# Meeting of the Technical Advisory Council (TAC)

December 3, 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: lfaidata.foundation
- Wiki: wiki.lfaidata.foundation
- GitHub: github.com/lfaidata
- Landscape: landscape.lfai.foundation or l.lfai.foundation
- Mail Lists: https://lists.lfaidata.foundation
-
- LF AI Logos: https://github.com/lfaidata/artwork/tree/master/lfaidata
- LF AI Presentation Template: https://drive.google.com/file/d/1eiDNJvXCqSZHT4Zk_-czASIz2GTBRZk2/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 (2 mins)
- Project Contribution Proposal + Q&A + TAC Vote (35 minutes)
  - JanusGraph (Oleksandr Porunov)
- Invited Presentation (15 minutes)
  - RosaeNLG (Ludan Stoecklé)
- LF AI General Updates (3 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 |
| 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 |

# Approval of November 19th, 2020 Minutes

Draft minutes from the November 19th TAC call were previously distributed to the TAC members via the mailing list

**Proposed Resolution:**

- That the minutes of the November 19th meeting of the Technical Advisory Council of the LF AI & Data Foundation are hereby approved.

# Project Contribution Proposal - JanusGraph

LF AI & DATA

# Project Contribution Proposal Review & Discussion: JanusGraph

JanusGraph is a highly scalable graph database optimized for storing and querying large graphs with billions of vertices and edges distributed across a multi-machine cluster. JanusGraph is a transactional database that can support thousands of concurrent users, complex traversals, and analytic graph queries. JanusGraph is an open source, distributed graph database under The Linux Foundation. JanusGraph is available under the Apache License 2.0 license. The project is supported by IBM, Google, Hortonworks and Grakn Labs.

**Presenter:** Oleksandr Porunov

**Resources:**

Github: https://github.com/JanusGraph/janusgraph
Project Level: Incubation
Proposal: https://github.com/lfai/proposing-projects/blob/master/proposals/janusgraph.adoc



# JanusGraph
# LF AI & DATA Incubation Project Proposal

Oleksandr Porunov
Principal Software Engineer
JanusGraph Technical Steering Committee

December 3, 2020

# Presentation plan

1. Project overview
2. Architectural overview
3. Development roadmap
4. Community stats
5. Governance
6. Events
7. Production users
8. Collaboration opportunities
9. Benefits
10. Conclusion

# 1. Project overview

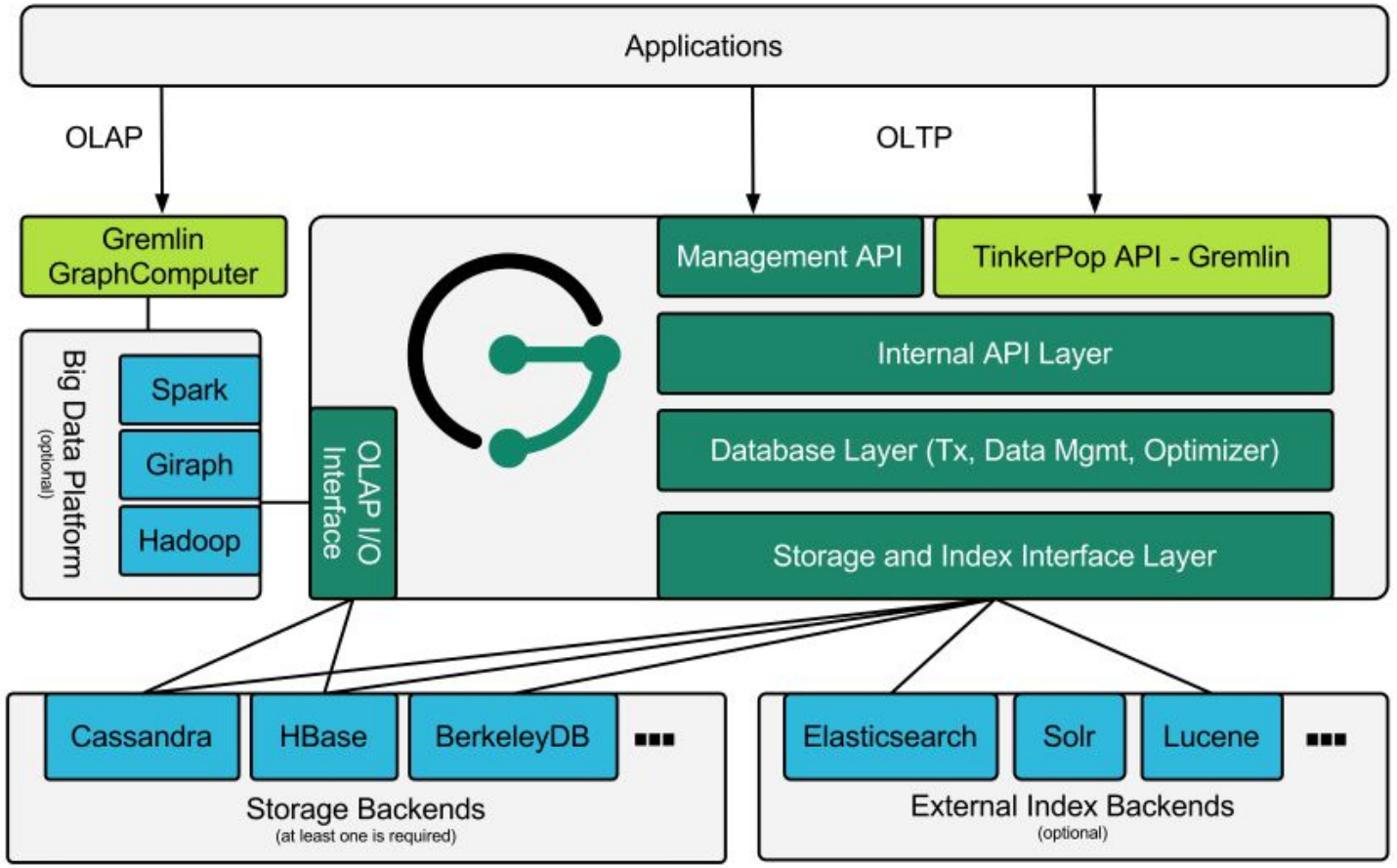
**JanusGraph** – distributed, open source, massively scalable graph database under The Linux Foundation.

Fork of TitanDB (being developed since 2012).
The initial release of JanusGraph was made in April 2017.

Current repositories:

- **janusgraph** – Main database repository. Includes Java Gremlin Language Variant (GLV)
- **janusgraph-dotnet** - JanusGraph .NET Gremlin Language Variant (GLV)
- **janusgraph-python** - JanusGraph Python Gremlin Language Variant (GLV) (initial stage,
- not officially released but can be compiled from `initial-pr` branch)
- **janusgraph-docker** - JanusGraph Docker images
- **janusgraph-ambari** - JanusGraph Ambari plugin
- **janusgraph-cassandra** - Cassandra Thrift and Embedded Backend Adapter for JanusGraph
- **janusgraph-foundationdb** - FoundationDB storage adapter for JanusGraph
- **janusgraph.org** – Main web page repository
- **docs.janusgraph.org** – Documentation repository
- **logos** – JanusGraph logos
- **blog.janusgraph.org** – was intended to be a blog but the development didn't start
- **legal** – Repository to support CLA paper signers (deprecated as it was replaced by EasyCLA)

# 2. Architectural Overview




Supported storage backends: Cassandra, HBase, Bigtable, BerkeleyDB, ScyllaDB
Supported storage backends via adapters: Aerospike, DynamoDB, FoundationDB
Supported index backends: ElasticSearch, Solr, Lucene

# 3. Development roadmap

JanusGraph is a community driven project.

Currently JanusGraph roadmap is defined per release by TSC members.

**Roadmap for the next major release (0.6.0):**

- Performance optimization (index selection strategy, count query, index repair, cql iteration, mixed index support for `has(key)`, etc.)
- Add support for Solr 8 and Lucene 8
- JanusGraph server improvements (gRPC with Protocol Buffers for JanusGraph management, replacement of default GremlinServer by JanusGraph server with supported java configurations)
- Add support for GraphBinary serialization format
- DataStax driver upgrade to 4.x version
- Support drop: Thrift protocol, Cassandra 2.x

# 4. Community stats

Stats as of December 1, 2020.

GitHub main repository stats:

- Stars: 3692
- Forks amount: 923
- Contributors amount: 129
- Master branch commits amount: 6119
- Closed Pull Requests: 977
- Opened Pull Requests: 41
- Closed Issues: 846
- Opened Issues: 395

Gitter users' chat participants: 696
Gitter developers' chat participants: 27
Amount of JanusGraph users Google group threads: 1535
Amount of JanusGraph developers Google group threads: 295
Amount of StackOverflow tagged questions with [janusgraph]: 637
JanusGraph Twitter followers: 1386
JanusGraph LinkedIn followers: 229
JanusGraph takes 5 place out of 32 by DB-Engines Ranking of Graph DBMS

# 5. Governance

Technical Steering Committee (TSC) members can elect new TSC members and new Committer members.

TSC members are responsible for the direction of the project and for the release lifecycle. Committer members have direct write access to the project repositories.

Organization stats as of December 1, 2020:

Committers team includes 24 people
Maintainers team (TSC members) includes 9 people
Total amount of organization members: 30

# 6. Events

In addition to general events where JanusGraph has a chance to be presented, the community members often start online Meetups with several JanusGraph sessions and Q&A.

- JanusGraph Online Meetup, Bruno Berriso, Florian Grieskamp, & Ted Wilmes - 2020.10.07
- JanusGraph Online Meetup, Ryan Stauffer, Ted Wilmes, Aaron Ploetz, Becky Nelson, & Rick Paiste – 2019.08.07
- JanusGraph Online Meetup, Chris Hupman, Ryan Stauffer, Jan Jansen, John Mertic, & Ted Wilmes - 2019.03.27
- DataWorksJun2017: Large Scale Graph Analytics with JanusGraph, P. Taylor Goetz, 2017.06.13
- HBaseCon2017 Community-Driven Graphs with JanusGraph, Jing Chen He & Jason Plurad, 2017.06.12

# 7. Production users

There are next confirmed current or past production users who gave permission to list them in JanusGraph:
Times Internet, Target, RedHat, 360.cn, G Data, Finc, Compose (an IBM company), Celum, Netflix, Uber

There are next confirmed projects which are currently powered or were powered by JanusGraph and gave permission to list them in JanusGraph:
Apache Atlas, Eclipse Keti, Exakat, Express-Cassandra,Unifi Catalog & Discovery, Uber Knowledge Graph, Windup by RedHat, Open Network Automation Platform (ONAP)

# 8. Collaboration opportunities

- Egeria (graduate project) has a number of open JanusGraph connectors. A direct collaboration could be done to improve these connectors.
- Amundsen (incubating project) has “Amundsen Metadata Service” project which can use either Neo4j or Apache Atlas as a persistent layer. Apache Atlas uses JanusGraph internally which creates indirect collaboration between Amundsen and JanusGraph but we can go even farther and add direct JanusGraph support to the project.

# 9. Benefits

- Legal and management support of the project sponsorship systems (creation and management of sponsorship platforms)
- Presenting JanusGraph on LF AI & Data events. Possible usage of LF AI & Data booth to promote the project. Presence in LF AI & Data Day in China
- Marketing services and possible management / co-management of existing media channels
- Enterprise GitHub Actions to increase the number of concurrent jobs

# 10. Conclusion

- JanusGraph is strongly supported project by different organizations. Thus, it is expected that the project will see see the ongoing development.
- JanusGraph takes a big role in AI because it can be used to store and query data in a distributed fashion. This database suites well to represent neural networks and process large sets of data using distributed computation.
- JanusGraph suites not only to AI projects, but also to any project which uses strongly connected data (IOT, Social Networks, Malware & Fraud detection, Identity and access management, etc.).



# JanusGraph
# LF AI & DATA Incubation Project Proposal

Oleksandr Porunov
Principal Software Engineer
JanusGraph Technical Steering Committee

December 3, 2020

# TAC Vote on Project Proposal: JanusGraph

**Proposed Resolution:**

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

# Next Steps

LF AI & Data staff will work with JanusGraph to onboard the project leading to the announcement of the project joining LF AI & Data

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

Integrate the project with LF AI & Data operations

# Invited Presentation - RosaeNLG

Tech Lead -

- L udan Stoeckle - ludan.stoeckle@rosaenlg.org

LF AI & DATA

# Project Presentation: RosaeNLG

RosaeNLG is an open source Natural Language Generation (NLG) project. It aims to offer the same NLG features as product NLG solutions, to be **developer and IT friendly** for template configuration, and to provide NLG on both server-side and browser-side.

RosaeNLG is mainly implemented using TypeScript and JavaScript under Apache 2.0 license.

- GitHub: https://github.com/RosaeNLG/rosaenlg
- Presenter: Ludan Stoecklé ludan.stoeckle@rosaenlg.org
- Supporter: Jamil CHAWKI, Chair of LF AI Outreach Committee
- Contributors: Ludan Stoecklé (original author), Marco Riva (Italian), RedLab Paris (5 PhDs engagement)
- 57 000 lines of code, 100+ commits since first public version in Sept. 2019

# History & Context

- **In France we love written language and literature!** Albert Camus, Jean-Paul Sartre, Marcel Proust, Victor Hugo, George Sand, Émile Zola, Jules Verne, Simone de Beauvoir...
- Strong tradition of academic codification of the language
- Early French NLG ecosystem:
  - Fundamental research on NLG: *Génération automatique de textes en langues naturelles*, Laurence Danlos 1985
  - Yseop founded in 2007 (Arria 2013, Narrative Science 2010, Automated Insights 2007)
  - Strong adoption by the French banks
  - CoreNLG built by Société Générale
  - Specialized NLG service companies like Addventa, P-Val
- Europe is a linguistic playground: 24 official languages, 60 regional languages

# Two NLG Techniques

## Machine Learning NLG: GPT-2, GPT-3 etc.

- can learn on data or on text
- requires training data
- produces very nice to read texts, but riddled with errors

## Classic NLG with vendors Narrative Science, Arria NLG, Automated Insights, Yseop

- automates the production of relatively repetitive texts
- input is data
- requires a significant setup effort, defining explicitly what to say and how to say it (business rules, text templates)
- makes no errors
- used in production (Société Générale, BNP Paribas, Moodys)
- requires a **NLG engine** (like RosaeNLG)

# What are the Use Cases for NLG

- describe a product based on its features (SEO Search Engine Optimization)
- produce structured reports: risk reports, fund performance in the financial industry
- describe a situation: client summary before a meeting
- generate well formed chatbot answers

# Financial Fund Performance

- comment the monthly performance of financial funds
- performance in general and compared to the benchmark
- contributors and detractors per sector, per country

## Uni-Global - Equities Emerging Markets - AA-USD

### Fund performance

The fund returned +3.0% (gross of fees, in USD terms) in July, strongly outperforming its benchmark by 80bp (gross of fees, in USD terms), which increased by 2.2%. From a country point of view, our stock selection was a positive contributor to relative performance while country allocation was a positive contributor to excess returns.

### Largest contributors of the month

China (CHINA TELECOM CORP LTD-H and AGRICULTURAL BANK OF CHINA-H) and Brazil stocks selected for the portfolio added the most to the fund's performance.

In terms of absolute performance our positions in WALMART DE MEXICO SAB DE CV, INDIAN OIL CORP LTD and SOUTHERN COPPER CORP --- US were the standout gainers rising by 0.0%, 0.0% and 0.0%, respectively.

With a rise of 0.0%, ZHEN DING TECHNOLOGY HOLDING (Taiwan, Technology) was the top contributor to excess returns.

With a rise of 0.0%, WALMART DE MEXICO SAB DE CV (Food Retailing in Mexico) was also a solid contributor.

Finally, with a rise of 0.0%, INFOSYS LTD (India, Software) was also a solid contributor.

At sectorial level, our selection of Materials, Software and Banks sectors produced gains.

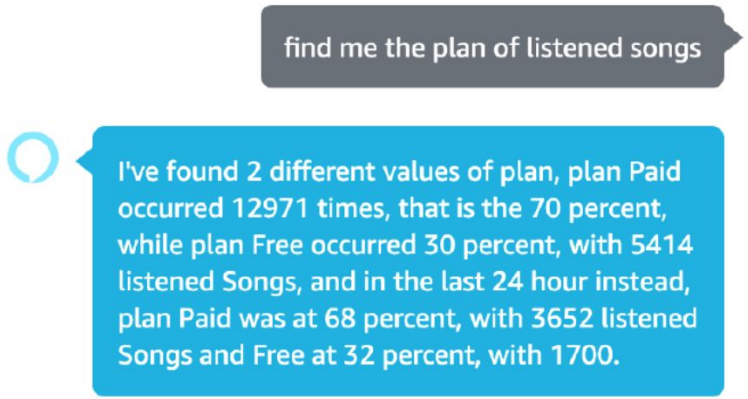
# Tires Description (SEO)

- generate unique original tire descriptions for 10.000 tire references
- duplicate content challenge over different resellers

| 838 MT #1 | Dueler H/P Sport #1 |
| --- | --- |
| Le 838 MT est destiné à des véhicules 4 roues motrices, il est développé pour un usage en tout-chemin.<br><br>Le 838 MT inséré par la marque Achilles est un pneumatique "été" prévu pour des véhicules de franchissement. Cette enveloppe est spécifiquement dédiée à être utilisée en dehors des sentiers battus.<br><br>Ce produit comporte des dimensions allant du 14 au 16 pouces. Il est disponible en 5 versions dimensionnelles. Ce pneu ne dispose pas de version spécifique pour un constructeur et est donc adapté à toutes les marques.<br><br>• Pneumatique au bon rapport performances/prix pour véhicules 4 roues motrices<br>• Économie mise en avant<br>• Robustesse en conditions off-road<br><br>random seed : 671 | Le pneumatique Dueler H/P Sport est un produit "été" du manufacturier Bridgestone. Ce pneumatique est prévu pour des véhicules de type SUV. Cette enveloppe est principalement dédiée à être utilisée en tout-chemin mais peut cependant s'adapter pour un usage routier.<br><br>Destiné à des véhicules de type SUV, le pneu "été" Dueler H/P Sport de la marque Bridgestone est spécifiquement développé pour une finalité sur routes et chemins cassants. Il peut malgré tout s'apprêter pour un usage routier.<br><br>Ce pneu bénéficie de diamètres qui vont du 16 au 21 pouces. Il possède dans sa gamme 123 dimensions. Ce pneu est disponible en versions spécifiques Porsche, Audi (AO), Mercedes (Mo), Maserati et BMW (*).<br><br>En cas de crevaison, la technologie "Roulage à plat" (autrement dit "Runflat" ) vous permettra de continuer à rouler de manière modérée sur une courte distance.<br><br>Ce pneumatique dispose de performances optimales pour une conduite familiale. Il bénéficie d'aptitudes importantes en termes de confort de conduite et de sécurité. C'est un produit fiable, ses capacités en matière de longévité sont excellentes.<br><br>• La Dueler H/P Sport haut de gamme appropriée aux SUV<br>• Prestations globales élevées pour une conduite familiale<br>• Freinage haute-performance sur sols secs et mouillés<br><br>random seed : 205 |

# Chatbot Answers

- querying database using natural language
- query result is transformed into text using RosaeNLG

# NLG Software Landscape

Main pain point of NLG products is **proprietary languages**: hard to learn, favor lock-in, not compatible with standard dev tools like VSCode or Git.

Open source alternatives are scarce:

- SimpleNLG (Mozilla Public License 2.0) focuses on a specific part of the NLG pipeline, and requires low-level Java coding (no templates)
- CoreNLG (Apache 2.0) requires coding in Python (no templates) and does not contain linguistic resources

# RosaeNLG - the Current Project Features

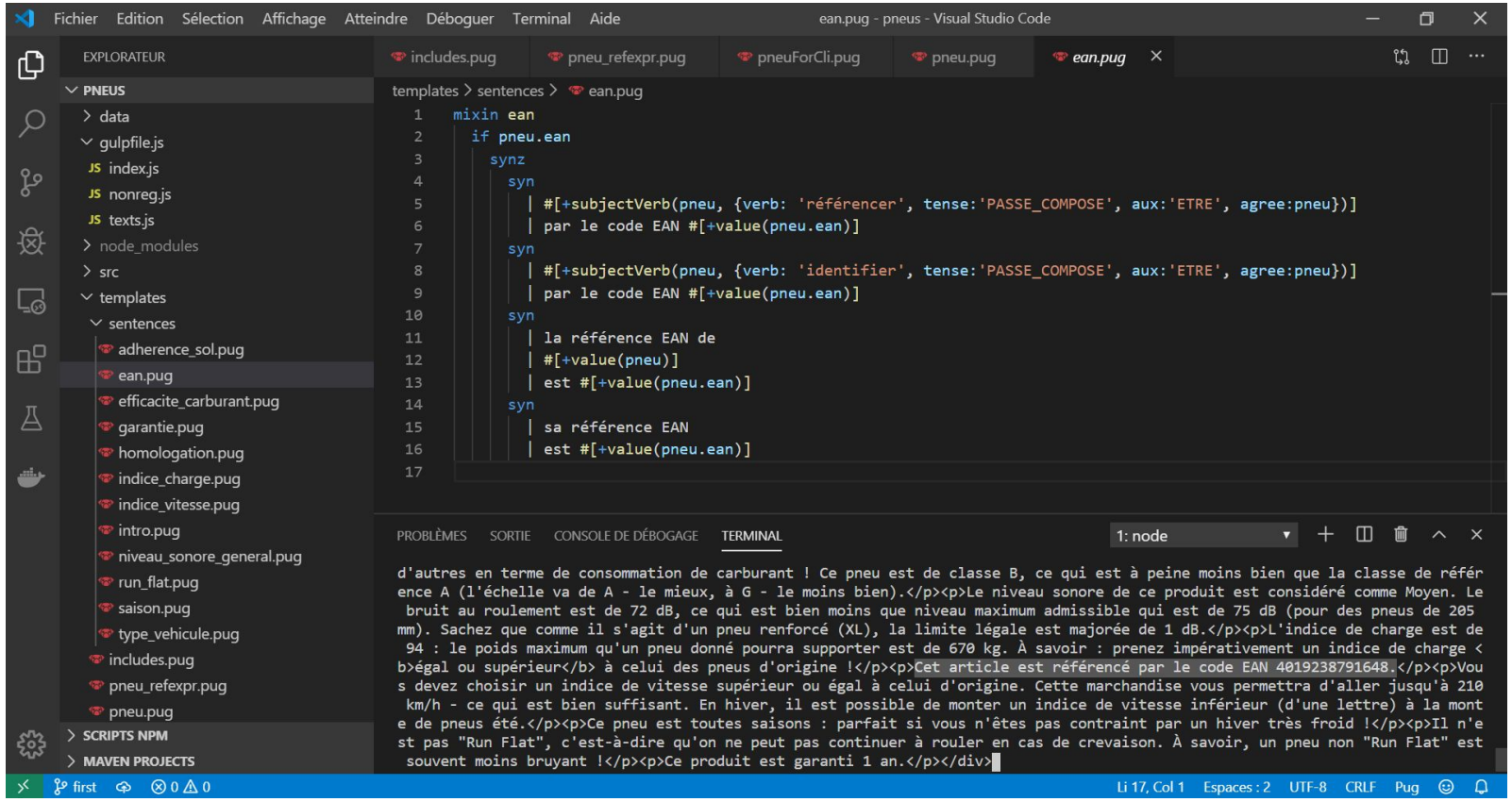
RosaeNLG is an open source NLG project with main developer Ludan Stoecklé (13 years in NLG):

- Designed to be **developer and IT friendly** for template configuration, using VSCode
- Supports multiple languages (currently English, French, German, Italian and Spanish)
- Provides NLG on **both server-side** (using node.js REST API) **and browser-side**
- Extensive documentation

RosaeNLG.org

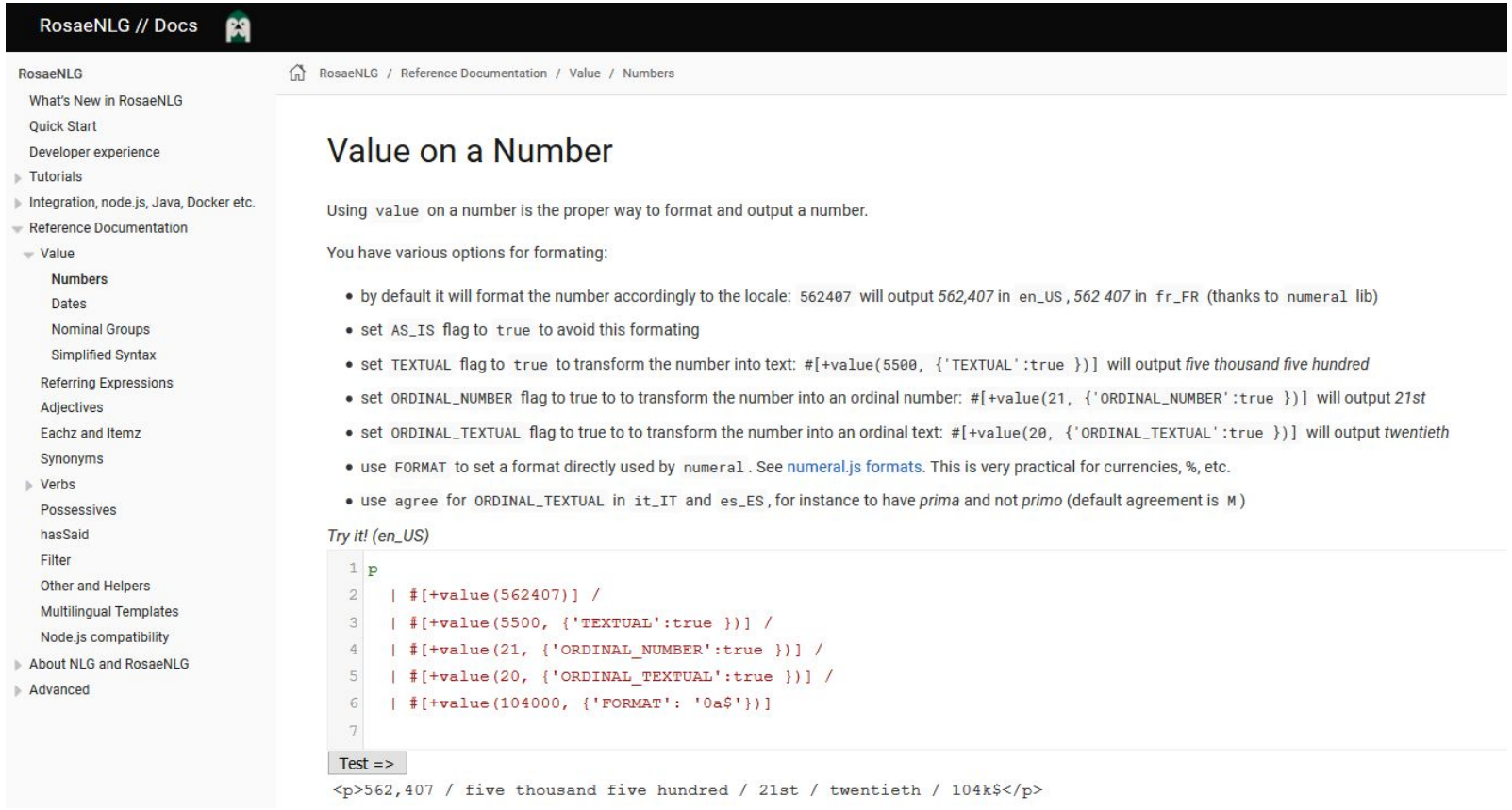
# Develop using VSCode

- templates development with syntax highlighting
- render texts
- regression testing



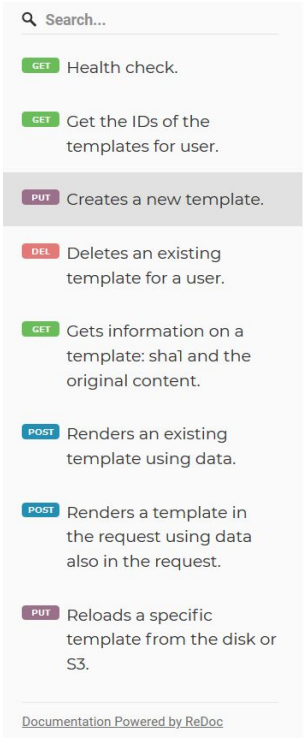
# Reference Documentation

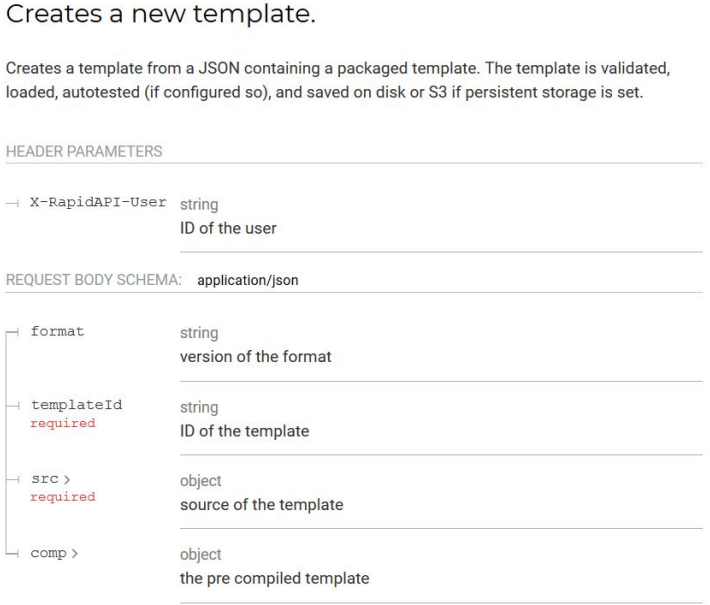
- extensive documentation
- versioned
- search engine
- test snippets of code directly in the browser

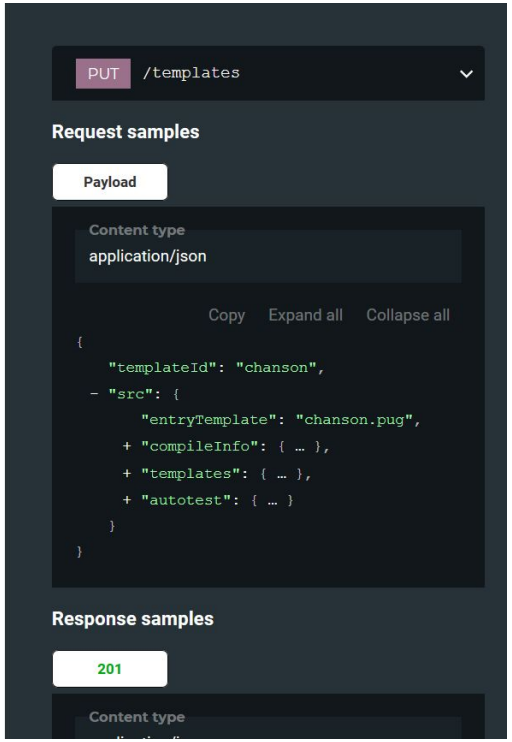


# RosaeNLG node.js API

- REST API: load a template, then render using data
- templates can be stored either on disk or S3
- ideal for a NLG micro service
- packaged in a Docker image




*there is also a Java version*

# Possible Collaborations with LF AI Projects

RosaeNLG currently runs on Acumos for Orange AI Marketplace.

RosaeNLG can be used at the end of the AI pipeline, to **explain a decision** to non-experts:

- **AI Explainability 360:** provide a clear, readable, summarized explanation for an end user (e.g. Bank Customer) asking for explanations
- **AI Fairness 360:** generate comprehensive compliance reports on fairness (initial situation, what was done, final situation)

RosaeNLG is positioned at the same level as LF AI Delta: use data to create a business service.

# RosaeNLG Burgeoning Ecosystem

## Corporate:

- Addventa (company specialized in NLG, based in Paris) provides commercial support on RosaeNLG (support with SLA and Professional Services)
- RosaeNLG is available for commercial usage on Orange AI marketplace
- Specialized technology companies: Lizeo (tires descriptions), Radicalbit (natural language querying of databases)
- Financial corporations on POCs and production: Exane, BNP Paribas

## Academic:

- Used in thesis (Marco Riva, *Making a Time-Series Database "smart": human and machine communication towards conversational analytics*, Laurea in Informatica, Università degli Studi di Milano, 2020), also **contributor on Italian version** of RosaeNLG
- Official commitment from RedLab Paris to dedicate 5 PhD to contribute to the open source version of RosaeNLG

# Ambition & Roadmap

## Ambition

- To **become the widely used NLG open source project**:
  - in corporate custom NLG projects
  - to power NLG features of any software
  - Be embedded in dashboarding software
  - Power NLG products
- To **support more than 50 commonly spoken languages**
- Standardize NLG templating language (whatever the underlying implementation)

## Roadmap

- NLG library to ease number analysis
- Improve ability to add new languages
- More languages: Arabic, Chinese, Indian languages, Finnish, etc. - depending on contributors
- Increase code quality
- Dedicated VSCode plugin, with template debug support
- Industrialize Java version
- Collaborate with current LF AI projects (AI Explainability 360 & AI Fairness 360)
- Onboard contributors like Redlab Paris

# Questions

Monthly community meeting every first Thursday of the month.
Next (and first) will take place Thursday 7th of January 2021 18:00 CET.

The time of the meeting changes every month:

- 18:00 CET January / March etc. - which is friendly for Europe and the US.
- 9:00 CET February / April etc. - which is friendlier for Russia and Asia.

# Appendix

# NLG Engine - Features

The main features of a NLG engine are:

1. the ability to properly enumerate (xxx, yyy and zzz)
2. the proper agreement of verbs, nouns, adjectives
3. the use of synonyms and referring expression to avoid repetitions
4. proper punctuation, spacing, capitalization and contractions

Some features depend of the output language.

String concatenation or standard template engines can be used to generate texts. But a NLG project without a NLG engine is a nightmare.

# NLG Text Templates

NLG templates combine:

- static texts
- structures (e.g. conditions, loops, lists)
- NLG functions (e.g. agreements, conjugations)
- local processing using code (e.g. filtering, sorting)

These templates are run by a specific template engine: the **NLG engine**.

Edit your RosaeNLG template: *based on "noms et adjectifs"* fr_FR

```
1 | ils #[+verb(getAnonMP(), {verb:'vouloir', tense:'PASSE_COMPOSE'})]
2 itemz {begin_with_general: 'à la fois', separator: ',', last_separator:'et', end:'.', mix:true}
3   item
4     | un moteur de NLG
5   item
6     | #[+value('caillou', {det:'DEFINITE', adj:'beau', adjPos:'BEFORE', number:'P'})]
7   item
8     | #[+value('plage', {det:'DEFINITE', adj:'beau', adjPos:'BEFORE', number:'P'})]
9
```

☑ Render automatically

```
done!
```

Rendered texts: show as html

```
Ils ont voulu à la fois un moteur de NLG, les
beaux cailloux et les belles plages.
```

# LF AI & Data - General Updates

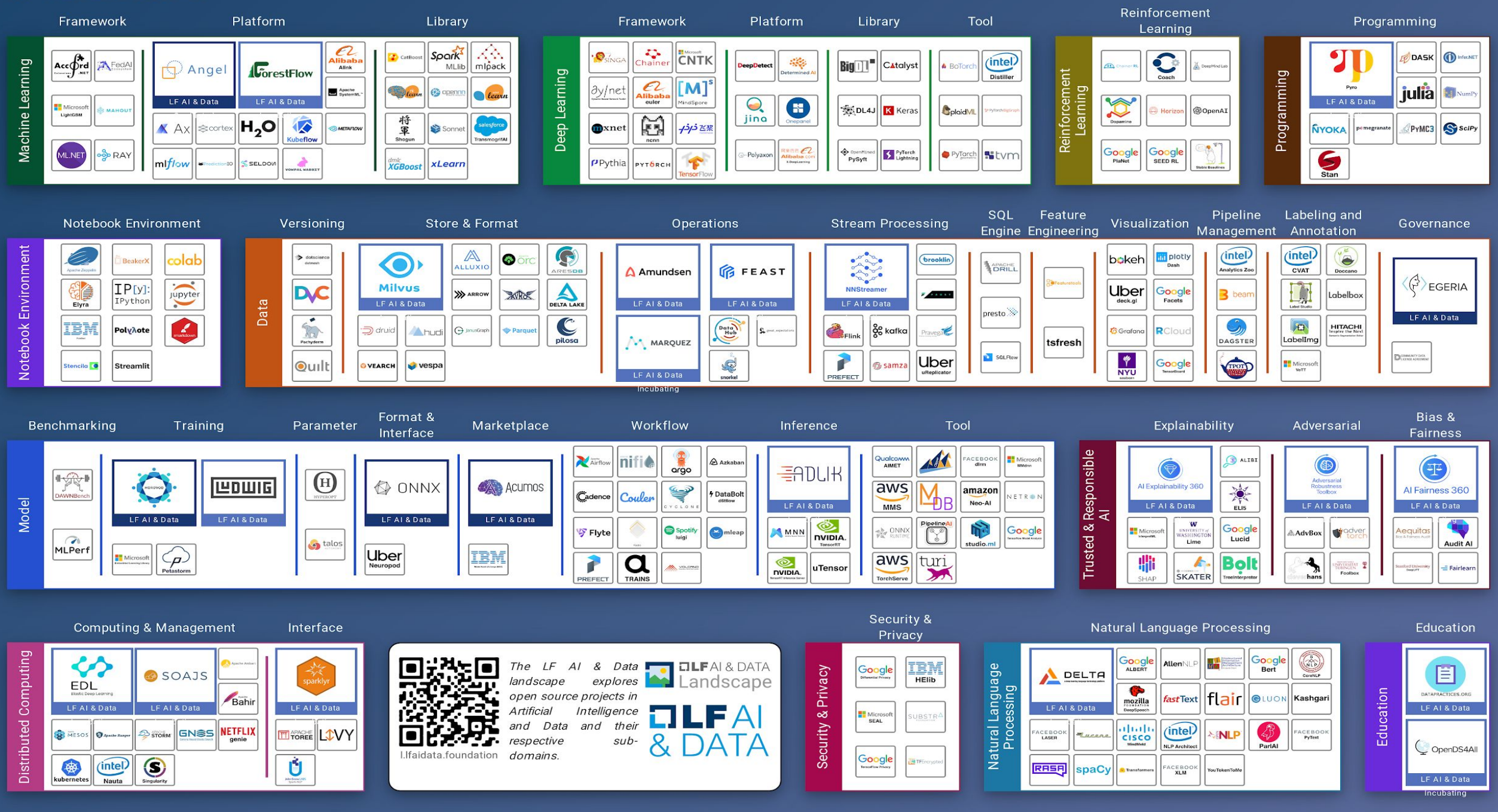
LF AI & DATA

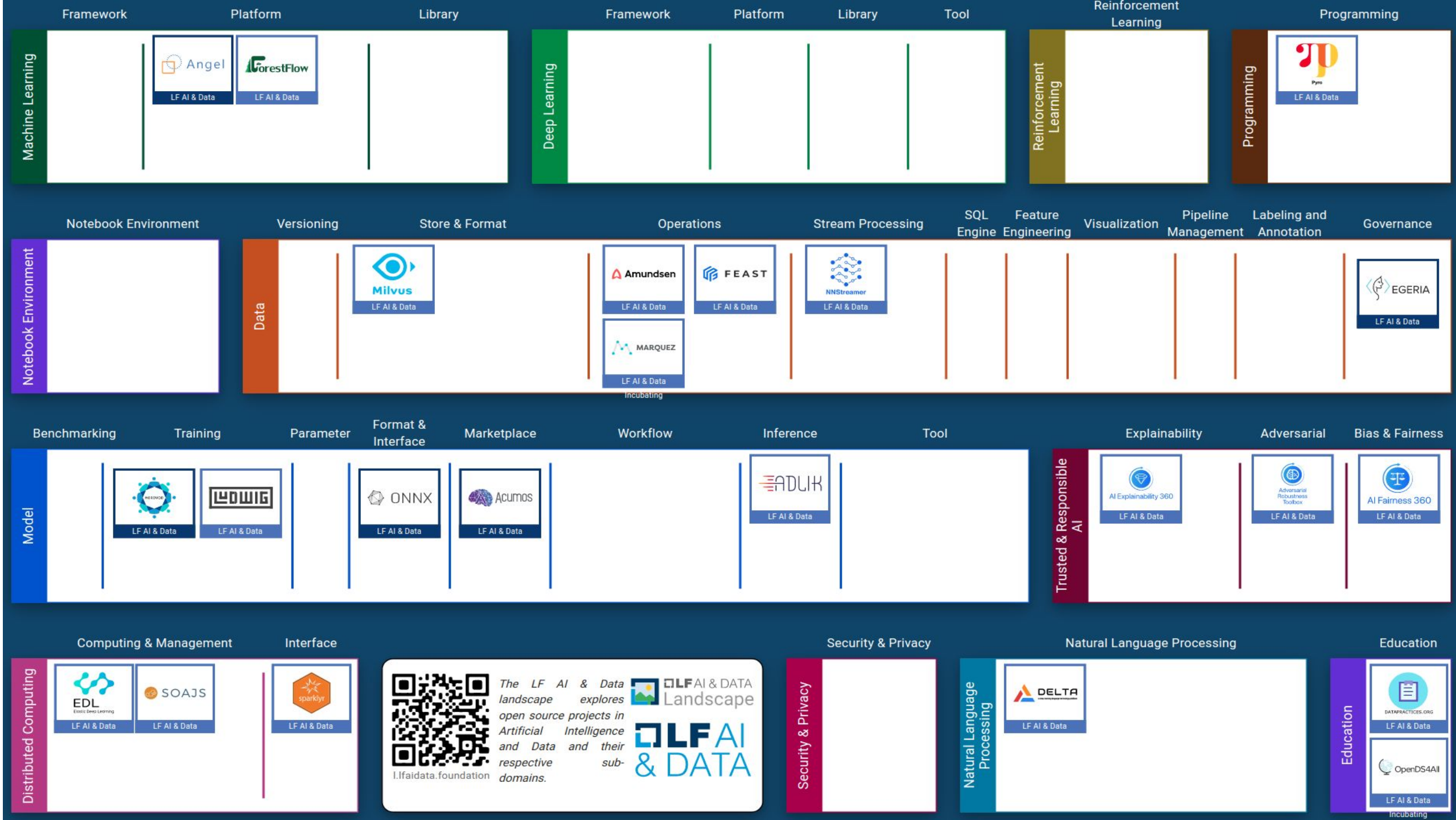
# Linux Foundation AI Landscape

2020-11-22T22:28:27Z b08f33f

See the interactive landscape at l.lfai.foundation

Greyed logos are not open source




Framework
Platform
Library
Machine Learning
Angel
LF AI & Data
ForestFlow
LF AI & Data
Framework
Platform
Library
Tool
Deep Learning
Reinforcement Learning
Reinforcement Learning
Programming
Programming
Pyro
LF AI & Data
Notebook Environment
Notebook Environment
Versioning
Store & Format
Operations
Stream Processing
SQL Engine
Feature Engineering
Visualization
Pipeline Management
Labeling and Annotation
Governance
Data
Milvus
LF AI & Data
Amundsen
LF AI & Data
FEAST
LF AI & Data
MARQUEZ
LF AI & Data
Incubating
NNStreamer
LF AI & Data
EGERIA
LF AI & Data
Benchmarking
Training
Parameter
Format & Interface
Marketplace
Workflow
Inference
Tool
Model
LF AI & Data
LF AI & Data
ONNX
LF AI & Data
Acumos
LF AI & Data
ADLIK
LF AI & Data
Explainability
Adversarial
Bias & Fairness
Trusted & Responsible AI
AI Explainability 360
LF AI & Data
Adversarial Robustness Toolbox
LF AI & Data
AI Fairness 360
LF AI & Data
Computing & Management
Interface
Distributed Computing
EDL
LF AI & Data
SOAJS
LF AI & Data
sparklyr
LF AI & Data
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
Security & Privacy
Natural Language Processing
Natural Language Processing
DELTA
LF AI & Data
Education
Education
DATAPRACTICES.ORG
LF AI & Data
OpenDS4All
LF AI & Data
Incubating

# Suggested Additions

**Project Key**

Yellow = not in Landscape, maybe should be added

## Programming
Numpy
Numba
SciPy
Dask
Julia (*)
Python
Rstudio

## Notebooks
Elyra
I-python
Jupyter Notebooks
PixieDust
Rmarkdown

## Security & Privacy
HE-Lib (*)
TensorFlow Privacy
TF-Encrypted

## Distributed Computing

*Management*
OpenShift
Kubernetes
Mesos
Ranger
Storm

*Interface*
Sparklyr
Toree
Livy
Spark-NLP

## Data

*Versioning*
Pachyderm (*)

*Store & Format*
Alluxio
Arrow
Avro
Delta Lake (*)
Druid
JanusGraph
Parquet
Ceph

*Stream Processing*
Flink
Kafka
Logstash (*)
FluentD (*)

*Relational DB*
Postgres
MySQL
CouchDB

*SQL Engine*
Presto (*)

*Visualization*
Bokeh
D3
Plotly
Facets
Grafana
Seaborn
Superset (*)
TensorBoard
Prometheus

## Data

*Governance*
Egeria
CLDA

*Feature Engineering*
Tsfresh

*Operations*
FEAST (*)
Amundsen (*)
Hive (*)
Snorkel (*)

*Pipeline Management*
Beam

*Labeling & Annotation*
Vott (*)

*Exploration*
Hue
Kibana

## Machine Learning

*Framework*
LightGBM
Mahout
Ray (*)

*Platform*
Kubeflow
H2O
SystemML
Mlflow (*)
Seldon (*)
Marvin-AI (*)

*Library*
Scikit-learn
XGBoost
cat-boost
SparkML

## Deep Learning

*Framework*
TensorFlow
PyTorch
MX-Net

*Libarary*
Keras

## Reinforcement Learning
DeepMind Lab (*)
OpenAI Gym (*)

## Model

*Inference*
TensorRT
TensorRT Inference

*Benchmarking*
MLPerf

*Training*
Horovod (*)

*Parameter*
HyperOpt
Katib

*Format & Interface*
ONNX

*Marketplace*
MAX (*)

*Workflow*
Kubeflow Pipelines
Tekton
Airflow (*)
Nifi (*)
Argp (*)
Mleap (*)
Volcano (*)

*Tool*
KFServing
ONNX Runtime
TorchServe (*)
Clipper (*)
MMS (*)

## Trusted AI

*Explainability*
AI Explainability 360
Alibi (*)
LIME
SHAP

*Bias & Fairness*
AI Fairness 360

*Adversarial Attacks*
Adversarial Robustness Toolbox

## Natural Language Processing
UIMA
BERT
Core NLP
Lucene
PyText
Spacy
Transformers (*)

## Education
OpenDS4All

LF AI & DATA

# 2020 TAC Meetings Summary

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

(Entity)* = incubating vote

** **bold = graduate vote**

*Italics = invited project presentation*

# 2021 TAC Meetings Pipeline Summary

| Months | | | |
|---|---|---|---|
| Jan<br>Feb<br>Mar | ?: DataHub (LinkedIn) | ?: Ray (Anyscale.io) | ?: Couler (Ant Financial) |
| Apr<br>May<br>Jun | ?: Kubeflow-On-Prem (Google, Arrikto, Intel) | ?: Data Lifecyle Framework (IBM)<br>?: Pachyderm (Pachyderm) | ?: Common Knowledge (Code Reef) |
| Jul<br>Aug<br>Sep | ?: KubeflowServing (Google, Arrikto, Seldon) | ?: Kubeflow Pipeline (Google, Bloomberg) | ?: Open Data Hub (Red Hat) |
| Oct<br>Nov<br>Dec | ?: Vespa (Verizon Media) | ?: Snorkle (Snorkle)<br>?: Plotly (DASH)<br>?: Mellody (Substra)<br>?: mloperator (Polyaxen)<br>?: SnapML (IBM) | ?: PMML/PFA (DMG.org)<br>?: Mindspore, Volcano (Huawei)<br>?: TransmorgrifAI (Salesforce)<br>?: AIMET (Qualcomm)<br>?: Elyra-AI (IBM) |

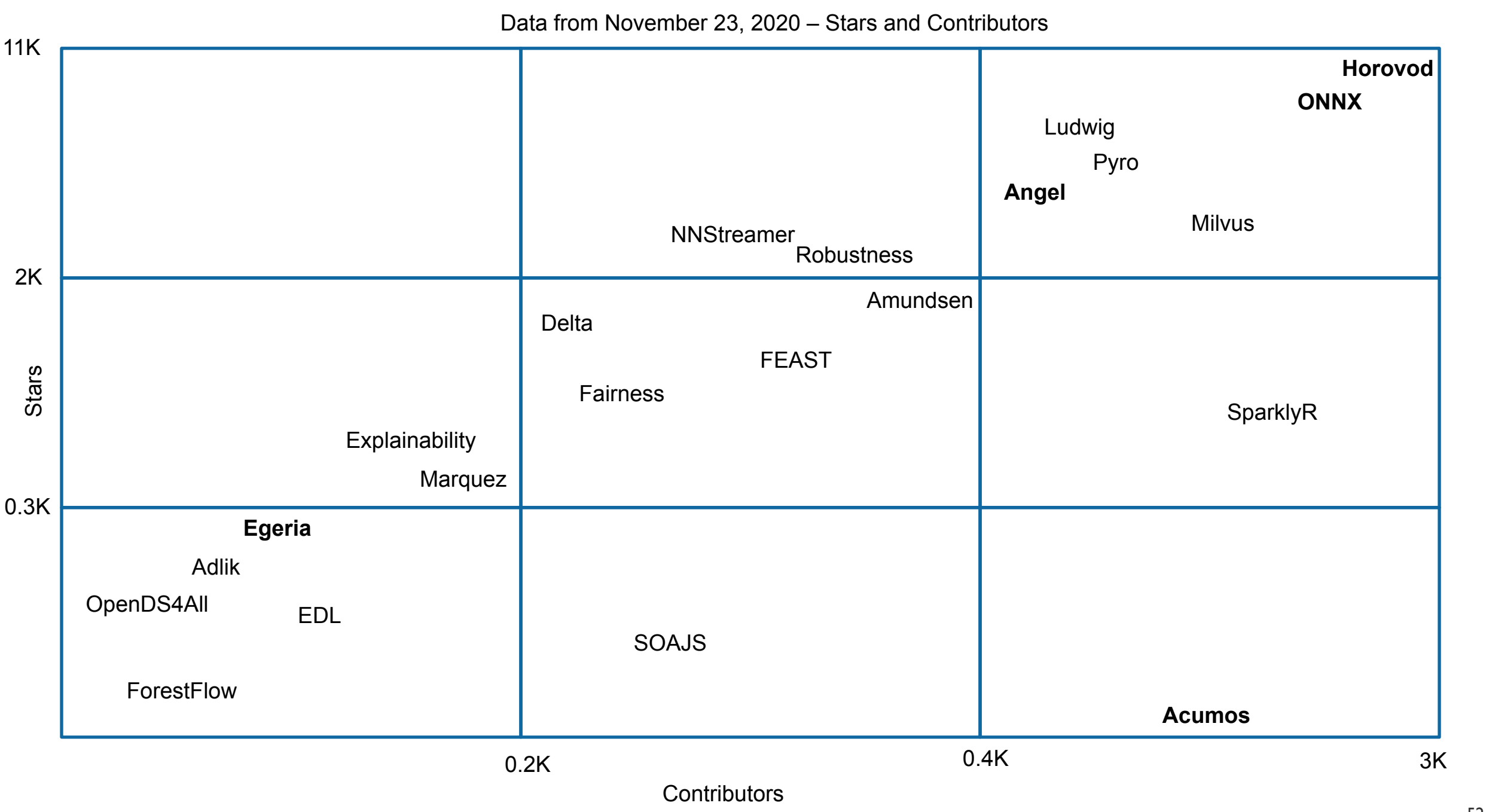
(Entity)* = incubating vote

** **bold = graduate vote**

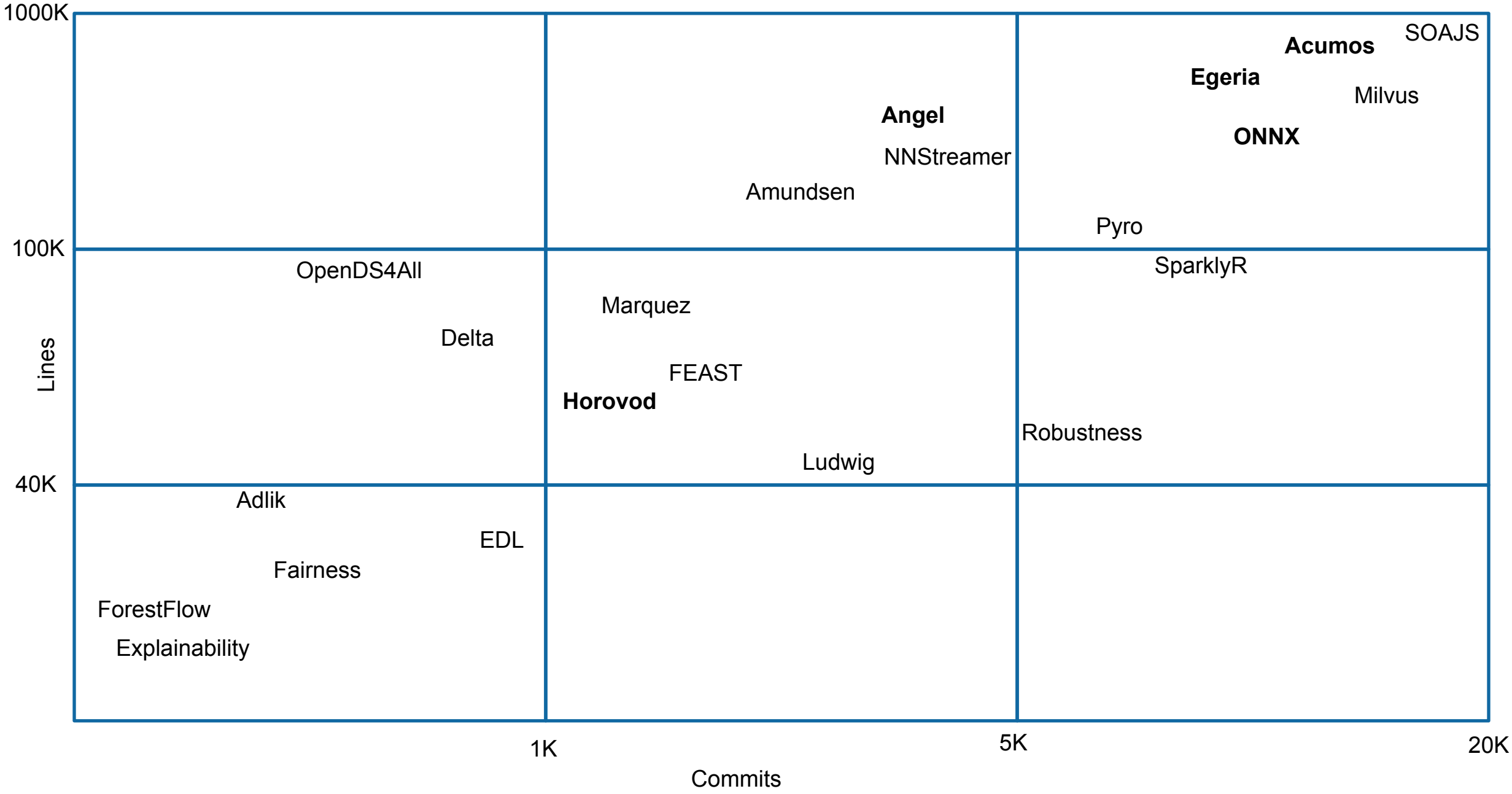
*Italics = invited project presentation*

# Getting to know the projects more

LF AI & DATA


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 from November 23, 2020 – Lines of Code and Commits



# 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@lfaidata.foundation

- Please consider holding virtual events

To discuss participation, please email events@lfaidata.foundation

# Upcoming Events

**Upcoming**

**“AI/ML/DL presented by LF AI Foundation” Track at OSS Japan – Dec 2-4**

**Just completed**

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

# 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
- **Github:** https://github.com/lfai/trusted-ai
- **Wiki:** https://wiki.lfai.foundation/display/DL/Trusted+AI+Committee
- **To participate:** https://lists.lfaidata.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

# ML Workflow & Interop

- **Leadership:** Huang “Howard” Zhipeng (Huawei)
- **Goal:** Define an ML Workflow and promote cross project integration
- **Wiki:** https://wiki.lfaidata.foundation/pages/viewpage.action?pageId=10518537
- **To participate:** https://lists.lfaidata.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

# 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 direction for your organization.
- **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.lfaidata.foundation/g/aiethics-training**

# Upcoming TAC Meetings

LF AI & DATA

# Upcoming TAC Meetings

- **December 17th:** Tentative Pyro graduated project proposal and additional new incubation project
- **December 31st :** No meeting - New Year's Eve Holiday
- **January 14th:** TBD

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

LF AI & DATA