatible across different platforms, mobile, and different suppliers

Disadvantages

- Low level with rich access to components
- Explicit and verbose on what is being achieved, with C-style API as core
- A broad range of top players leading the development of the framework
- Highly compatible across different platforms, mobile, and different suppliers


Only takes about
500-2000+ lines of C++ code...

Major Projects Code Replication

Alibaba Group 阿里巴巴集团

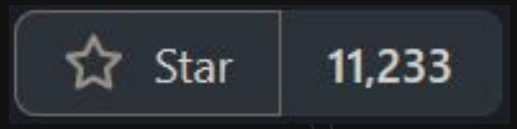
MNN
Mobile Neural Network

facebook

PyTorch

Tencent 腾讯

ncnn



Google

TensorFlow 2.0

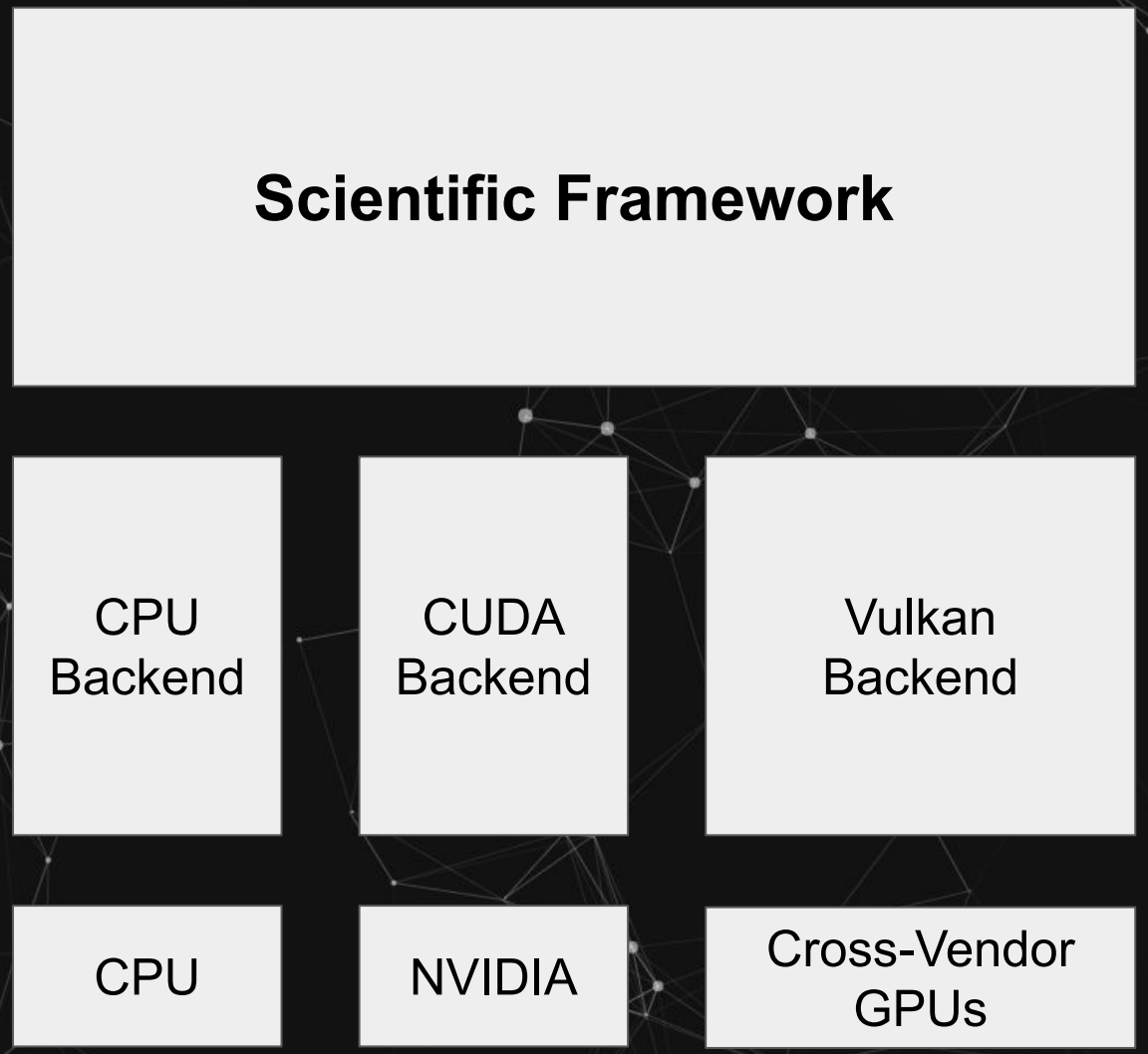
Implementations of Vulkan

- Each repository has hundreds/thousands of lines of replicated logic across
- Overhead of maintenance of custom and complex vulkan wrapper code
- Code replication leads to lack of standardisation in the underlying interaction with Vulkan interface
- Potential bugs can be introduced in replicated code with replicated efforts to fix them

Vulkan Location

Where does Vulkan sit in the stack?

- Vulkan SDK is being adopted as a backend for introducing cross-vendor GPU capabilities
- The Vulkan SDK is being adopted despite high barrier entry
- Vulkan SDK provides access to thousands of GPUs and specialised hardware.



Enter Kompute

The General Purpose Cross-Vendor GPU Computing Framework.

- **Dozens** instead of thousands of lines of code required
- **Augments** Vulkan interface instead of abstracting it
- **BYOV**: Bring-your-own-Vulkan design to play nice with existing Vulkan applications
- **Non-Vulkan name convention** to disambiguate components

Kompute Location

Where does Kompute sit in the stack?

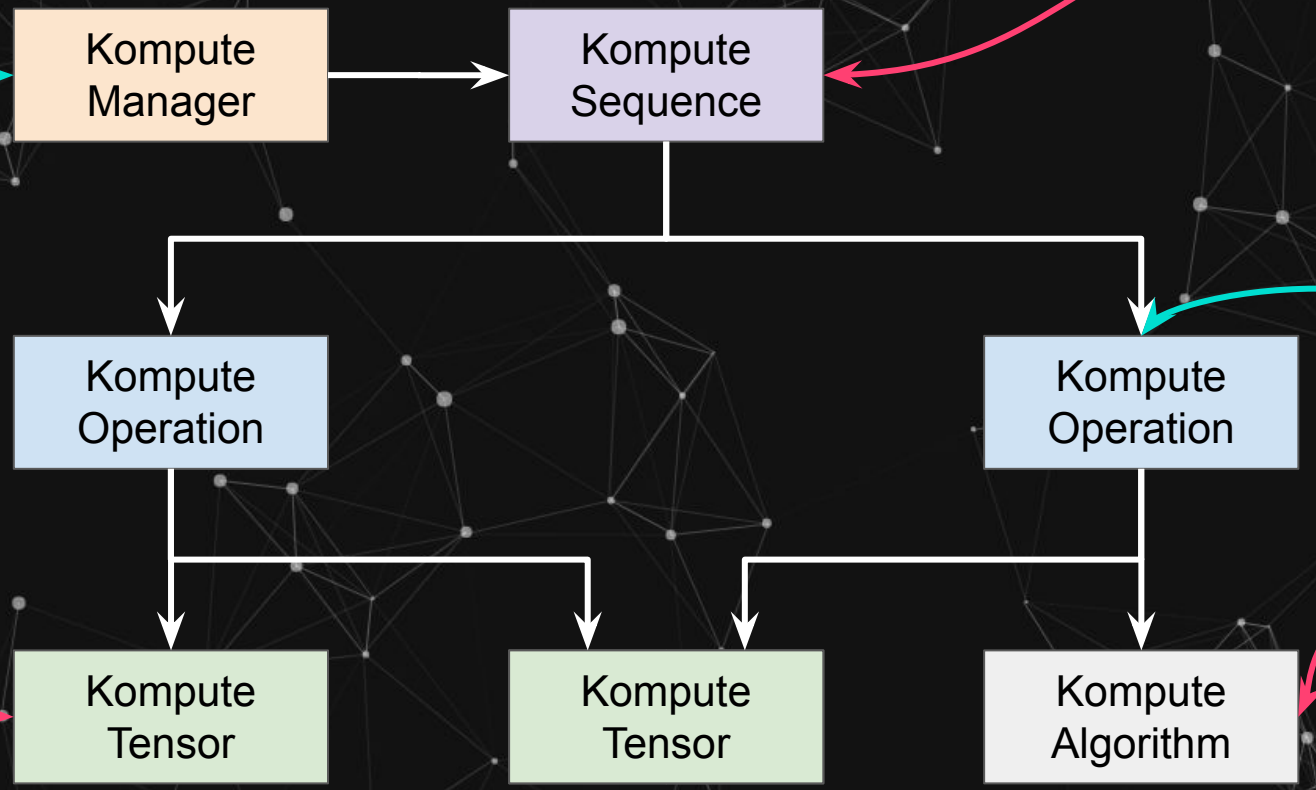
Scientific Framework

CPU Backend	CUDA Backend	Vulkan Backend with Kompute
CPU	NVIDIA	Cross-Vendor GPUs

High Level Overview of Features

- C++ Interface with Python Bindings
- Extensible operation-based architecture
- Robust testing with 90% code cov
- Exposes low level GPU resources
- Edge integration with Raspberry Pi
- Integration with Mobile Apps
- Integration with Game Engines

Vulkan Kompute: Architecture



Top level resource that manages Vk Device and Vk Queue

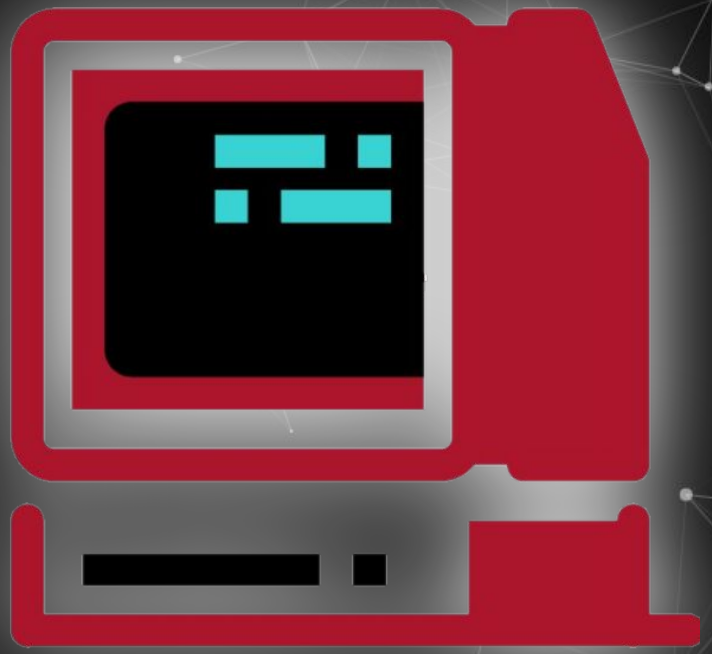
Manages & executes operations as batch in GPU as record commands and queue submits

Core data unit component to transfer and process via GPU memory and buffers

Extensible operation that performs instructions in GPU with tensor and optional shader

Abstracts the compute pipeline, shader modules and descriptor set bindings for processing

Kompute Mission



- Further the GPGPU ecosystem for scientific and industry applications through cross-vendor graphics card tooling and capabilities (across AMD, Qualcomm, NVIDIA & Friends)
- Foster ecosystem of parallel & distributed frameworks, standards and applications that enables for efficient and robust GPGPU
- Standardisation of underlying cross-vendor GPGPU computing across advanced data processing frameworks
- Enable high performance processing for GPU accelerated mobile and edge processing use-cases

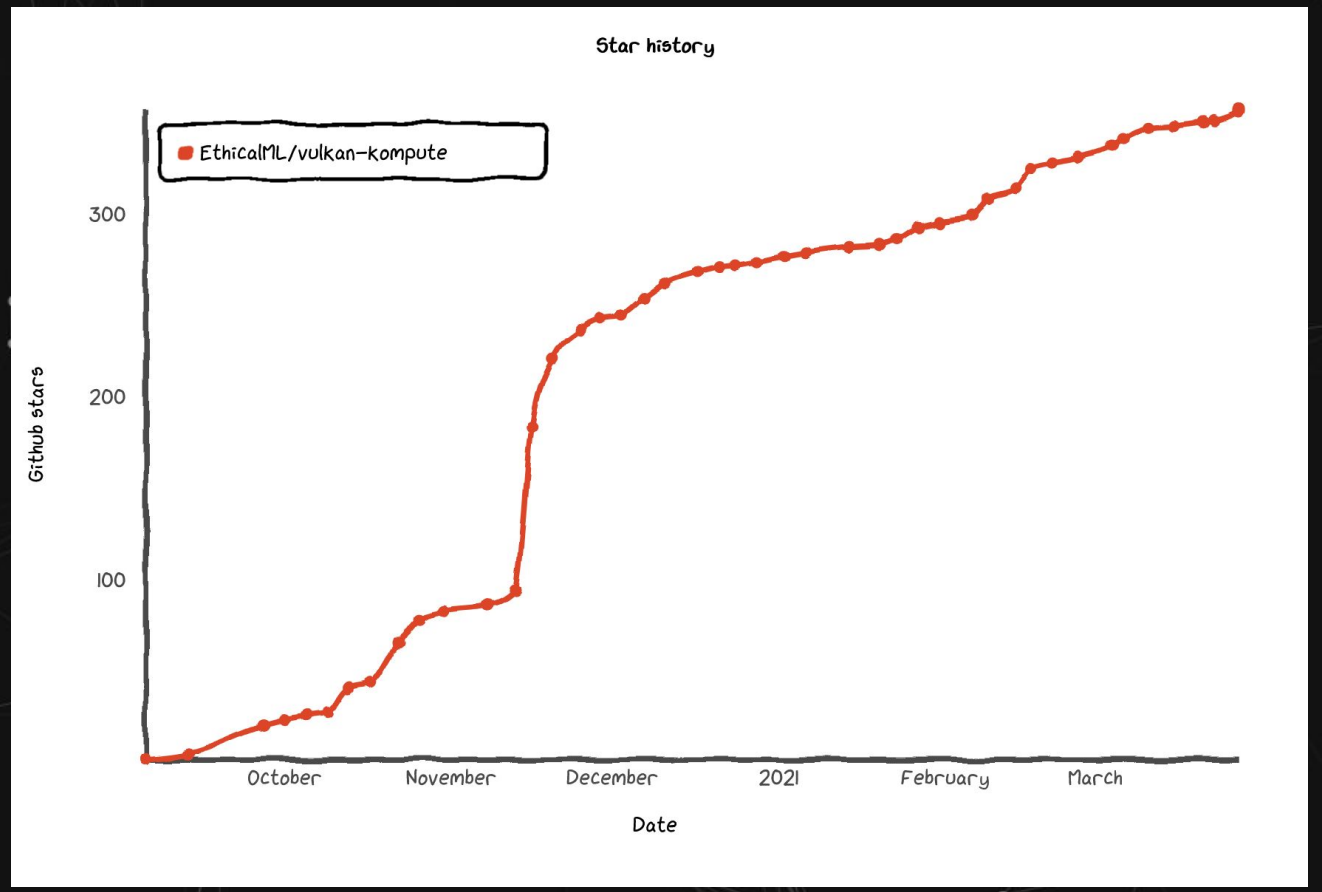
Kompute Reach (1/3)

4 Main Contributors, 9 Total

Beyond CUDA: GPU Accelerated Python for Machine Learning on Cross-Vendor Graphics Cards Made Simple

A practical deep dive into GPU Accelerated Python on cross-vendor graphics cards (AMD, Qualcomm, NVIDIA & friends) building machine learning algorithms using the Vulkan Kompute Python Framework

Articles have 30k+ Views



Organic Popularity Growth

Kompute Reach (2/3)

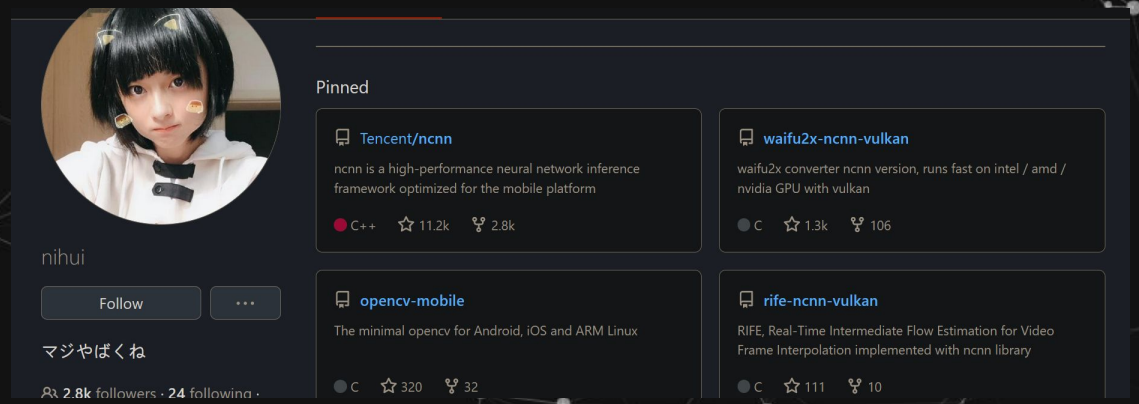
vkJAX

JAX interpreter based on Vulkan Kompute

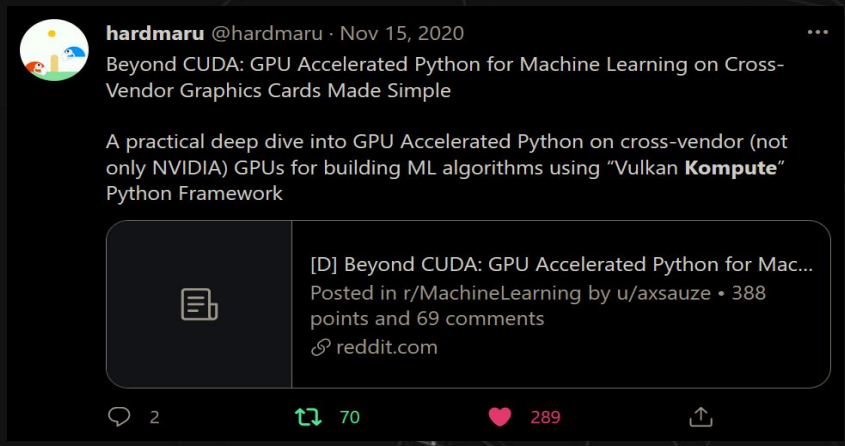
Minimal Example

```
import numpy as np, jax.numpy as jnp
import vkjax

def jax_fun(x,W,b):
  return jnp.dot(x, W) + b

vkfun = vkjax.wrap(jax_fun)

#this runs on the GPU, powered by vulkan
y = vkfun(
    np.random.random([8,128]),
    np.random.random([128,16]),
    np.random.random([16])
)
```

Backend for ML Libraries



Tencent NCNN Author Recognition



Social Media Traction

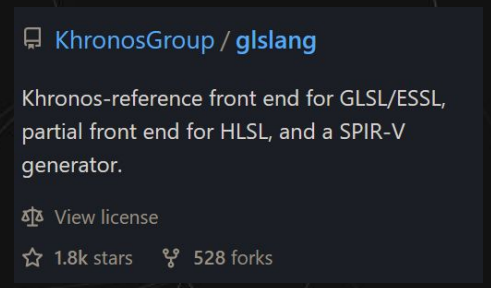
Kompute Reach (3/3)

Orgs. Hiring for Kompute Skill

KHRONOS GROUP android

UNREAL ENGINE GODOT Game engine

Broad Community Collaboration



Upstream Contributions

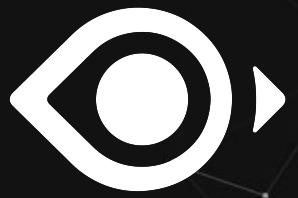
Potential of Kompute at LFAI

- Become the backend of large projects looking to integrate with Vulkan or add mobile capabilities
- Provide interoperability for applications to introduce GPGPU through higher level C++, Python, or other
- Enable edge processing capabilities through current capabilities in Android, IOS, Raspberry Pi, etc
- Build ecosystem of higher level tools for specialised functionality
- Serve as an anchor to the Khronos Group as LFAI has core DL/ML application

LF Project Collaborations

ONNX

Angel



Milvus

KubeEdge

OpenYurt

kubernetes

mlflow™

Pyro

DELTA LAKE™

LF Project Collaborations Detail

Project Name	Potential Collaboration Description
ADLIK	GPU for cross vendor graphics cards as well as edge and mobile integrations
ONNX	Extensions to the server for cross vendor GPU support
Acumos AI	Marketplace entry for deploying Kompute acceletated applications
Angel	Cross vendor GPU support (using a JNI interface extended from Android support)
Milvus	GPU for cross vendor graphics cards as well as edge and mobile integrations
KubeEdge	Integration for GPU support on edge device and examples
OpenYurt	Integration for GPU support on edge device and examples
Kubernetes	General initiative to add support for cross-vendor GPU integration

LF Project Collaborations Detail

Project Name	Potential Collaboration Description
MLFlow	Runtime / backend for exporting / deploying kompute accelerated algorithms
Pyro	Adding backend for GPU acceleration in cross-vendor GPUs
Deltalake	Cross vendor GPU Acceleration for data processing
RISC-V	Contributing from the compute perspective to the ongoing collaboration [link] towards fully open source GPU (or display adapter) to the hardware level
AI Fairness 360	Backend integration for GPU accelerated algorithms in cross-vendor GPUs
AI Explainability 360	Backend integration for GPU accelerated algorithms in cross-vendor GPUs
AI Adv Robustness	Backend integration for GPU accelerated algorithms in cross-vendor GPUs
Horovod	Example deploying Kompute powered application on non-standard GPU device

a@ethical.institute

Vulkan Kompute

Linux Foundation Proposal for Vulkan Kompute as Sandbox project focused on advancement of cross-vendor GPGPU

Alejandro Saucedo

@AxSaucedo

ANNEX

Enter Vulkan Kompute (Simple Python Example)

```python
# Create Kompute Manager
mgr = Manager()
sq = mgr.sequence()
```

```python
# Initialize tensors with List[] or np.Array
t_a = mgr.tensor([2, 2, 2])
t_b = mgr.tensor([1, 2, 3])
t_out = mgr.tensor([0, 0, 0])

sq.eval(kp.OpTensorSyncDevice([t_a, t_b, t_out]))
```

```python
sq.eval(kp.OpAlgoDispatch(
    [t_a, t_b, t_out],
    compute_shader_multiply.to_spirv()))
```

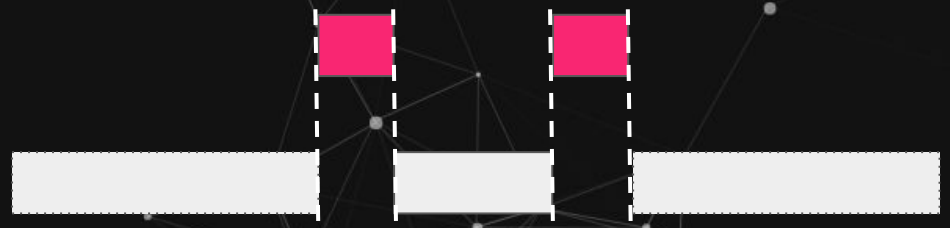
```python
# Sync to CPU host memory
sq.eval(kp.OpTensorSyncLocal(
        [t_out]))
```

```python
# Prints [2.0, 4.0, 6.0]
print(t_out.data())
```

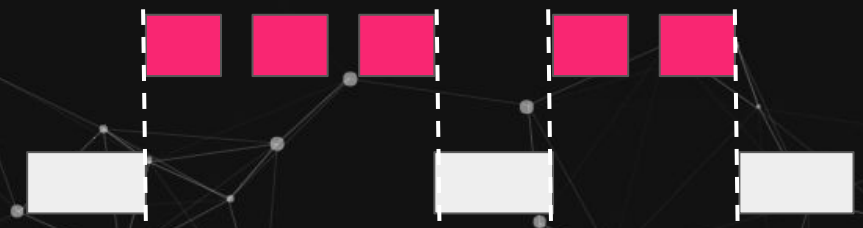
```python
# Define the function via PyShader as glsl string / spirv bytes
@python2shader
def compute_shader_multiply(index=("input", "GlobalInvocationId",
ivec3),
                            buffer_a=("buffer", 0, Array(f32)),
                            buffer_b=("buffer", 1, Array(f32)),
                            buffer_out=("buffer", 2, Array(f32))):
    i = index.x
    buffer_out[i] = buffer_a[i] * buffer_b[i]
```

Deeper Optimizations

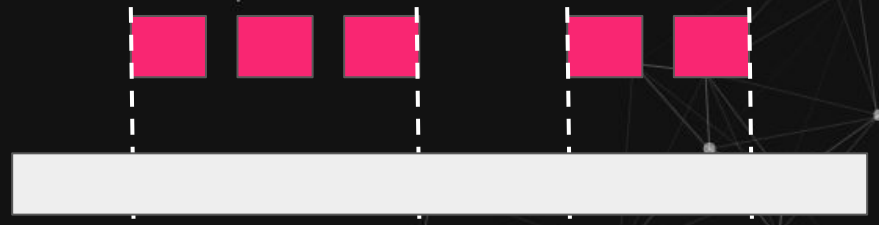
CPU GPU



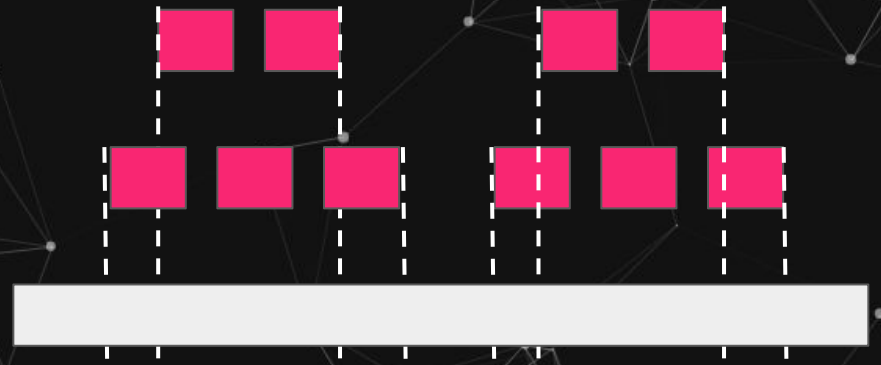
Run a single command/operation in a sequence with manager



Reuse multiple sequences in same Tensors with pre-recorded cmds



Asynchronous execution of Sequences



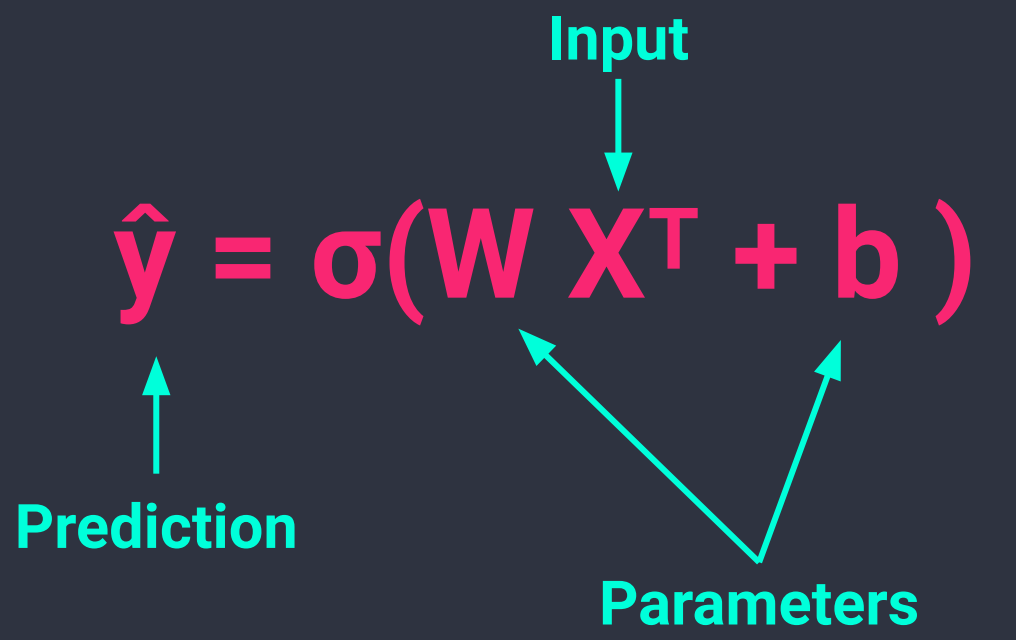
Concurrent execution of Sequences across GPU queues

Kompute ML Example

Kompute

1. Create Tensors
2. Initialise Sequence to init Tensors
3. Initialise Sequence to run Algorithm (with Shader)
4. Iterate 100 times to “learn” the parameter Tensors with learning rate
5. Print the learned parameter Tensors

Shader (GPU)



LR Shader Logic

```python
@ps.python2shader
def compute_shader(
        index   = ("input", "GlobalInvocationId", ps.ivec3),
        x_i   = ("buffer", 0, ps.Array(ps.f32)),
        x_j   = ("buffer", 1, ps.Array(ps.f32)),
        y     = ("buffer", 2, ps.Array(ps.f32)),
        w_in  = ("buffer", 3, ps.Array(ps.f32)),
        w_out_i = ("buffer", 4, ps.Array(ps.f32)),
        w_out_j = ("buffer", 5, ps.Array(ps.f32)),
        b_in  = ("buffer", 6, ps.Array(ps.f32)),
        b_out   = ("buffer", 7, ps.Array(ps.f32)),
        l_out   = ("buffer", 8, ps.Array(ps.f32)),
        M     = ("buffer", 9, ps.Array(ps.f32))):

    i = index.x # Fetch the current run index

    m = M[0]
```

```python
@ps.python2shader
def compute_shader(...):
        # ...
        w_curr = vec2(w_in[0], w_in[1])
        b_curr = b_in[0]

        x_curr = vec2(x_i[i], x_j[i])
        y_curr = y[i]

        z_dot = w_curr @ x_curr
        z = z_dot + b_curr
        y_hat = 1.0 / (1.0 + exp(-z))

        d_z = y_hat - y_curr
        d_w = (1.0 / m) * x_curr * d_z
        d_b = (1.0 / m) * d_z

        loss = -((y_curr * log(y_hat)) +
((1.0 + y_curr) * log(1.0 - y_hat)))

        w_out_i[i] = d_w.x
        w_out_j[i] = d_w.y
        b_out[i] = d_b
        l_out[i] = loss
```

Kompute Logic: Create Tensors

```
# First we create input and ouput tensors for shader
tensor_x_i = kp.Tensor([0.0, 1.0, 1.0, 1.0, 1.0])
tensor_x_j = kp.Tensor([0.0, 0.0, 0.0, 1.0, 1.0])

tensor_y = kp.Tensor([0.0, 0.0, 0.0, 1.0, 1.0])

tensor_w_in = kp.Tensor([0.001, 0.001])
tensor_w_out_i = kp.Tensor([0.0, 0.0, 0.0, 0.0, 0.0])
tensor_w_out_j = kp.Tensor([0.0, 0.0, 0.0, 0.0, 0.0])

tensor_b_in = kp.Tensor([0.0])
tensor_b_out = kp.Tensor([0.0, 0.0, 0.0, 0.0, 0.0])

tensor_l_out = kp.Tensor([0.0, 0.0, 0.0, 0.0, 0.0])

tensor_m = kp.Tensor([ tensor_y.size() ])

# We store them in an array for easier interaction
params = [tensor_x_i, tensor_x_j, tensor_y, tensor_w_in, tensor_w_out_i,
    tensor_w_out_j, tensor_b_in, tensor_b_out, tensor_l_out, tensor_m]
```

Tensors will be used as buffers in the GPU shader code created earlier

Kompute Logic: Init Tensors

```
mgr = kp.Manager(0)
```

```
mgr.eval_tensor_create_def(params)
```

Create manager with Device 0 and initialise all the Tensors

Kompute Logic: Main Sequence

Create sequence explicitly by recording multiple batch commands efficiently

```
# Create a managed sequence
sq = mgr.create_sequence()

# Clear previous operations and begin recording for new operations
sq.begin()

# Record operation to sync memory from local to GPU memory
sq.record_tensor_sync_device([tensor_w_in, tensor_b_in])

# Record operation to execute GPU shader against all our parameters
sq.record_algo_data(params, compute_shader.to_spirv())

# Record operation to sync memory from GPU to local memory
sq.record_tensor_sync_local(
    [tensor_w_out_i, tensor_w_out_j, tensor_b_out, tensor_l_out])

# Stop recording operations
sq.end()
```

Kompute Logic: “Learn” LR Params

```
ITERATIONS = 100
learning_rate = 0.1

# Perform ML training and inference across all input X and Y
for i_iter in range(ITERATIONS):

    # Execute an iteration of the algorithm
    sq.eval()

    # Calculate the parameters based on the respective derivatives
    for j_iter in range(tensor_b_out.size()):
    tensor_w_in[0] -= learning_rate * tensor_w_out_i.data()[j_iter]
    tensor_w_in[1] -= learning_rate * tensor_w_out_j.data()[j_iter]
    tensor_b_in[0] -= learning_rate * tensor_b_out.data()[j_iter]
```

Iterate 100 times updating the learned parameters using the learning rate of 0.1

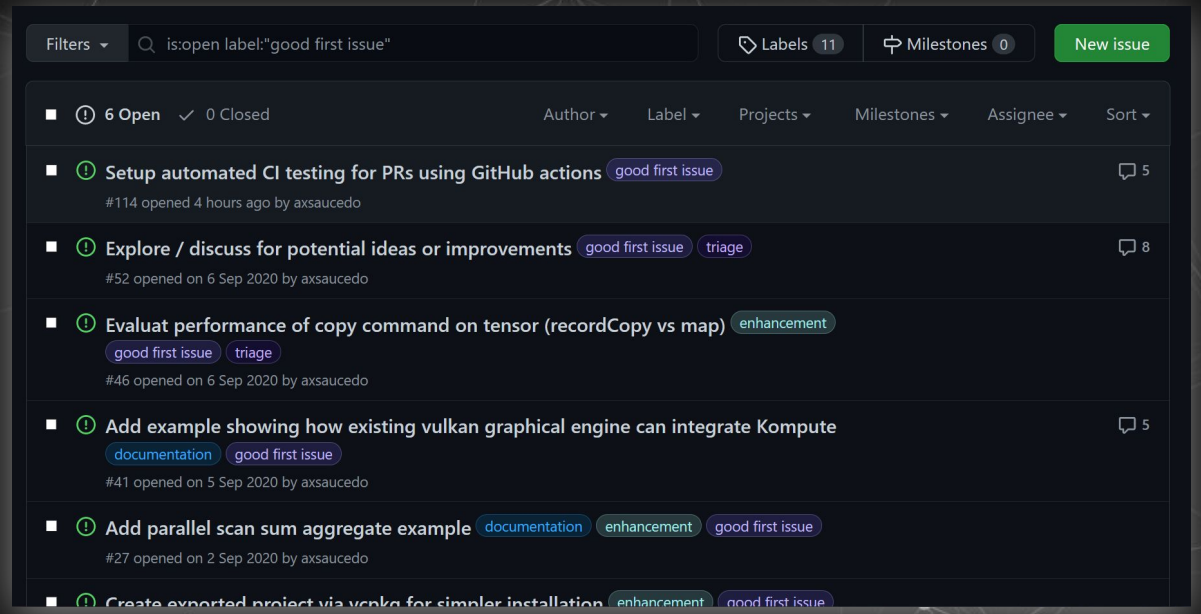
Kompute Logic: Print LR Params

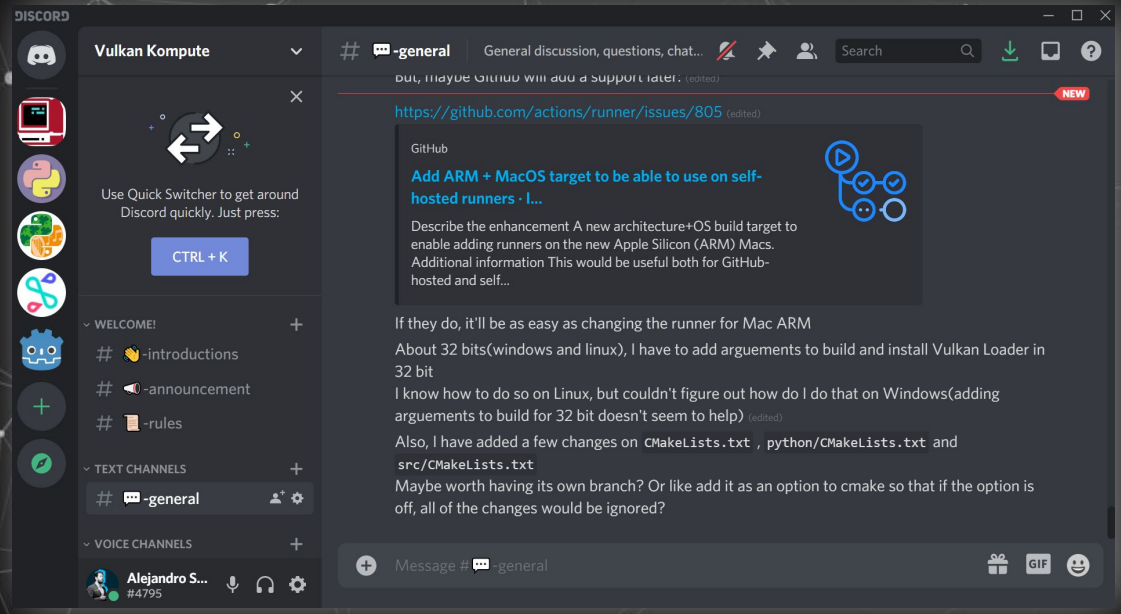
Finally print the learned parameters which represent our “trained model” and can be used to predict unseen datapoints

```
# Prints ~< 0.01
print(tensor_w_in.data()[0])

# Prints ~> 1.5
print(tensor_w_in.data()[1])

# Prints ~< 0.7
print(tensor_b_in.data()[0])
```

Get Involved!

github.com/EthicalML/vulkan-kompute





Pick up one of the good-first-issues

Join the discord chat

High level Roadmap

Integrate as backend of ML / scientific-computing frameworks

Create more default kp::Operations to have out of the box commands

---

Examples running Kompute across other platforms and frameworks

a@ethical.institute

Vulkan Kompute

Linux Foundation Proposal for Vulkan Kompute as Sandbox project focused on advancement of cross-vendor GPGPU

Alejandro Saucedo

@AxSaucedo

CI Working Group - Project CI usage analysis

LF AI & DATA

Background

- John surveyed all projects to get a sense of what CI systems and methodologies were in place. Questions asked:
  - Do you have a CI/CD environment?
  - Describe the environment and technologies used
  - What works well with this environment?
  - What things would you like to improve on or add to your environment?
- Responses came from 11 projects:
  - Angel, DataPractices.org, Datashim, Flyte, Ludwig, Marquez, Milvus, NNStreamer, OpenDS4All, Pyro, RosaeNLG

Key findings

- All code projects have a CI/CD environment
  - Datapractices and OpenDS4All do not
- GitHub Workflows and TravisCI most popular solutions used
  - Some usage of CircleCI, homegrown solutions.
  - Some on TravisCI moving to GitHub
  - Docker, SonarCloud, Kubernetes, readthedocs also being used
- Challenges
  - Limitations of free tier resources; need for cluster environments and different types of deployment architectures
  - Common asset management tooling (Dockerhub, etc)

Raw data at https://docs.google.com/spreadsheets/d/e/2PACX-1vSgNRr4WhH5rTIQxfcd3mhTIqaxdn6a1nqfg3rJcH-cLk8rzDq1_bAVAsiaB8i__pkxZbAay-5fPszb/pubhtml

Next steps (suggested)

- Collaborate on “CI Best Practices” document outlining recommendations and resources for projects.
  - API Documentation hosting
  - Test case result hosting
- Look at funding specialized infrastructure (clusters, GPUs, different architectures).
- Other ideas?

LF AI & Data - General Updates

LF AI & DATA

Linux Foundation AI & Data Landscape

2021-04-06T06:33:16Z 27aa395

See the interactive landscape at l.lfaidata.foundation

Greyed logos are not open source

Machine Learning

Framework: Accord.NET, FedAI, Microsoft LightGBM, MAHOUT, ML.NET, RAY

Platform: Angel (Graduated), ForestFlow (Incubating), Alibaba Alink, Apache SystemML, METAFLOW

Library: CatBoost, mlpack, Torch, OpenNN, scikit-learn, 将軍 Shogun, Sonnet, Salesforce Transmogrif AI, dmlc XGBoost, xLearn

Deep Learning

Framework: SINGA, Chainer, Microsoft CNTK, dy/net, Alibaba euler, [M]³ MindSpore, mxnet, ncnn, 飞桨, Pythia, PYTORCH, TensorFlow

Platform: DeepDetect, Determined AI, jina, Cnvrg.io, Polyaxon, Alibaba XDeepLearning

Library: BigDL, Catalyst, DL4J, Keras, PyTorch Ignite, PyTorch Lightning

Tool: BoTorch, Intel Distiller, plaidML, PyTorch Glow, tvm

Reinforcement Learning

Reinforcement Learning: Coach, DeepMind Lab, Dopamine, Horizon, OpenAI, Google PlaNet, Google SEED RL, Stable Baselines

Programming

Programming: Pyro (Graduated), DASK, Intel nGraph, julia, MARS, NumPy, ÑYOKA, pemegranate, PyMC3, SciPy, Stan

Notebook Environment

Notebook Environment: Apache Zeppelin, BeakerX, colab, Elyra, IPython, jupyter, IBM, Polynote, Stencila, Streamlit

Data

Store & Format: JanusGraph (Incubating), Milvus (Incubating), ALLUXIO, ICEBERG, ORC, ARES DB, ARROW, DELTA LAKE, druid, hudi, Parquet, pilosa, VEARCH, vespa

Versioning: data science pipeline, DVC, Pachyderm, Quilt

Operations: Amundsen (Incubating), datashim (Incubating), MARQUEZ (Incubating), HIVE, ckan, Data Hub, snorkel, WHYLABS whylogs

Feature Engineering: FEAST (Incubating), tsfresh

Stream Processing: NNStreamer (Incubating), RocketMQ, beam, brooklin, Flink, kafka, Pravega, PREFECT, PULSAR, samza, Uber uReplicator

SQL Engine: APACHE DRILL, HAWQ, presto, SQLFlow

Visualization: bokeh, IBM, D3, plotly Dash, Uber deck.gl, Ecco, Google Facets, Grafana, RCloud, NYU, Superset, Google TensorBoard

Pipeline Management: Intel Analytics Zoo, DAGSTER, TPOT

Labeling & Annotation: Intel CVAT, Doccano, Label Studio, Labelbox, LabelImg, HITACHI Inspire the Next, Microsoft VoTT

Governance: EGERIA (Graduated), Community Data License Agreement

Model

Inference: ADLIK (Incubating), KFServing, MNN, NVIDIA TensorRT, NVIDIA, uTensor

Benchmarking: DAWNBench, MLPerf

Training: Horovod (Graduated), LUDWIG (Incubating), Microsoft Embedded Learning Library, Petastorm

Parameter: Hyperopt, talos

Format & Interface: ONNX (Graduated), Uber Neuropod

Marketplace: Acumos (Graduated), IBM Model Asset Exchange

Workflow: Flyte (Incubating), Airflow, nifi, argo, Azkaban, Cadence, Couler, CYCLONE, DataBolt dataflow, Spotify luigi, mleap, PREFECT, TRAINS, Volcano

Tool: Qualcomm AIMET, FACEBOOK dlrm, Microsoft MMdnn, aws MMS, MDB, amazon Neo-AI, NETRON, ONNX Runtime, PipelineAI, studio.ml, Google TensorFlow Model Analysis, aws TorchServe, turi

Trusted & Responsible AI

Explainability: AI Explainability 360 (Incubating), ALIBI, ELI5, Microsoft InterpretML, University of Washington Lime, Google Lucid, SHAP, SKATER, Bolt Trustinterpreter

Adversarial: Adversarial Robustness Toolbox (Incubating), AdvBox, advertorch, hans, Foolbox

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

Distributed Computing

Computing & Management: EDL Elastic Deep Learning (Incubating), SOAJS (Incubating), Bahir, MESOS, Apache Kyuubi, Spark, STORM, GNSS, NETFLIX genie, kubernetes, intel Nauta, Singularity

Interface: sparklyr (Incubating), APACHE TOREE, LIVY, Jupyter Enterprise Gateway

l.lfaidata.foundation

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

LF AI & DATA Landscape

LF AI & DATA

Security & Privacy

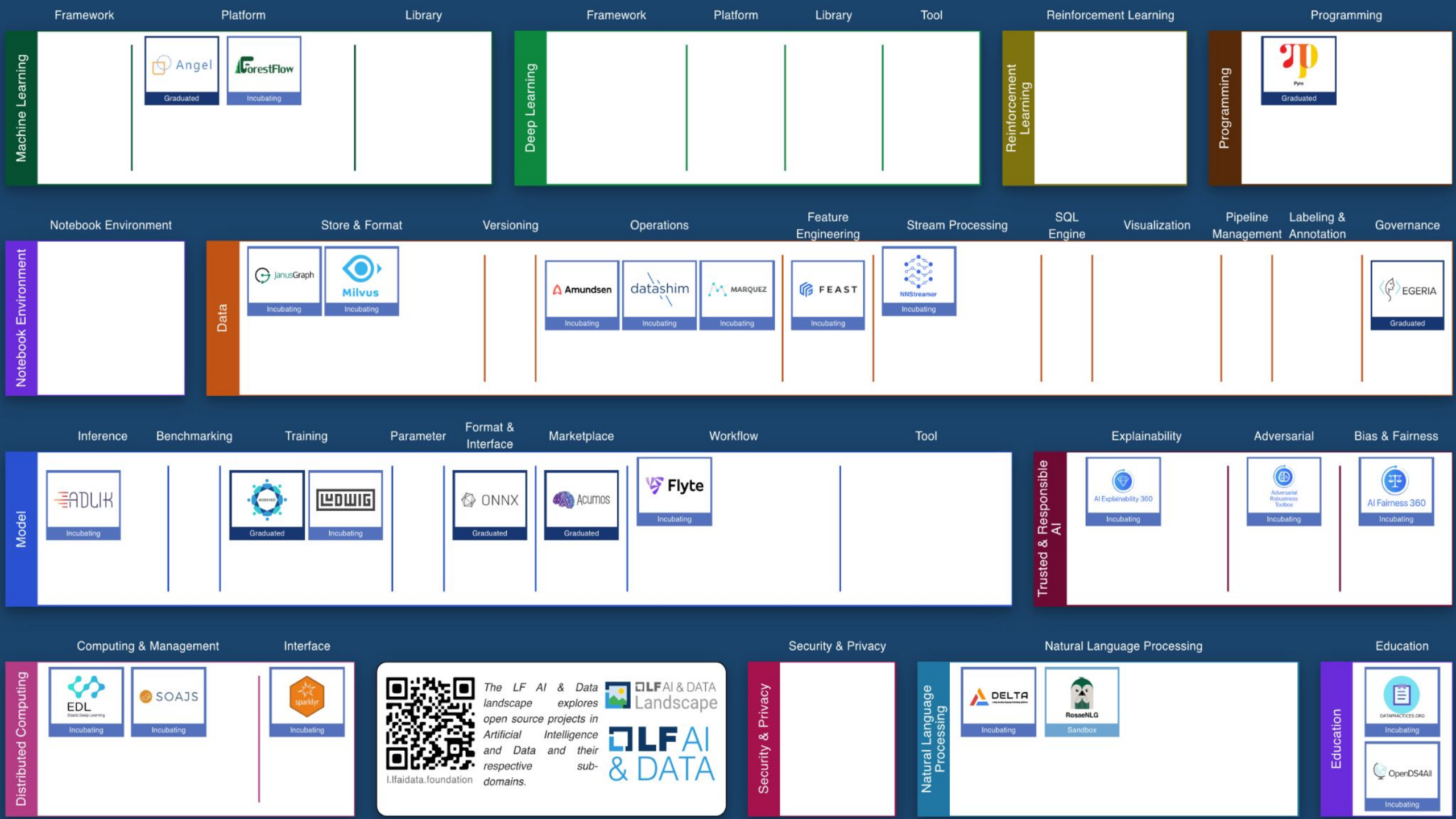
Security & Privacy: Google Differential Privacy, IBM HElib, Microsoft SEAL, SUBSTRA, Google TensorFlow Privacy, TFEncrypted

Natural Language Processing

Natural Language Processing: DELTA (Incubating), RosaeNLG (Sandbox), Google ALBERT, AllenNLP, Google Bert, CoreNLP, mozilla DeepSpeech, fastText, flair, LUON, Kashgari, FACEBOOK LASER, Facebook fairseq, cisco MindMeld, intel NLP Architect, NLP, ParlAI, FACEBOOK PyText, RASA, spaCy, Transformers, FACEBOOK XLM, YouTokenToMe

Education

Education: DATAPRACTICES.ORG (Incubating), OpenDS4All (Incubating)


Framework
Platform
Library
Machine Learning
Angel
Graduated
ForestFlow
Incubating
Framework
Platform
Library
Tool
Deep Learning
Reinforcement Learning
Reinforcement Learning
Programming
Programming
Pyro
Graduated
Notebook Environment
Notebook Environment
Store & Format
Versioning
Operations
Feature Engineering
Stream Processing
SQL Engine
Visualization
Pipeline Management
Labeling & Annotation
Governance
Data
JanusGraph
Incubating
Milvus
Incubating
Amundsen
Incubating
datashim
Incubating
MARQUEZ
Incubating
FEAST
Incubating
NNStreamer
Incubating
EGERIA
Graduated
Inference
Benchmarking
Training
Parameter
Format & Interface
Marketplace
Workflow
Tool
Model
ADLIK
Incubating
Graduated
LUDWIG
Incubating
ONNX
Graduated
Acumos
Graduated
Flyte
Incubating
Explainability
Adversarial
Bias & Fairness
Trusted & Responsible AI
AI Explainability 360
Incubating
Adversarial Robustness Toolbox
Incubating
AI Fairness 360
Incubating
Computing & Management
Interface
Distributed Computing
EDL
Incubating
SOAJS
Incubating
sparklyr
Incubating
l.lfaidata.foundation
The LF AI & Data landscape explores open source projects in Artificial Intelligence and Data and their respective sub-domains.
LF AI & DATA Landscape
LF AI & DATA
Security & Privacy
Security & Privacy
Natural Language Processing
Natural Language Processing
DELTA
Incubating
RosaeNLG
Sandbox
Education
Education
DATAPRACTICES.ORG
Incubating
OpenDS4All
Incubating

2020 TAC Meetings Summary

Months

Jan Feb Mar	16: Milvus (Zilliz)*	*13: MLOps Work (LF CD)* *27: Collective Knowledge (Coral Reef)*	12: NNStreamer (Samsung)* 26: ForestFlow (?)*
Apr May Jun	*9: Trusted AI & ML Workflow (LF)* *23: Open Data Hub (Red Hat)*	7: Ludwig (Uber)* *21: SnapML (IBM)*	*4: Trusted AI (AI for Good, Ambianic.ai, MAIEI)* 18: Fairness, Explainability, Robustness (IBM)*
Jul Aug Sep	*16: Mindspore (Huawei)* 30: Amundsen (Lyft)*	*16: Delta (Didi)* **16: Horovod (Uber/LF)**** *30: ModelDB (?)* *30: Egeria, OpenDS4All, BI&AI (LF ODPi)*	10: SOAJS (HeronTech)* 10: Delta (Didi)* 24: FEAST (Gojek)* **24: Egeria, (LF ODPi)**** 24: OpenDS4All (ODPi)* 24: BI&AI Committee (ODPi)
Oct Nov Dec	*8: Fairness, Explainability, Robustness (LF)* *22: OpenLineage (DataKins)* *22: IDA (IBM/Salesforce)*	5: DataPractices.Org (WorldData/LF)* *5: Kubeflow-On-Prem (Google,Arrikto/Intel)* *19: OpenDS4All, DataPractices.Org, edX Ethical AI (LF)*	3: TBD - JanusGraph (LF)* *3: TBD - RosaeGL (?)* 17: TBD – Seldon Core (Seldon)* **17: TBD – Pyro (Uber/LF)****

(Entity)* = incubating vote

** **bold = graduate vote**

Italics = invited project presentation

2021 TAC Meetings Pipeline Summary

Months

Jan Feb Mar	14: Datashim(IBM)* 28: Project Lifecycle Stages vote Invited talks Sedna & CIM	11: Invited talks Egeria CI & Mentorships 25: Flyte (Lyft) *	11: RosaeNLG () Sandbox Proposal Invited talk Elyra-AI (IBM) 25: Substra Framework (Substra)*
Apr May Jun	8: Invited talk JINA AI 22: Egeria & OpenDS4All - project update	6: ML eXchange (MLX) (IBM) Sandbox Vulcan Kompute () Sandbox 20: OpenLineage (Datakin) Sandbox	3: KOSA.ai 17: GSI
Jul Aug Sep	1: Canceled for holiday 15: TonY (Linkedin)	5: TBD - Project updates 19: TBD - Project updates	?: Open Data Hub (Red Hat) ? Ray (Anyscale.io) ?: Pachyderm (Pachyderm) ?: DataHub (LinkedIn) ?: Kubeflow-On-Prem (Google, Arrikto, Intel)
Oct Nov Dec	?: Vespa (Verizon Media) ?: KubeflowServing (Google, Arrikto, Seldon) ?: Kubeflow Pipeline (Google, Bloomberg) ?: Common Knowledge (Code Reef) ?: Couler (Ant Financial)	?: Snorkle (Snorkle) ?: Plotly (DASH) ?: Mellody (Substra) ?: mloperator (Polyaxen) ?: SnapML (IBM)	?: PMML/PFA (DMG.org) ?: Mindspore, Volcano (Huawei) ?: TransmorgrifAI (Salesforce) ?: AIMET (Qualcomm) ?: Elyra-AI (IBM)

(Entity)* = incubating vote

** **bold = graduate vote**

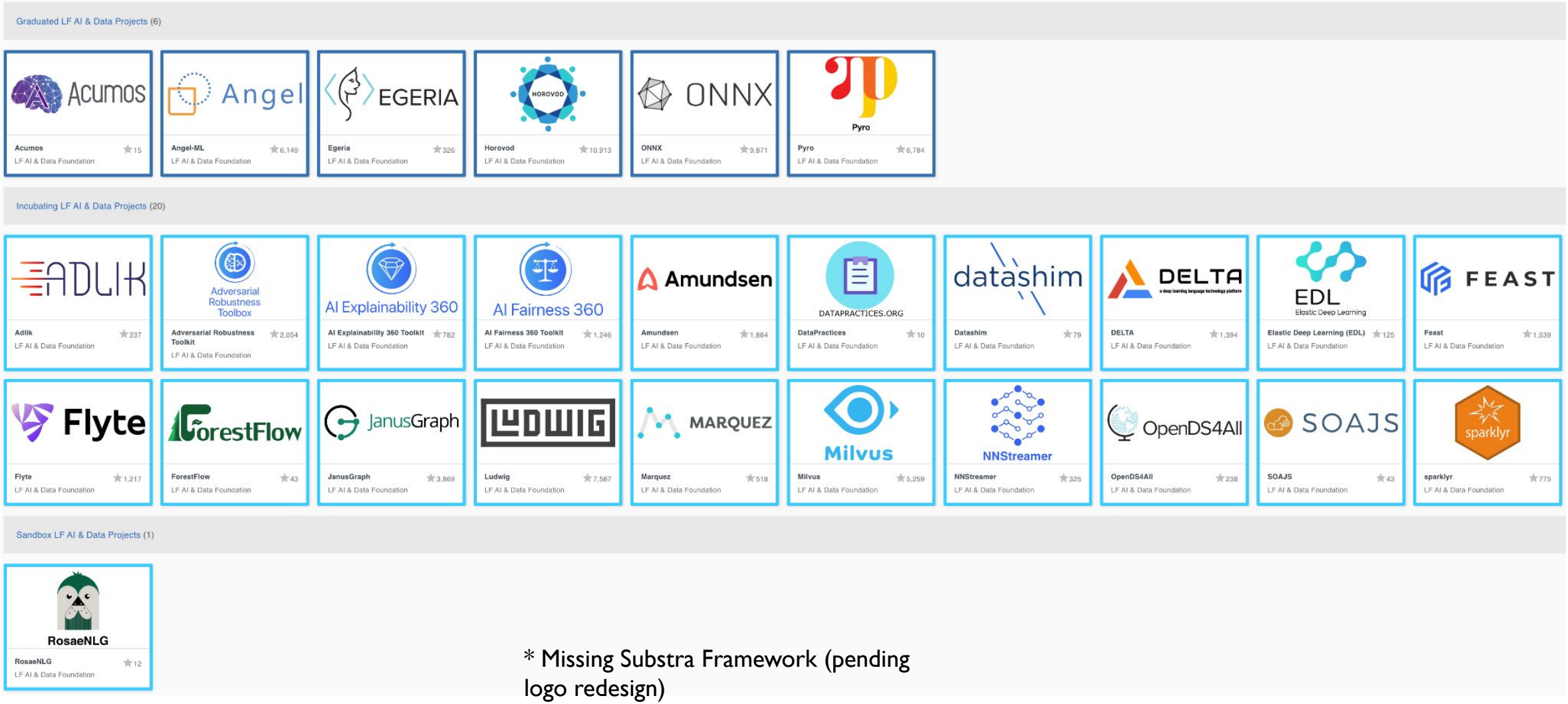
Italics = invited project presentation

Getting to know the projects more

LF AI & DATA

Projects (28)

https://landscape.lfai.foundation/card-mode?project=company



* Missing Substra Framework (pending logo redesign)

New projects in 2021

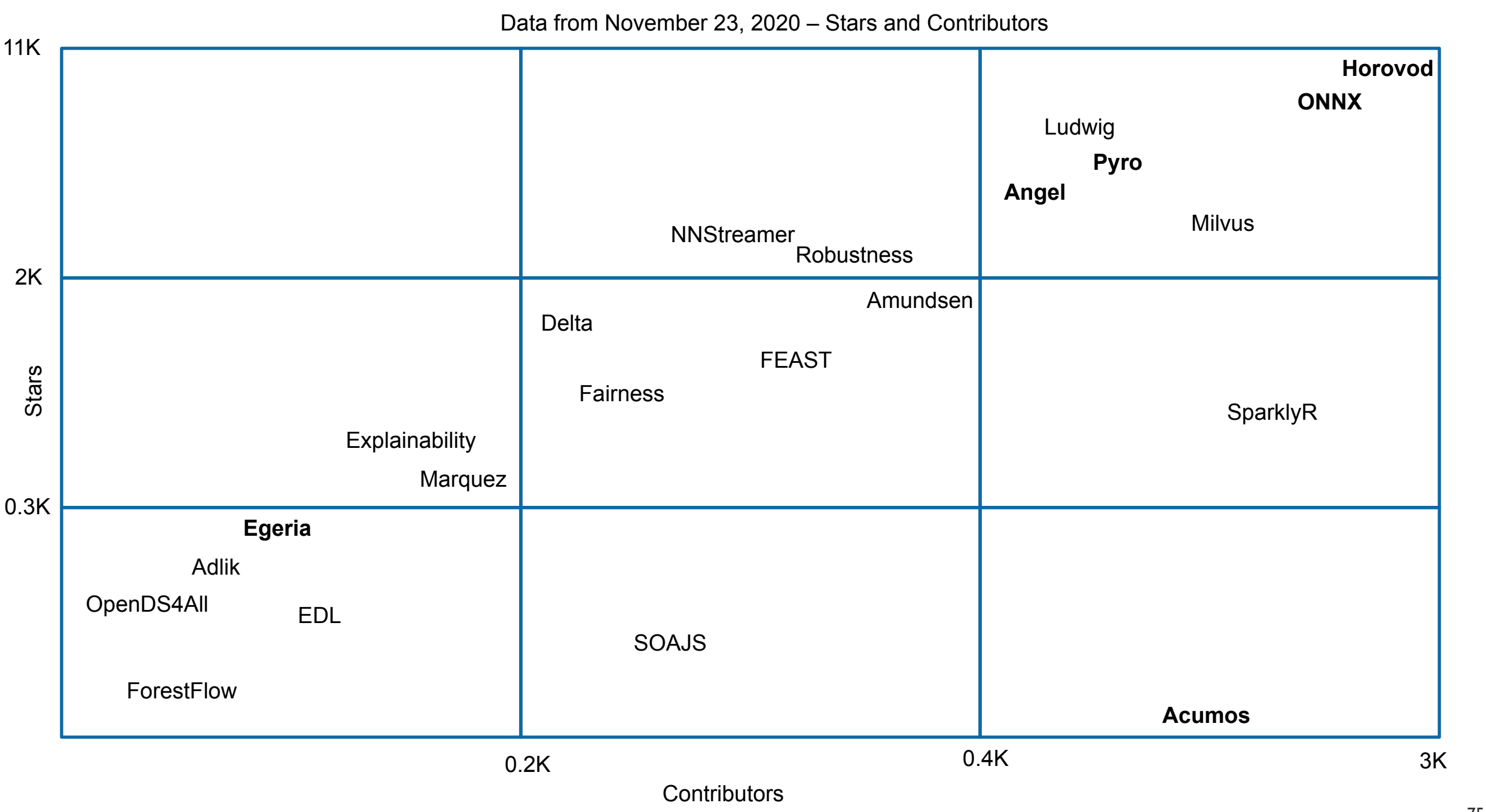
1. **Datashim:** Open source enablement and acceleration of data access for Kubernetes/Openshift workloads in a transparent and declarative way
2. **Flyte:** Production-grade, declarative, structured and highly scalable cloud-native workflow orchestration platform
3. **RosaeNLG:** Open source project, template-based Natural Language Generation (NLG) automating the production of relatively repetitive texts based on structured input data and textual templates, run by a NLG engine
4. **Substra Framework:** Low-layer framework, offering secure, traceable, distributed orchestration of machine learning tasks among partners.

→ Track incoming proposals via: https://github.com/lfai/proposing-projects

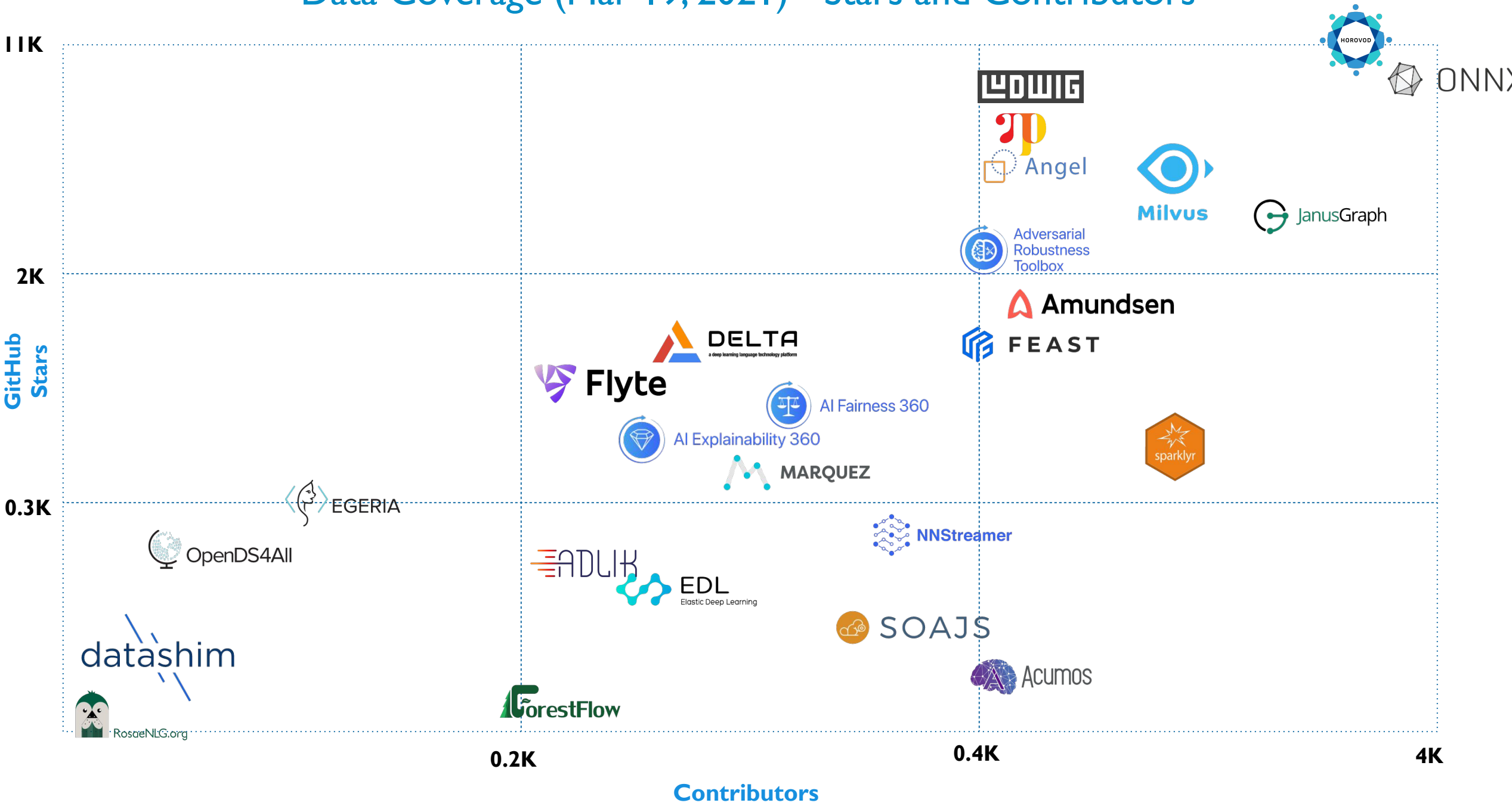
Active and growing developer community

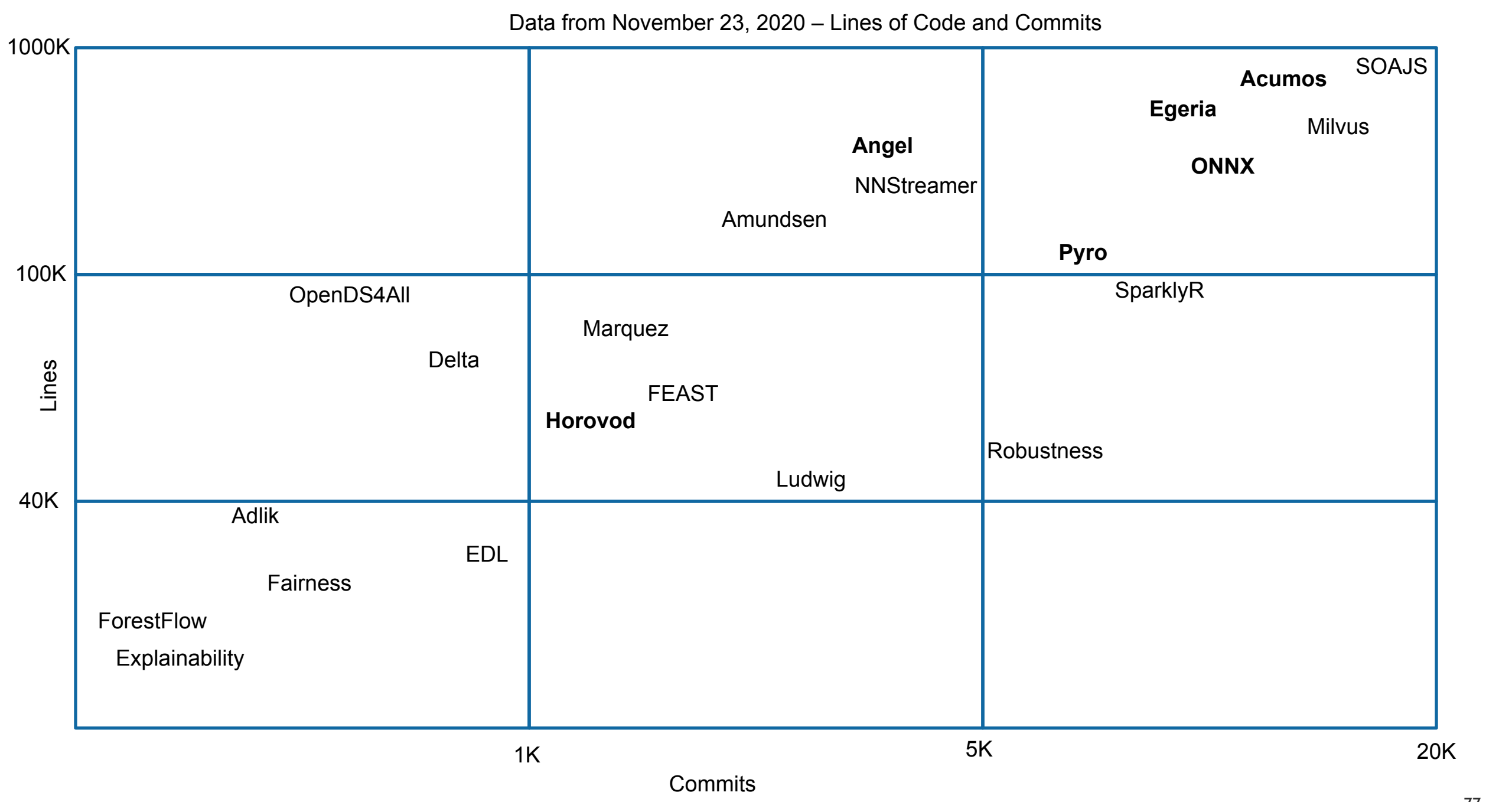
Cumm. Jan 1- Dec 31, 2020 vs. Jan 1, 2020 to Apr 2, 2021

8.92K Contributors	**34.65K** PRs/Changesets	**100.03K** Commits	**22.54K** Total issues	**358** Repositories	**2.08K** Slack messages
+7.25%	+7.21%	+11.03%	+5.19%	+3.07%	+68.75%
9.61K Contributors	**37.15K** PRs/Changesets	**111.07K** Commits	**23.71K** Total issues	**369** Repositories	**3.51K** Slack messages


Data from November 23, 2020 – Stars and Contributors
11K
2K
0.3K
Stars
0.2K
0.4K
3K
Contributors
Horovod
ONNX
Ludwig
Pyro
Angel
Milvus
NNStreamer
Robustness
Amundsen
Delta
FEAST
Fairness
SparklyR
Explainability
Marquez
Egeria
Adlik
OpenDS4All
EDL
SOAJS
ForestFlow
Acumos

Data Coverage (Mar 19, 2021) - Stars and Contributors




Data from November 23, 2020 – Lines of Code and Commits
1000K
100K
40K
Lines
SOAJS
Acumos
Egeria
Milvus
Angel
ONNX
NNStreamer
Amundsen
Pyro
SparklyR
OpenDS4All
Marquez
Delta
FEAST
Horovod
Robustness
Ludwig
Adlik
EDL
Fairness
ForestFlow
Explainability
1K
5K
20K
Commits

Data Coverage (Mar 19, 2021) - Line of Code and Commits

Looking to host a project with LF AI & Data

- Hosted project stages and life cycle:
  https://lfaidata.foundation/project-stages-and-lifecycle/
- Offered services for hosted projects:
  https://lfaidata.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 2021 Events wiki for a list of all events
  - To participate visit the LF AI & Data 2021 Events wiki page or email info@lfaidata.foundation

- Please consider holding virtual events

To discuss participation, please email events@lfaidata.foundation

Upcoming Events

https://lfaidata.foundation/events/

- **Kubernetes AI Day (virtual) - May 4th - Event Website**
  a. **Booth**
- **OSS Global (hybrid) - Seattle, WA, USA - Sept 27-30 - Event Website**
  a. **Mini-Summit, Booth, Track**

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@lfaidata.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

- **Slack conversation channel:**
  #trusted-ai-committee
  https://lfaifoundation.slack.com/archives/CPS6Q1E8G
- **Github:**
  https://github.com/lfai/trusted-ai
- **Wiki:**
  https://wiki.lfai.foundation/display/DL/Trusted+AI+Committee
- **Email lists:**
  https://lists.lfaidata.foundation/g/trustedai-committee/
- **Next call:** Monthly alternating times
  https://wiki.lfai.foundation/pages/viewpage.action?pageId=12091895

ML Workflow & Interop

- **Leadership:**
  Huang “Howard” Zhipeng (Huawei)
- **Goal:**
  Define an ML Workflow and promote cross project integration

- **Slack conversation channel:**
  #ml-workflow
  https://lfaifoundation.slack.com/archives/C011V9VSMQR
- **Wiki:**
  https://wiki.lfaidata.foundation/pages/viewpage.action?pageId=10518537
- **Email lists:**
  https://lists.lfaidata.foundation/g/mlworkflow-committee
- **Next call:** Monthly check calendar/slack
  https://wiki.lfai.foundation/pages/viewpage.action?pageId=18481242

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 business intelligence and better strategic direction for your organization.

- **Slack conversations channel:**
  #bi-ai-committee
  https://lfaifoundation.slack.com/archives/C01EK5ND073
- **Github:**
  https://github.com/odpi/bi-ai
  Wiki:
  https://wiki.lfaidata.foundation/pages/viewpage.action?pageId=35160417
  Email lists:
  https://lists.lfaidata.foundation/g/biai-discussion
  Next call: Monthly community call TBD

Ongoing 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.lfaidata.foundation/g/aiethics-training**

Upcoming TAC Meetings

LF AI & DATA

Upcoming TAC Meetings (Tentative)

- May 20: Julian Le Dem & Mandy Chessell “Open Lineage” sandbox proposal
- Jun 3: KOSA.ai
- June 17: GSI invited presentation
- July 1: Canceled Holiday
- July 15: TonY (LinkedIn)
- Aug 5: TBD - Annual project review
- Aug 19: TBD - Annual project review
-

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

TAC Meeting Details

- To subscribe to the TAC Group Calendar, visit the wiki: https://wiki.lfaidata.foundation/x/cQB2
- 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