# Meeting of the Technical Advisory Council (TAC)

March 11, 2021

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: 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  (5 mins)
- Approval of Minutes from Feb 25 (5 mins)

- Sandbox Proposal  (25 minutes)
  - RosaeNLG (Ludan Stoeckle)
- Invited Presentation  (15 minutes)
  - Elyra/Jupyter Ecosystem (Luciano Resende)

- LF AI General Updates (5 minutes)
- Open Discussion (5 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 |
| Pyro | Fritz Obermeyer* | fritz.obermeyer@gmail.com |

# Approval of February 25th, 2021 Minutes

Draft minutes from the February 25th TAC call were previously distributed to the TAC members via the mailing list

**Proposed Resolution:**

- That the minutes of the February 25th meeting of the Technical Advisory Council of the LF AI & Data Foundation are hereby approved.

# Sandbox Proposal - RosaeNLG

Ludan Stoecklé <ludan.stoeckle@gmail.com>

LF AI & DATA

# Project Contribution Proposal Review & Discussion: RosaeNLG

RosaeNLG is a Natural Language Generation library for node.js or client side (browser) execution, based on the Pug template engine. Based on Pug-like textual templates and on input data, RosaeNLG will generate high quality texts. Fully supported languages (with grammar, gender etc.) are *English*, *French*, *German*, *Italian* and *Spanish* but you can generate texts in any other language with less features. RosaeNLG is complete enough to write production grade real life NLG applications.

**Presenter:** Ludan Stoecklé"<ludan.stoeckle@gmail.com>

**Resources:**

Github: https://github.com/RosaeNLG/rosaenlg
Project Level: Sandbox
Proposal: https://github.com/lfai/proposing-projects/blob/master/proposals/RosaeNLG.adoc

# Project Contribution Proposal: 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 configuration, and to provide NLG on both server-side and browser-side.

RosaeNLG is 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)
- 60 000 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
  - 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, Pegasus etc.

- input is most often text (sometimes data)
- models have an out-of-the box knowledge about language and the world
- good at producing the text, but cannot (yet) learn intermediate steps of a reasoning (like calculations)
- hard to specialize on specific use cases
- texts are riddled with false information: the model invents facts

**Template-based NLG:** vendors Narrative Science, Arria NLG, Automated Insights, Yseop...

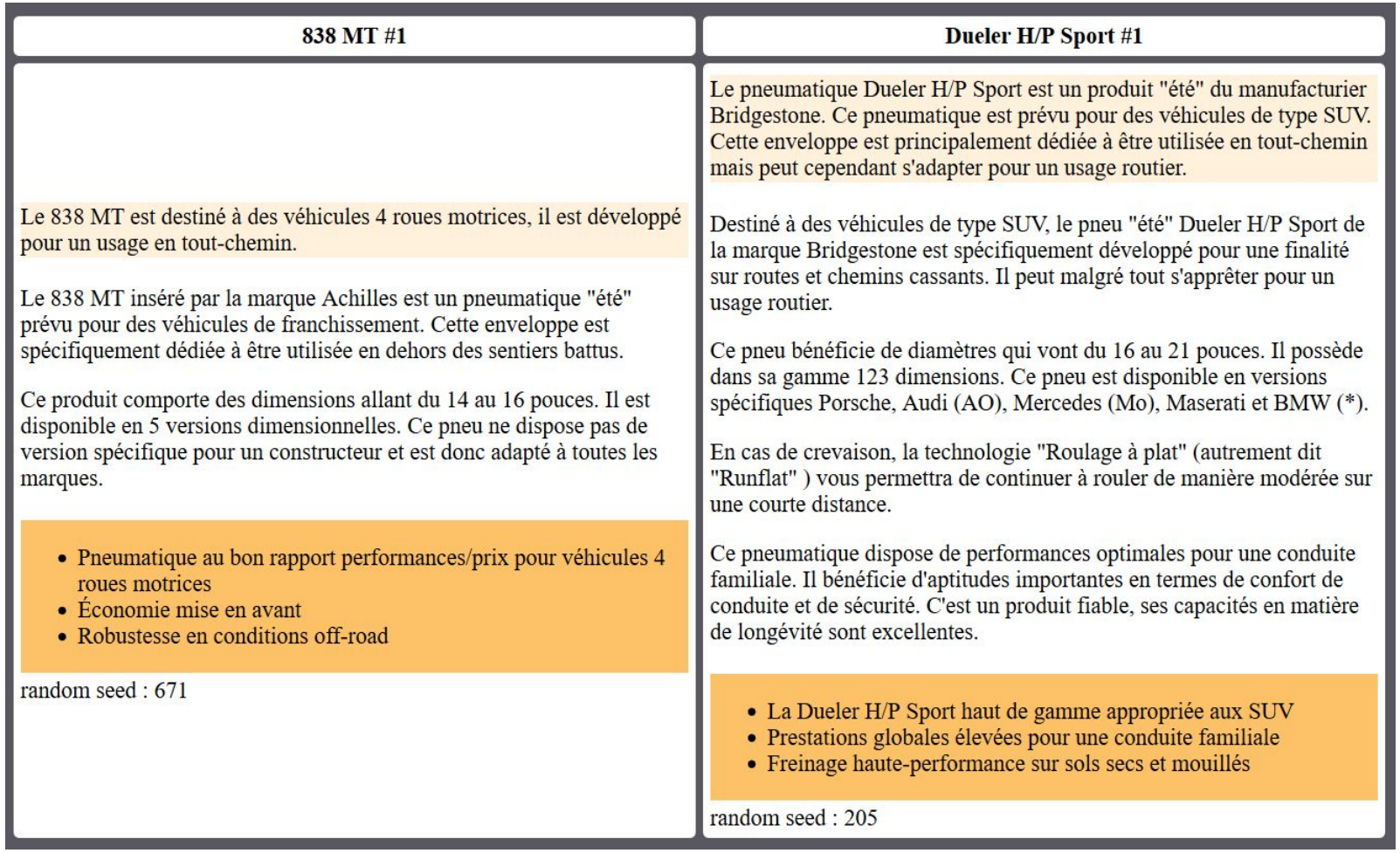
- input is always data
- automates the production of relatively repetitive texts
- requires a significant setup effort, defining explicitly what to say and how to say it, using text templates
- makes no errors
- used widely in production (Société Générale, BNP Paribas, Moodys)
- requires a **NLG engine**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

# What are the Use Cases for NLG

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

# Tires Description for SEO (Lizeo)

- generate unique original tire descriptions for 10.000 tire references



| 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 |

# 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.

# Automated comments of company financial statements data (Exane BNP Paribas)

- build a core knowledge base of financial analysis rules and apply them over a large set of industrial firms
- use a multilingual corpus of financial texts, rather than translating each document

**Accor** =

FINANCIAL HIGHLIGHTS

| Per Share data (EUR) | Dec. 17 | Dec. 18 | Dec. 19 | Dec. 20e | Dec. 21e | Dec. 22e |
|---|---|---|---|---|---|---|
| EPS restated, fully diluted | 1.50 | 1.27 | 0.89 | (1.79) | (0.33) | 0.70 |
| *% change* | *+46.4%* | *-15.3%* | *-30.3%* | *NC* | *+81.6%* | *NC* |
| Book value (BVPS) | 18.8 | 22.4 | 25.2 | 22.4 | 22.0 | 22.8 |
| Net dividend | 1.05 | 1.05 | 1.05 | 0.00 | 0.00 | 0.49 |
| Pay out | 70.1% | 82.7% | 118.6% | 0.0% | 0.0% | 70.0% |
| **Valuation metrics** | **Dec. 17** | **Dec. 18** | **Dec. 19** | **Dec. 20e** | **Dec. 21e** | **Dec. 22e** |
| P/E (x) | 27.0 | 34.2 | 42.9 | NC | NC | 43.2 |
| Net yield (%) | 2.6% | 2.4% | 2.8% | 0.0% | 0.0% | 1.6% |
| FCF yield (%) | 6.2% | 1.1% | 3.6% | -8.7% | -1.8% | 2.2% |
| EV/Sales (x) | 3.20 | 3.81 | 2.91 | 4.65 | 3.16 | 2.63 |
| EV/EBITDA (x) | 14.3 | 19.3 | 14.3 | NC | 41.9 | 15.4 |
| EV/EBITA (x) | 17.9 | 25.0 | 23.7 | NC | NC | 29.4 |
| **Income statement (EURm)** | **Dec. 17** | **Dec. 18** | **Dec. 19** | **Dec. 20e** | **Dec. 21e** | **Dec. 22e** |
| Sales | 2,773 | 3,610 | 4,049 | 2,146 | 3,193 | 3,759 |
| Organic sales growth | +7.9% | +8.8% | +3.8% | -52.0% | +64.5% | +28.7% |
| Restated EBITDA | 622 | 711 | 825 | (347) | 241 | 640 |
| Restated EBITDA margin | 22.4% | 19.7% | 20.4% | -16.2% | 7.5% | 17.0% |
| Restated EBITA | 496 | 549 | 497 | (521) | (18) | 335 |
| Restated EBITA margin | 17.9% | 15.2% | 12.3% | -24.3% | -0.6% | 8.9% |
| Net Profit (restated) | 432 | 367 | 241 | (485) | (89) | 191 |
| ROE | 7.9% | 5.8% | 3.5% | -8.0% | -1.5% | 3.1% |
| ROCE | 10.6% | 6.9% | 5.9% | -9.4% | -0.3% | 4.3% |
| **Cash Flow (EURm)** | **Dec. 17** | **Dec. 18** | **Dec. 19** | **Dec. 20e** | **Dec. 21e** | **Dec. 22e** |
| Operating cash flow | 1,110 | 621 | 749 | (390) | 204 | 698 |
| Capex | (189) | (308) | (209) | (140) | (140) | (200) |
| Free cash flow | 762 | 147 | 389 | (753) | (155) | 190 |
| Dividends paid | (163) | (323) | (294) | (282) | 0 | 0 |
| **Net debt (EURm)** | **Dec. 17** | **Dec. 18** | **Dec. 19** | **Dec. 20e** | **Dec. 21e** | **Dec. 22e** |
| Adjusted net debt | 2,776 | 2,040 | 1,912 | 2,183 | 2,305 | 2,083 |
| Net financial debt (cash) | 1,889 | 1,153 | 785 | 1,089 | 1,244 | 1,055 |
| Gearing | 48% | 32% | 35% | 44% | 46% | 41% |
| Adj. financial debt / EBITDA | 4.5x | 2.9x | 3.0x | NC | 11.9x | 4.1x |

White label

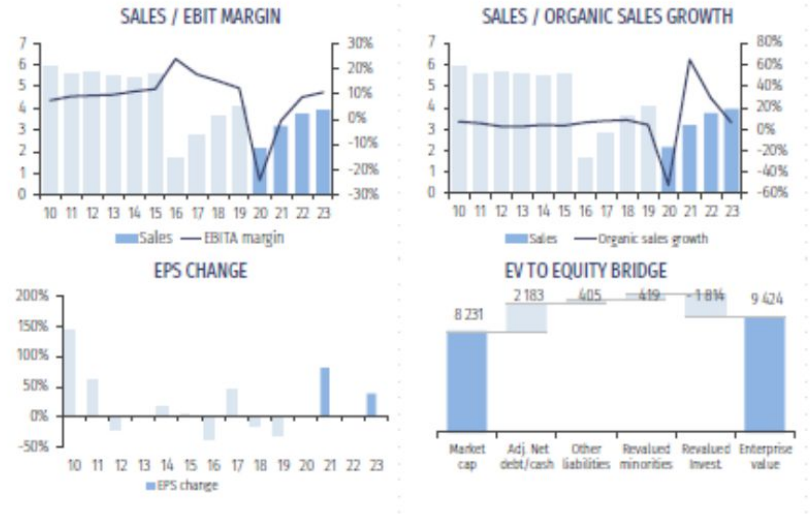
COMMENTS

**Valuation**

The company uses limited financial leverage.

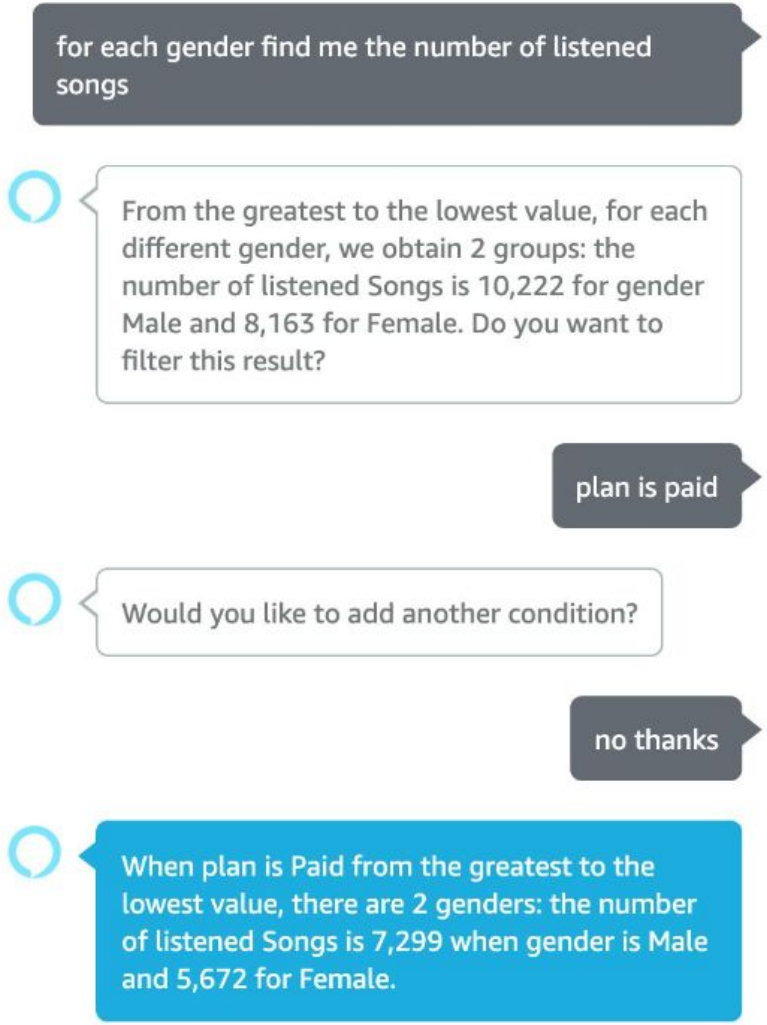
**Growth & Returns**

The company delivered 8.5 % sales growth in the past 5 years on average, mainly driven by organic growth. Free Cash Flow was fairly volatile over the last 10 years ranging between a minimum of -753 in 2020 and a maximum of 762 in 2017. The group has not created value for the past recent years and ROCE of -9.4 % in 2020 should stay significantly lower than the cost of capital of 6.2 %. Group profitability is low, with an ROE reaching -8 %, close to the average 5 past years.



# Chatbot Answers (Radicalbit)

- 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 development 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

There is a need for a **template-based, open source** NLG engine using pre packaged **linguistic resources**.

# RosaeNLG

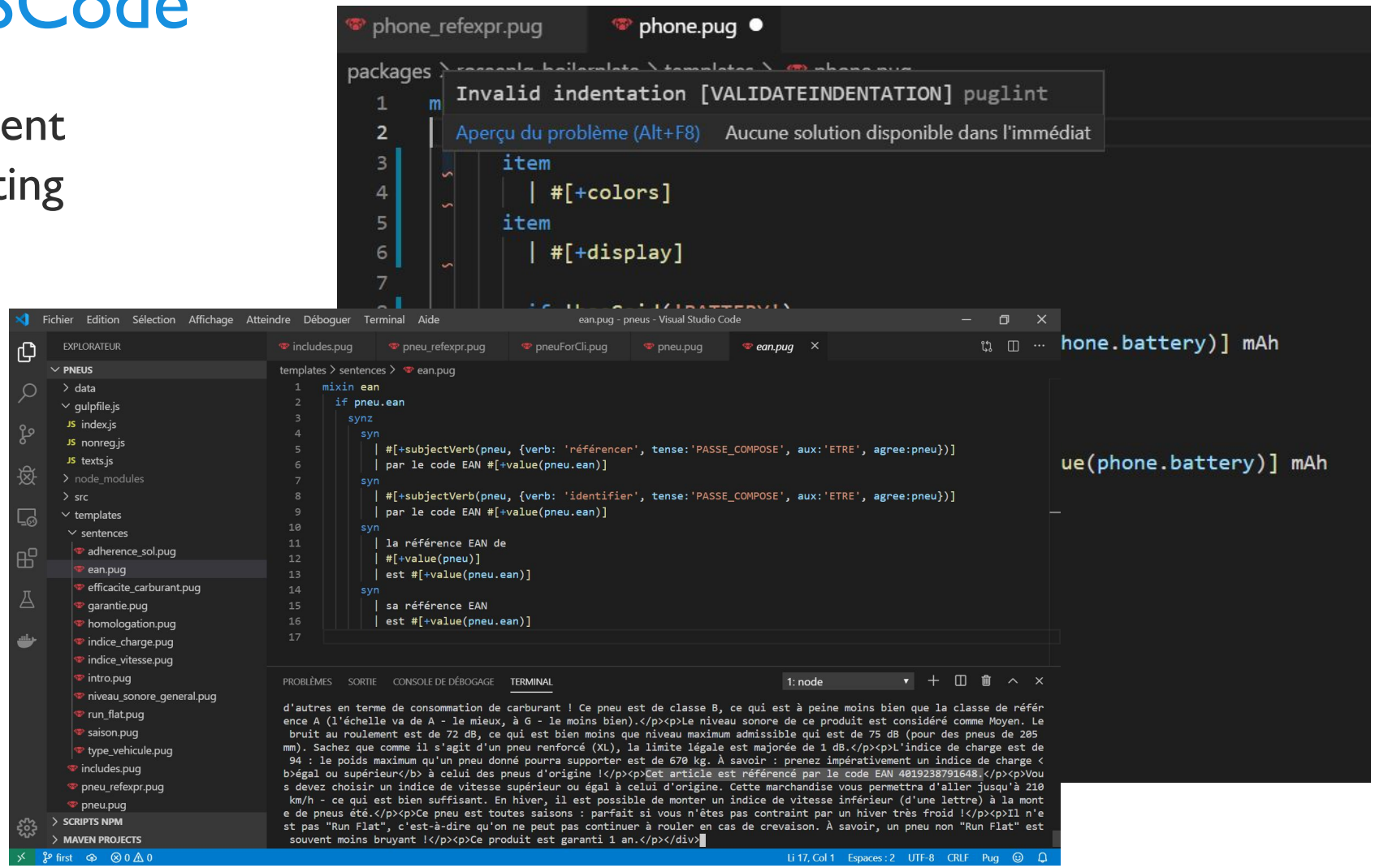
RosaeNLG is an open source NLG project with main developer Ludan Stoecklé (13 years in NLG):

- Designed to be **developer and IT friendly**: templates development on VSCode with a friendly syntax, easy to integrate
- Supports multiple languages with linguistic resources: currently English, French, German, Italian and Spanish
- Provides NLG on **both server-side** (using node.js REST API) **and browser-side**
- Extensive documentation
- and an awesome logo!

RosaeNLG.org

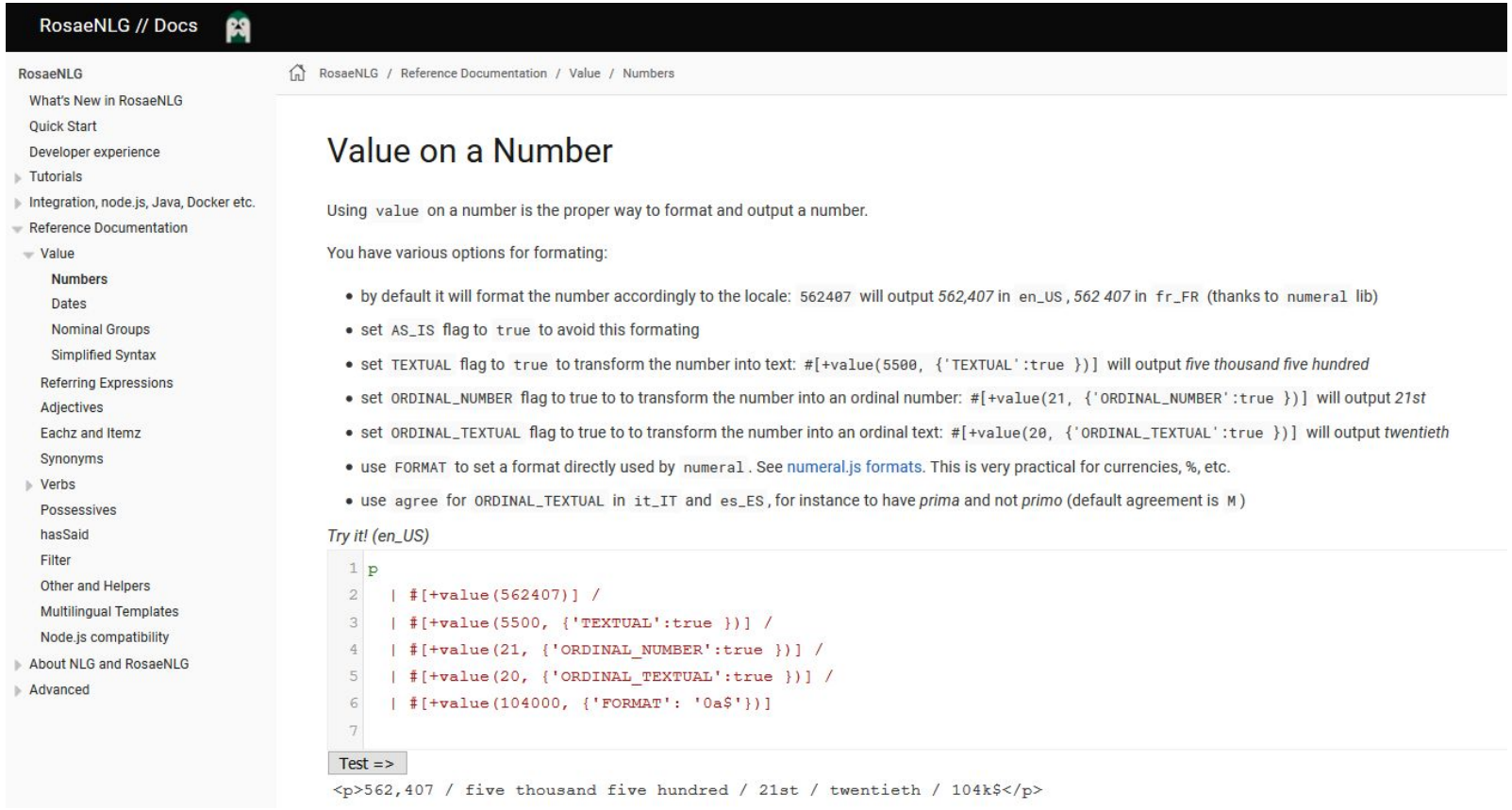
# Develop using VSCode

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



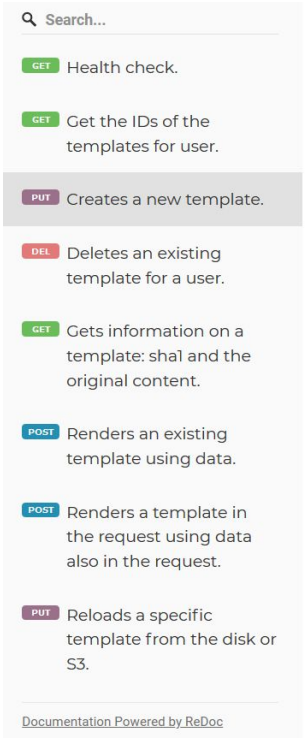
# Reference Documentation

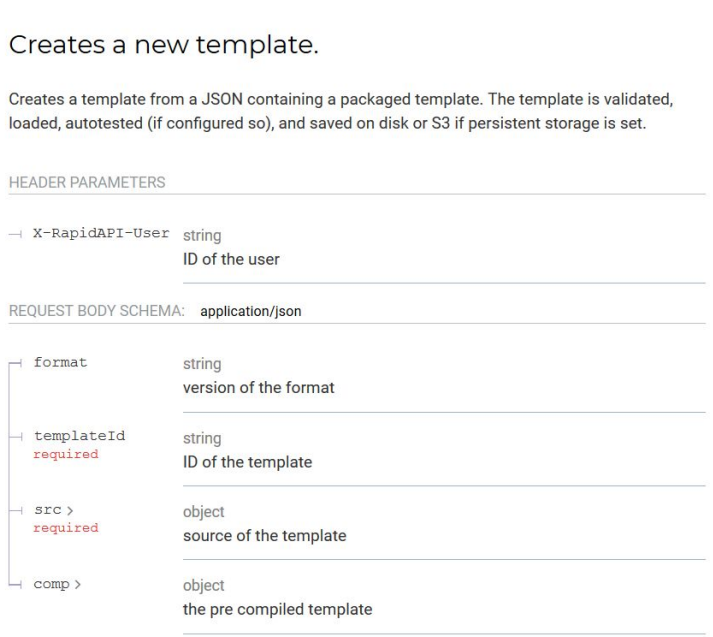
- extensive
- versioned
- search engine
- test snippets of code directly in the browser

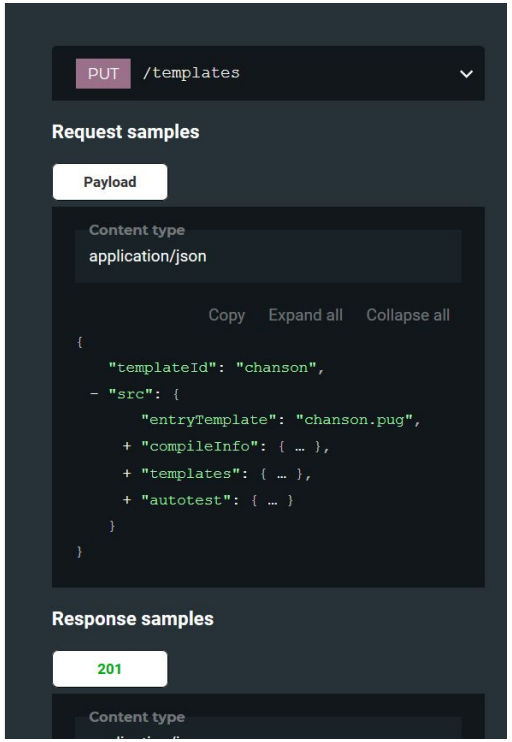


# RosaeNLG node.js API

- 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.

Potential usage of MLflow, especially Model Registry, to manage templates lifecycle.

# 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 for POC 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
- Have a strong community of users and contributors
- **Support more than 50 commonly spoken languages**
- In the long term: standardize NLG templating language (whatever the underlying implementation)

## Roadmap

- **More languages:** Arabic, Chinese, Indian languages, Finnish, etc. - depending on contributors
- **Dedicated VSCode plugin**, with syntax highlight, linter, template debug support
- Collaborate with current LF AI projects (AI Explainability 360 & AI Fairness 360)
- Create a real **Power BI plugin**
- NLG library to ease number analysis
- Onboard contributors like Redlab Paris

# Alignment with LF AI's mission

NLG is a brick to **build business services directly aimed at final users**. It is widely used in the industry.

NLG concurs to **democratization and understandability of AI**:

- Non-expert users don't understand figures and dashboards and **prefer textual explanations**
- Computer-generated texts can be superior (from the reader's perspective) to human-written texts
- At the end of an AI pipeline, to **automate and convey expertise**, **explain and summarize situations**, and communicate with end users

# Why sandbox RosaeNLG in the LF AI?

- Allow to **build business services** directly aimed at users: first service oriented project in the "Natural Language Processing" category of the landscape
- **Address new verticals**: banking, insurance, financial industry who could become LF AI members
- Concur to **democratization, understandability and thrusted AI**, typically in collaboration with **AI Explainability 360** and **AI Fairness 360**
- **Increase diversity:** first **French originated** project
- Foster usage, contributions and diversity in **NLG domain**, supporting languages presently not covered by any NLG system

# Questions

Monthly community meeting every first Thursday of the month.

The time of the meeting changes every month:

- 18:00 CET April / June etc. - which is friendly for Europe and the US.
- 9:00 CET May / July 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 and require linguistic resources.

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 "fruits"* 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     | #[+value('oiseau', {det:'DEFINITE', adj:'blanc'})]
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'})]
```

☑ Render automatically

```
done!
```

Rendered texts: show as html

Ils ont voulu à la fois les belles plages, les beaux cailloux et l'oiseau blanc.

# Example of recent changes

ability to follow the rendering path in the html output

```
templates/phoneForHtml.pug: 3
templates/phoneForHtml.pug: 4
templates/phoneForHtml.pug: 4
templates\phone.pug: 17
templates\phone.pug: 20
templates\phone.pug: 20
templates\sentences\intro.pug: 2
templates\sentences\intro.pug: 6
```

I really love the new

```
templates\sentences\intro.pug: 6
templates\phone_refexpr.pug: 2
```

OnePlus 5T

```
templates\sentences\intro.pug: 6

.
templates\phone.pug: 20
templates\phone.pug: 21
templates\phone.pug: 21
templates\phone.pug: 21
templates\phone.pug: 2
templates\phone.pug: 12
templates\phone.pug: 13
templates\phone.pug: 13
templates\phone_refexpr.pug: 5
templates\phone_refexpr.pug: 7
```

The phone

```
templates\phone.pug: 13
```

has a battery of

```
templates\phone.pug: 13
```

3,300

```
templates\phone.pug: 13
```

mAh

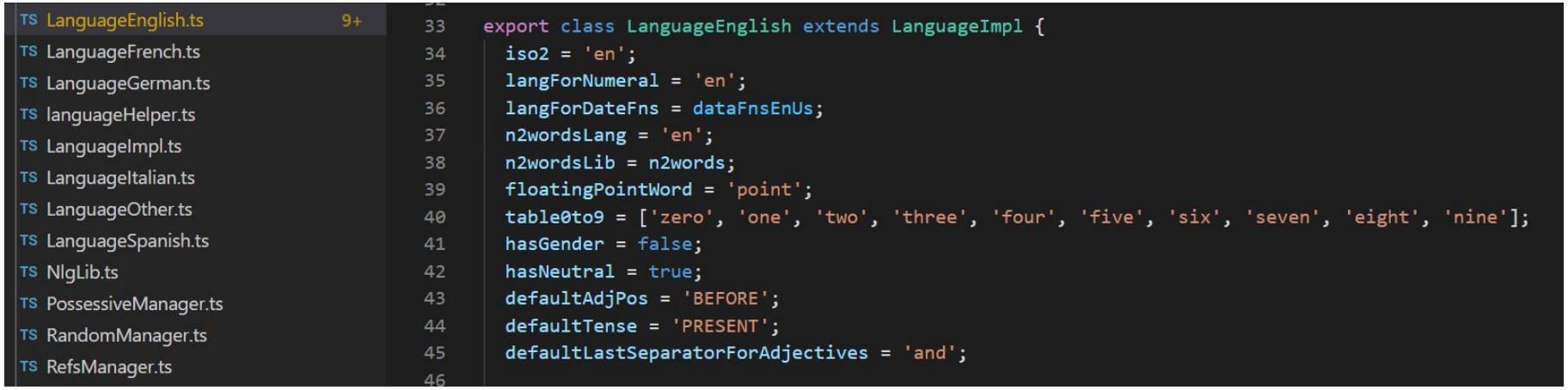
```
templates\phone.pug: 14

.
```

# Example of recent changes

new organisation of language specific code:

- before: switch/case per language everywhere, which is not scalable
- now: classes for each language
- will ease the addition of new languages and keep language specific packages small
- developer doc to add new languages



# TAC Vote on Project Proposal: RosaeNLG

**Proposed Resolution:**

The TAC approves the RosaeNLG as a Sandbox project of the LF AI & Data Foundation

# Next Steps

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

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

Integrate the sandbox project with LF AI & Data operations

# Invited Presentation:
# Elyra/Jupyter Ecosystem

Luciano Resende <lresende@us.ibm.com>

# Elyra - Extending JupyterLab for AI

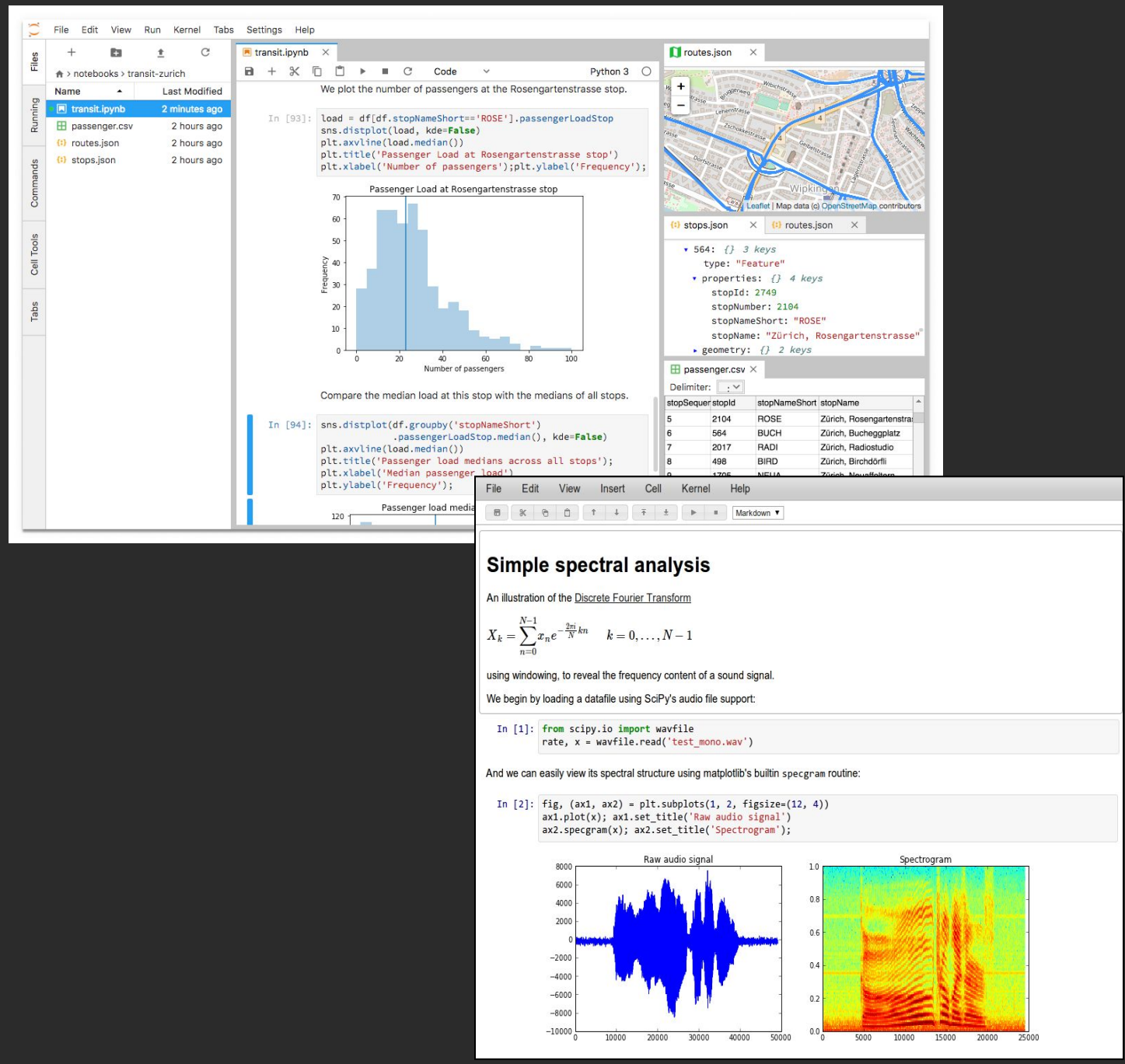
Luciano Resende
IBM - CODAIT



IBM

# Jupyter Notebooks

Notebooks are interactive computational environments, in which you can combine code execution, rich text, mathematics, plots and rich media.



# Languages, tools & Industry trends


**Simple, but Powerful**

As simple as opening a web page, with the capabilities of a powerful, multilingual, development environment.

**Interactive widgets**

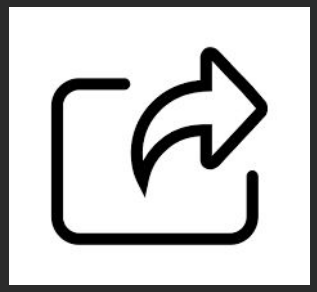
Code can produce rich outputs such as images, videos, markdown, LaTeX and JavaScript. Interactive widgets can be used to manipulate and visualize data in real-time.

**Language of choice**

Jupyter Notebooks have support for over 50 programming languages, including those popular in Data Science, Data Engineer, and AI such as Python, R, Julia and Scala.


**Big Data Integration**

Leverage Big Data platforms such as Apache Spark from Python, R and Scala.  Explore the same data with pandas, scikit-learn, ggplot2, dplyr, etc.

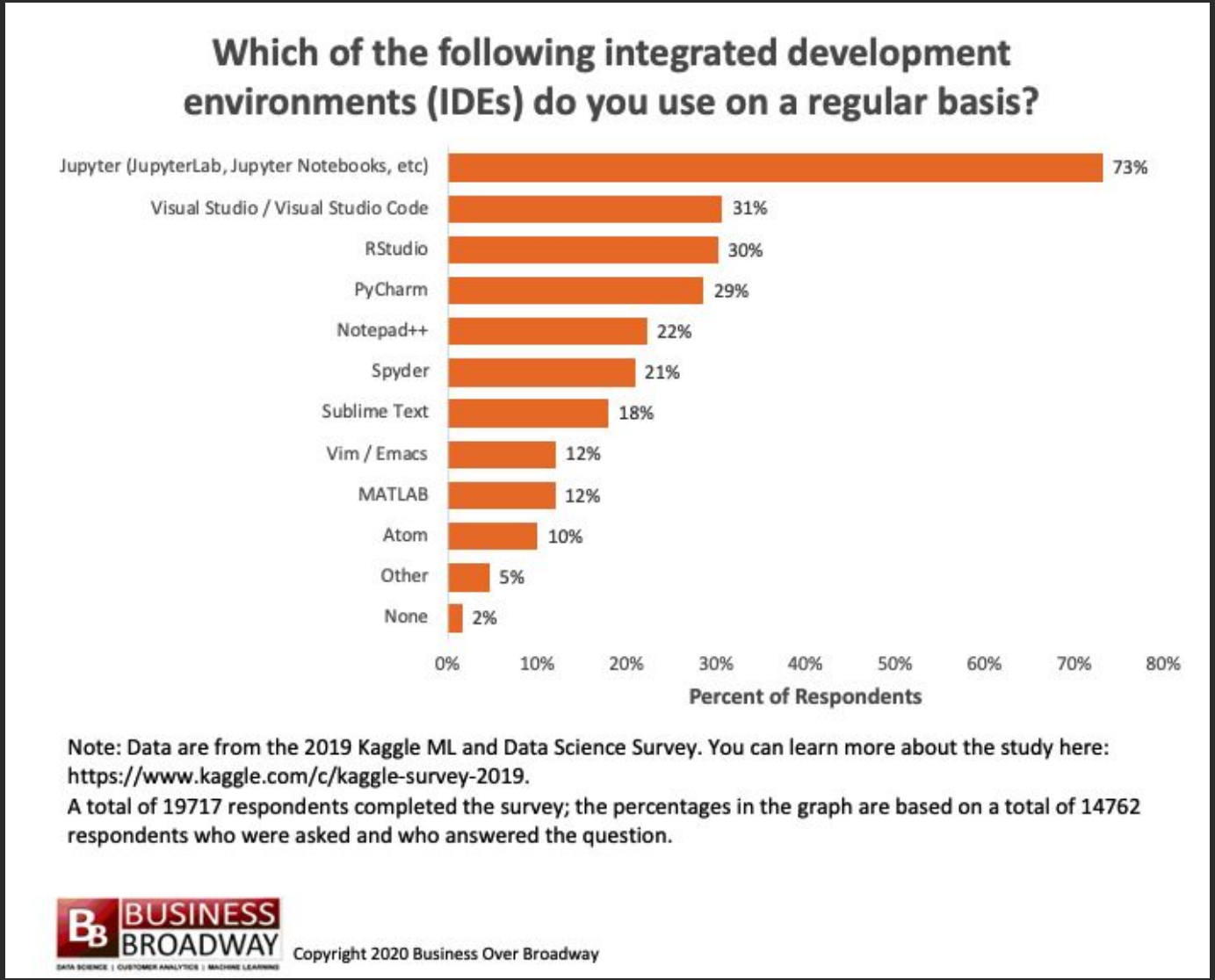


**Share Notebooks**

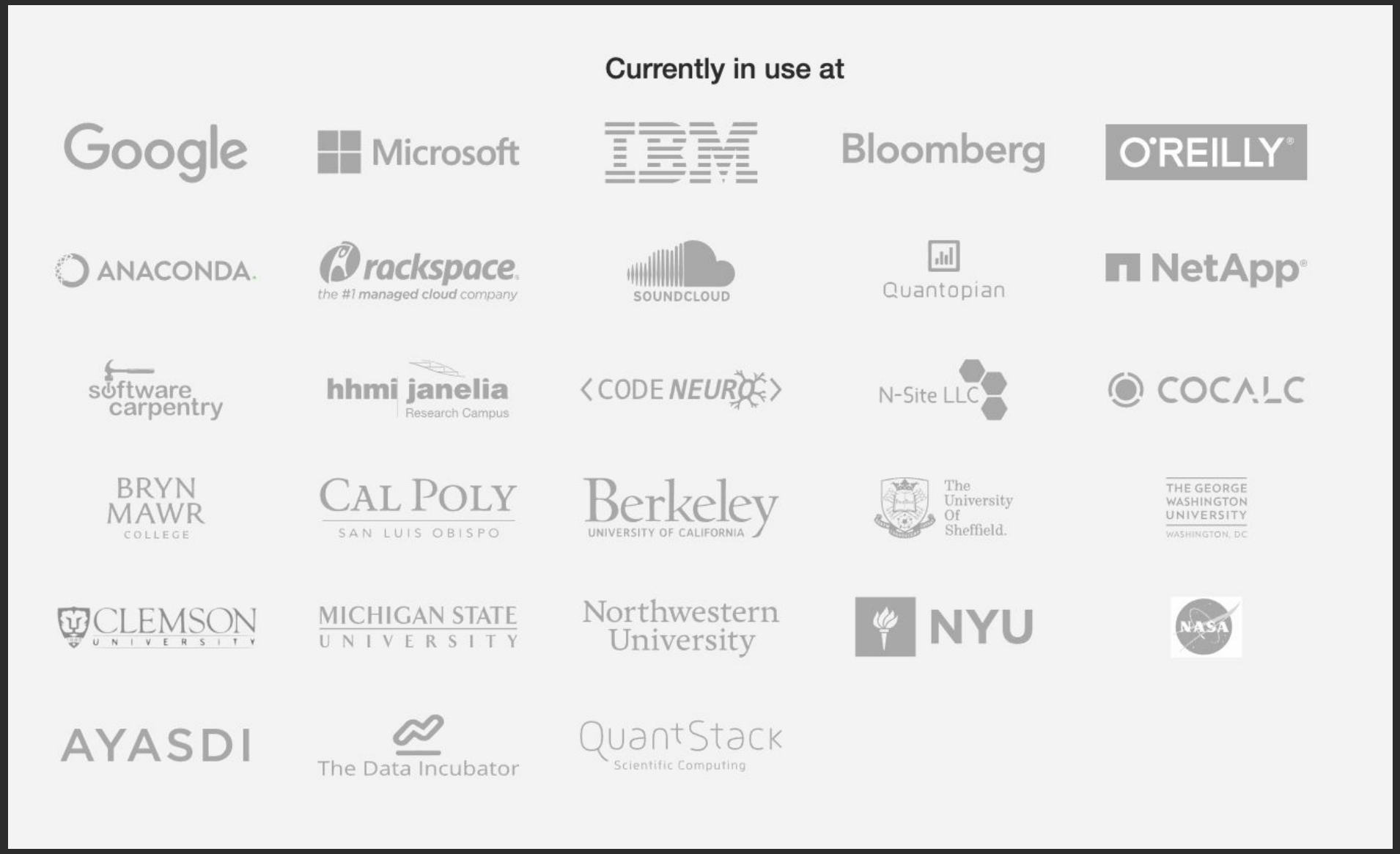
Notebooks can be shared with others using e-mail, Dropbox, Google Drive, GitHub, etc

# Jupyter Marketshare

Jupyter Ecosystem is de-facto standard tool in data science and AI



# Jupyter in the Industry



# ELYRA

## AI Pipelines editor

Elyra provides a visual editor for building AI pipelines, enabling the conversion of multiple notebooks and/or Python scripts into batch jobs or workflows.

## Notebook as batch job

Elyra extends the notebook UI to simplify the submission of notebooks as a batch job for model training

## Python script execution

Exposes Python Scripts as first-class citizens allowing users to edit their scripts and execute them against local or cloud-based resources seamlessly.

## Reusable Code Snippets

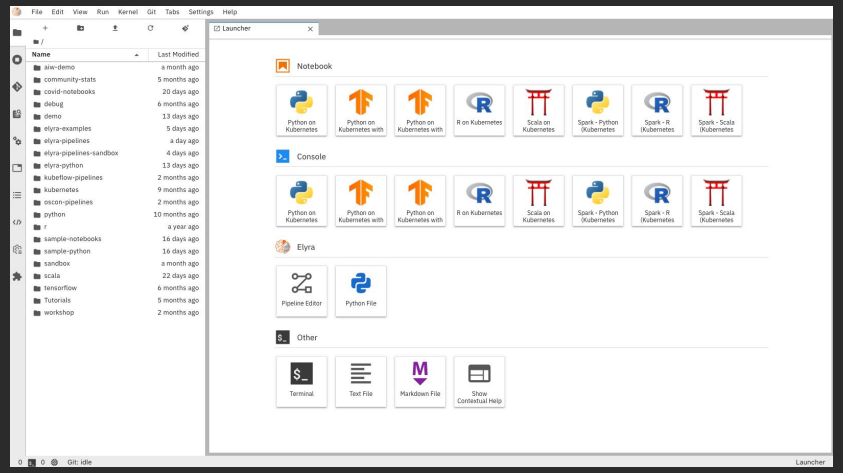
Code Snippets enables easy manipulation of reusable code snippets for various programming languages

## Versioning using git

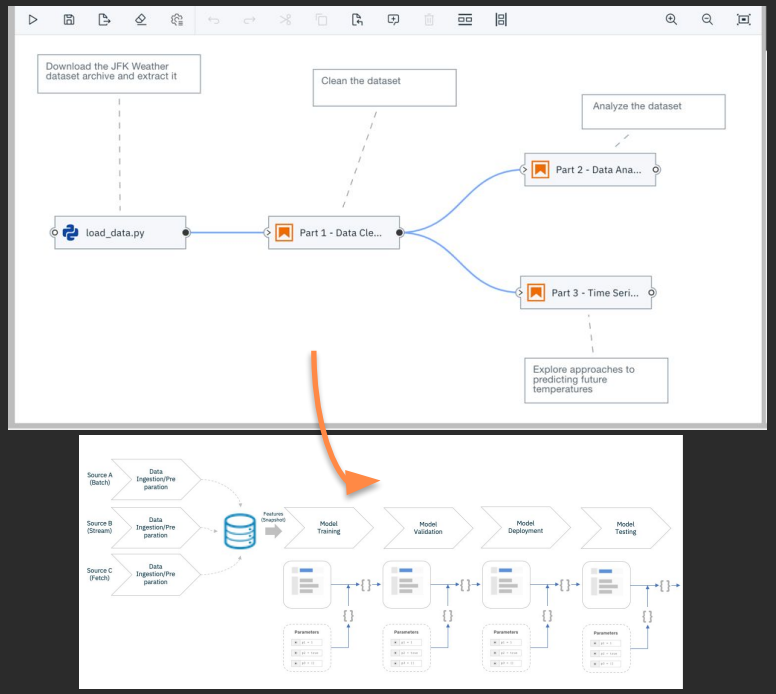
Simplify tracking changes, sharing contents among teammates

IBM **Developer**

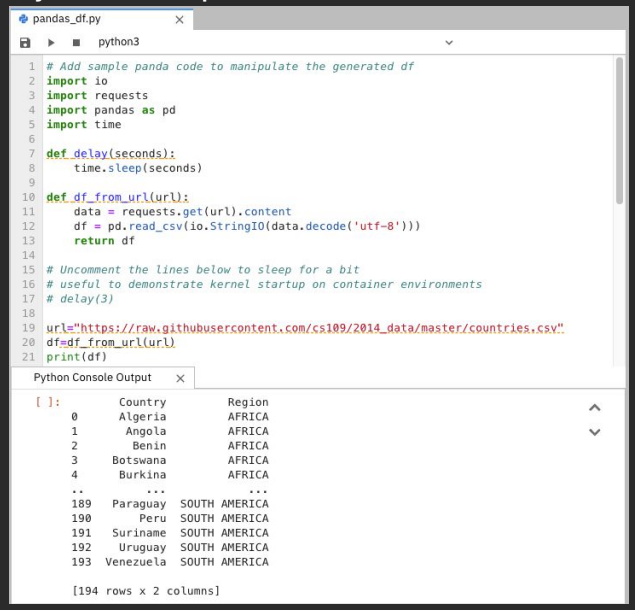
## JupyterLab Extensions



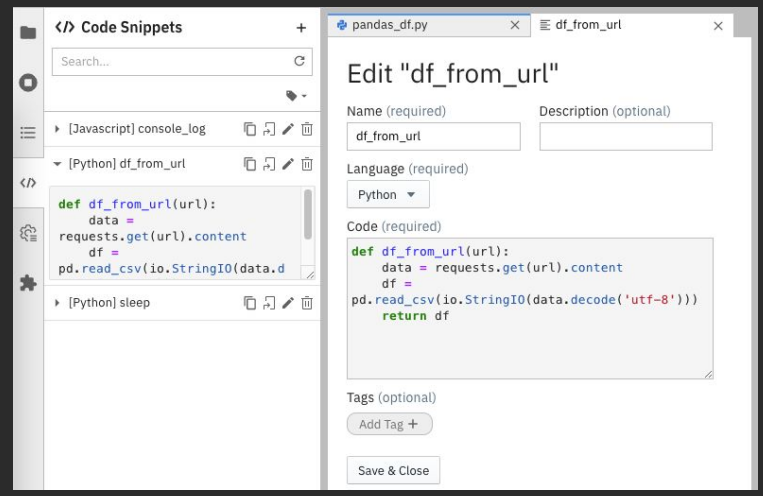

## AI Pipelines Editor




## Python Script Editor




## Code Snippets




ELYRA GITHUB github.com/elyra-ai
ELYRA DOCS elyra.readthedocs.io
GITTER gitter.im/elyra-ai/community
JUPYTERLAB ORG github.com/jupyterlab

# ELYRA

Elyra 2.0 announced today ->
https://developer.ibm.com/articles/what-is-new-in-elyra-2-0/

## JupyterLab 3.0

Enable support for JupyterLab 3.0 bringing access to all new capabilities such as Debugger, Enhanced TOC, Simple Interface mode, Globalization, etc

## Language Server Protocol (LSP)

The JupyterLab Language Server Protocol (LSP) integration enhances the development experience in the notebook editor and file editors, delivering features common in IDEs such as autocompletion, code navigation, hover suggestions, code linting, and renaming.

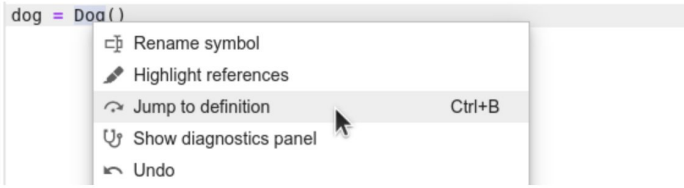
## KFP with Argo and/or Tekton

Enable pipelines to run on KFP using Argo or Tekton

IBM **Developer**

### Jump to Definition

Use the context menu entry, or `Alt` + to jump to definitions (you can change it to `Ctrl`/`⌘` in settings); use `Alt` + `O` to jump back

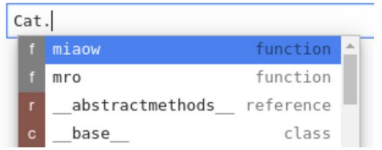



### Highlight References

Place your cursor on a variable, function, etc and all the usages will be highlighted

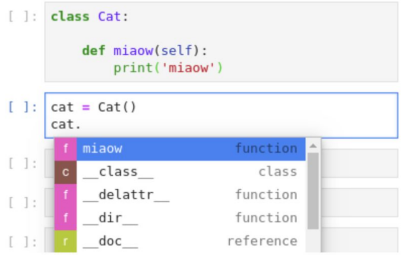
### Automatic Completion

Certain characters, for example '.' (dot) in Python, will automatically trigger completion




### Kernel-less Autocompletion

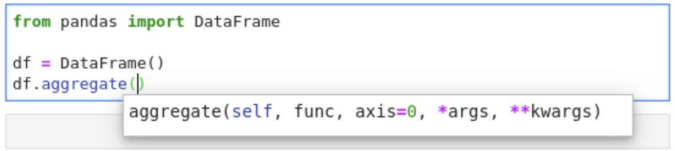
Advanced static-analysis autocompletion without a running kernel




When a kernel is available the suggestions from the kernel (such autocompletion) are merged with the suggestions from the Lang

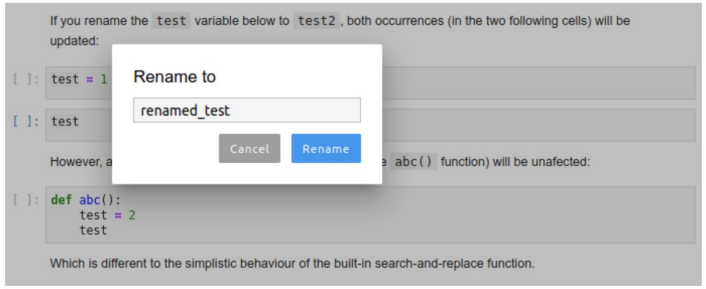
### Automatic Signature Suggestions

Function signatures will automatically be displayed




### Rename

Rename variables, functions and more, in both: notebooks and the file editor. Use the context menu option or the `F2` shortcut to invoke.

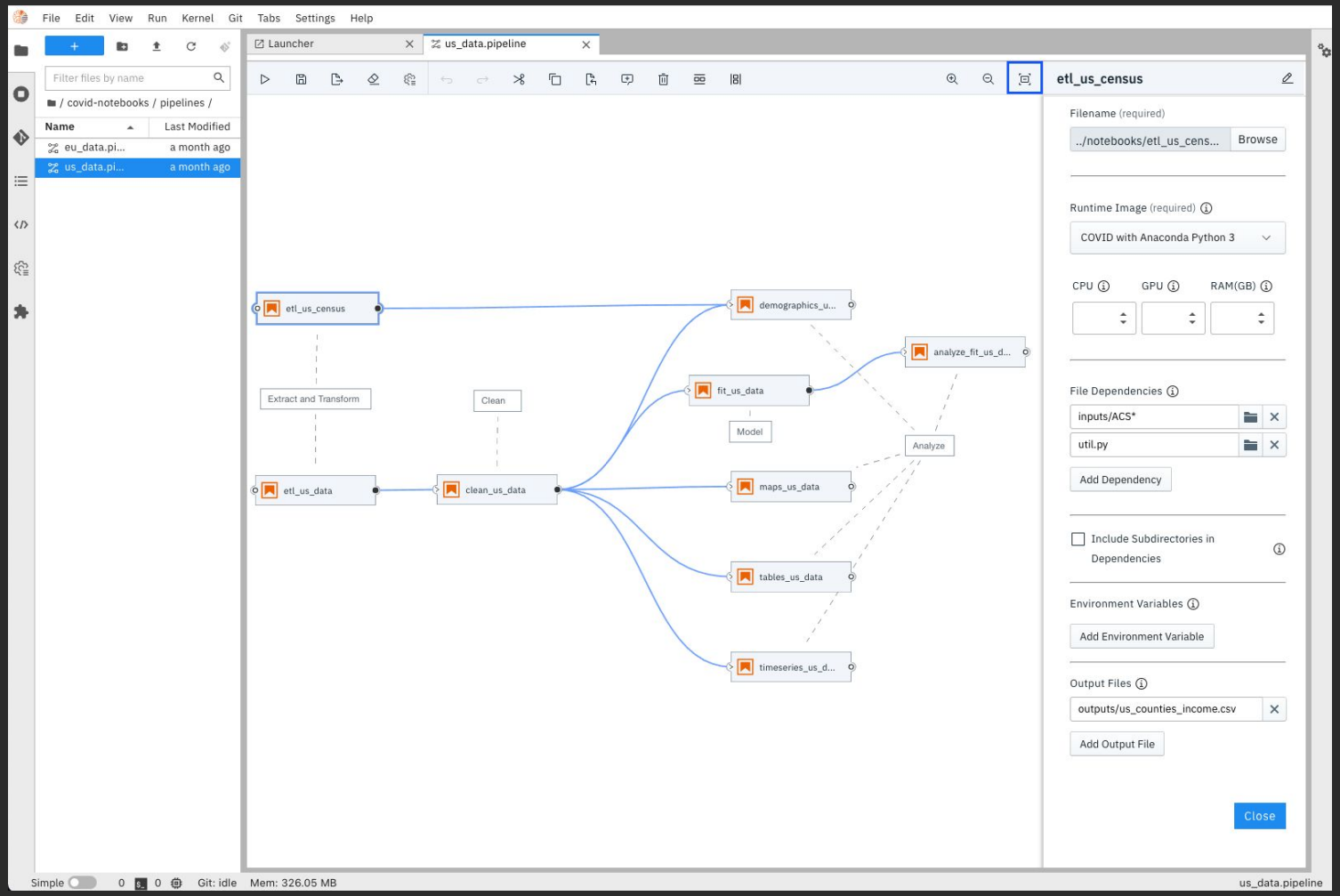


# COVID Notebook Scenario

An AI Pipeline that consists of multiple notebooks helping data scientists analyze real-time COVID-19 data.

Originally developed using JupyterLab notebooks and orchestrated with AI Pipelines created with the Elyra pipeline editor extension for JupyterLab.

Shared openly in GitHub to foster reproducibility and collaboration.




https://github.com/CODAIT/covid-notebooks

https://developer.ibm.com/blogs/open-source-jupyter-notebooks-analyze-covid-19-data/

IBM **Developer**

# Other resources

STAR US & FORK US ON GITHUB

## Elyra source code at GitHub

https://www.youtube.com/watch?v=PuGNijkV5PQ

## Elyra source code at GitHub

https://github.com/elyra-ai/elyra

## Elyra Documentation

https://elyra.readthedocs.io/en/latest/

## Elyra community chat on Gitter

https://gitter.im/elyra-ai/community

ELYRA

# LF AI & Data - General Updates

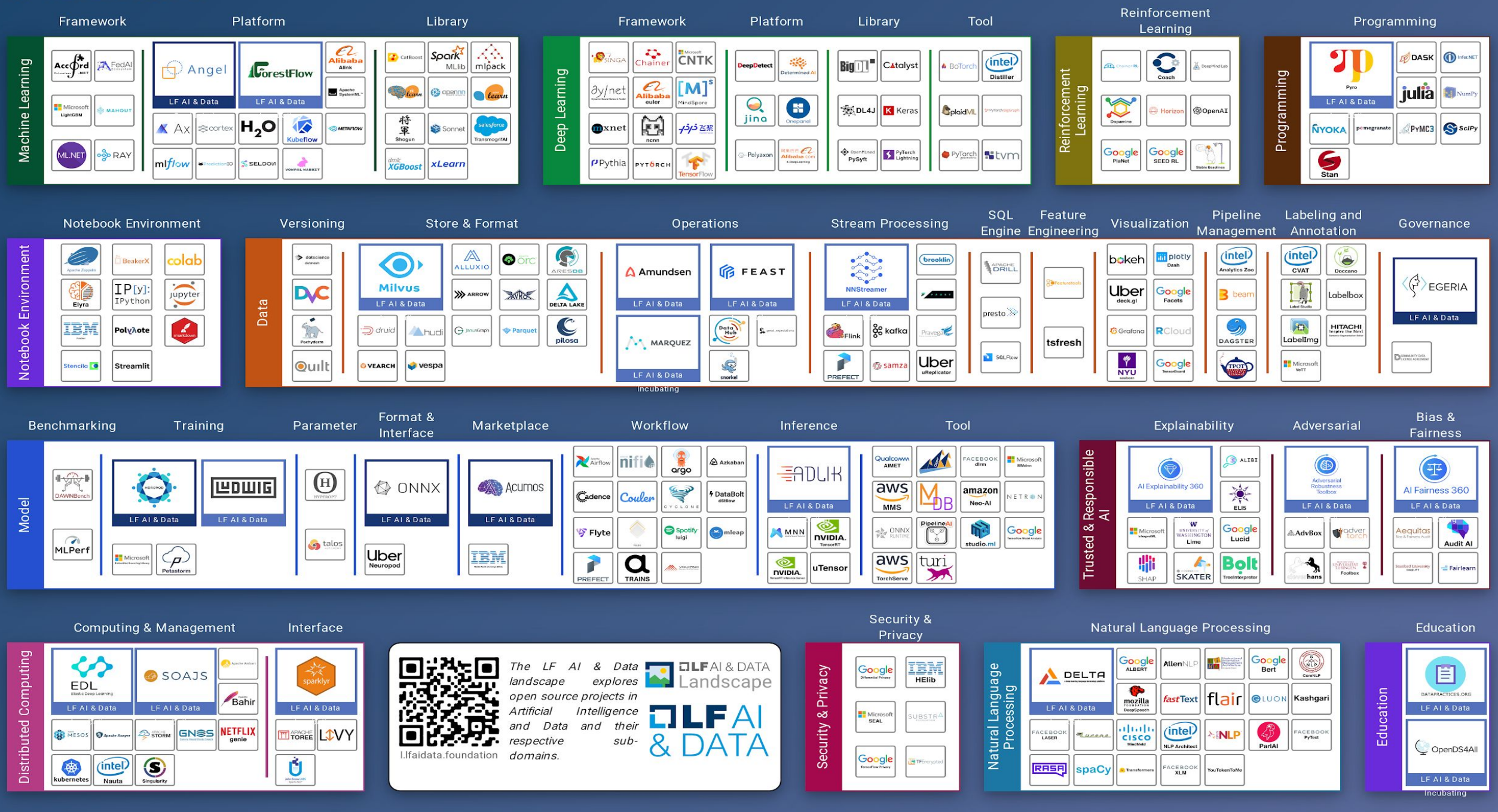
LF AI & DATA

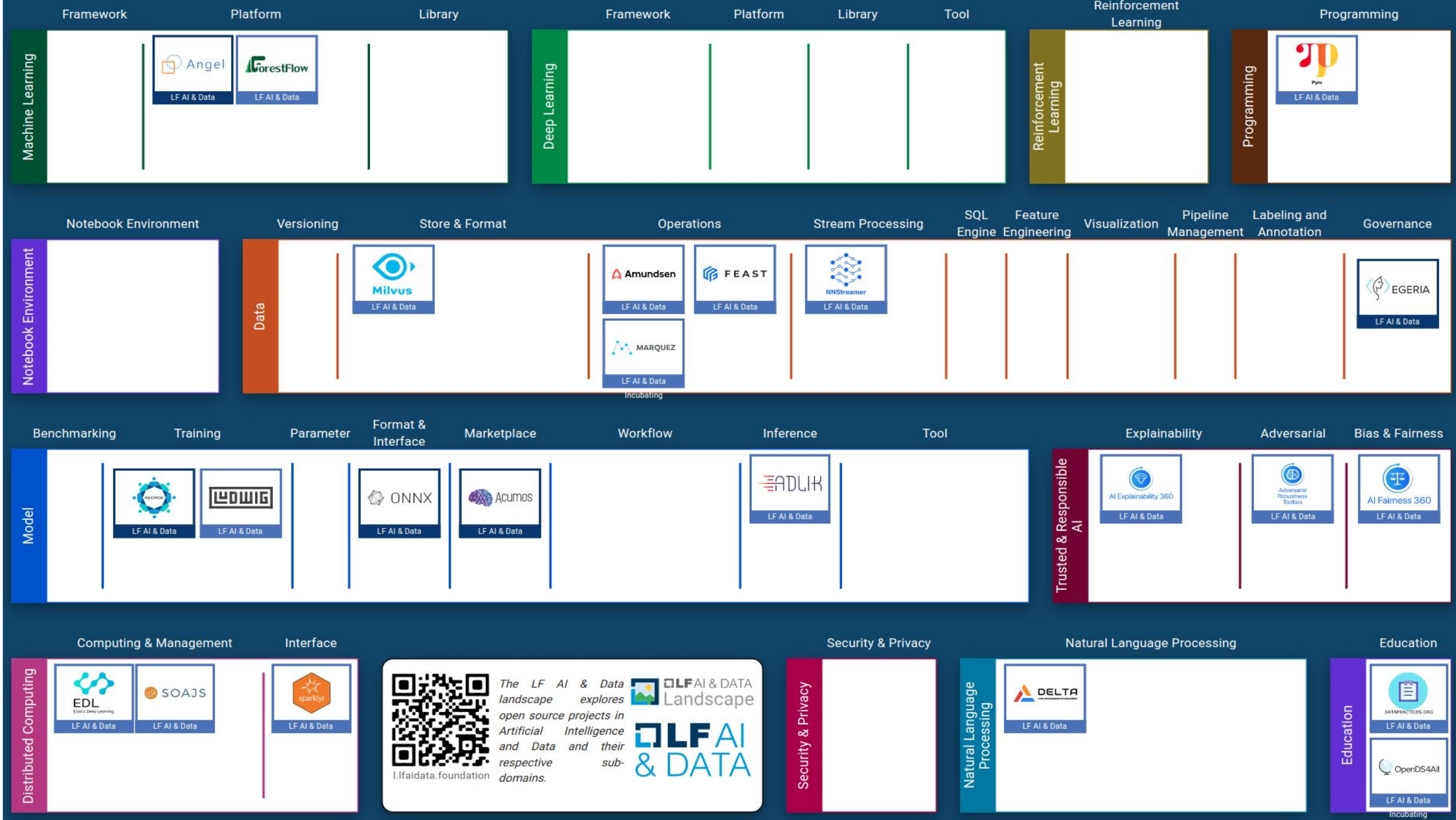
# Linux Foundation AI Landscape

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 | 14: Data Lifecyle Framework (IBM)*<br>28: Tentative: Verse (Seldon) | 11: MARS (Aliabab)<br>25: Flyte (Lyft) | 11: Streams (IBM)<br>25: Tentative: Substra Framework |
| Apr<br>May<br>Jun | 8: Adlik (ZTE)**<br>22: Kubeflow-On-Prem (Google, Arrikto, Intel) | ?: Ray (Anyscale.io)<br>?: Pachyderm (Pachyderm)<br>?: DataHub (LinkedIn) | ?: Common Knowledge (Code Reef)<br>?: Couler (Ant Financial) |
| 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) |

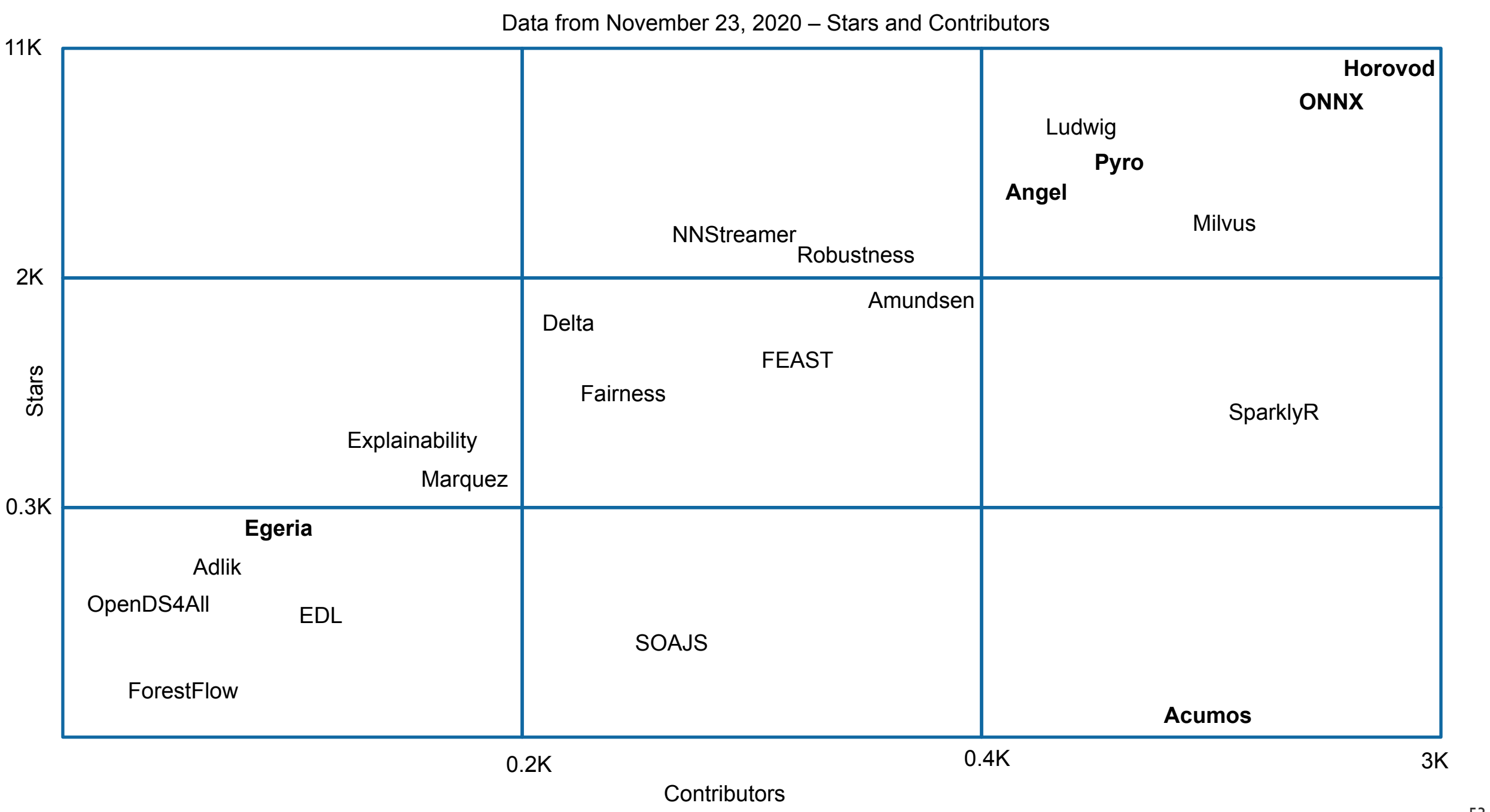
(Entity)* = incubating vote

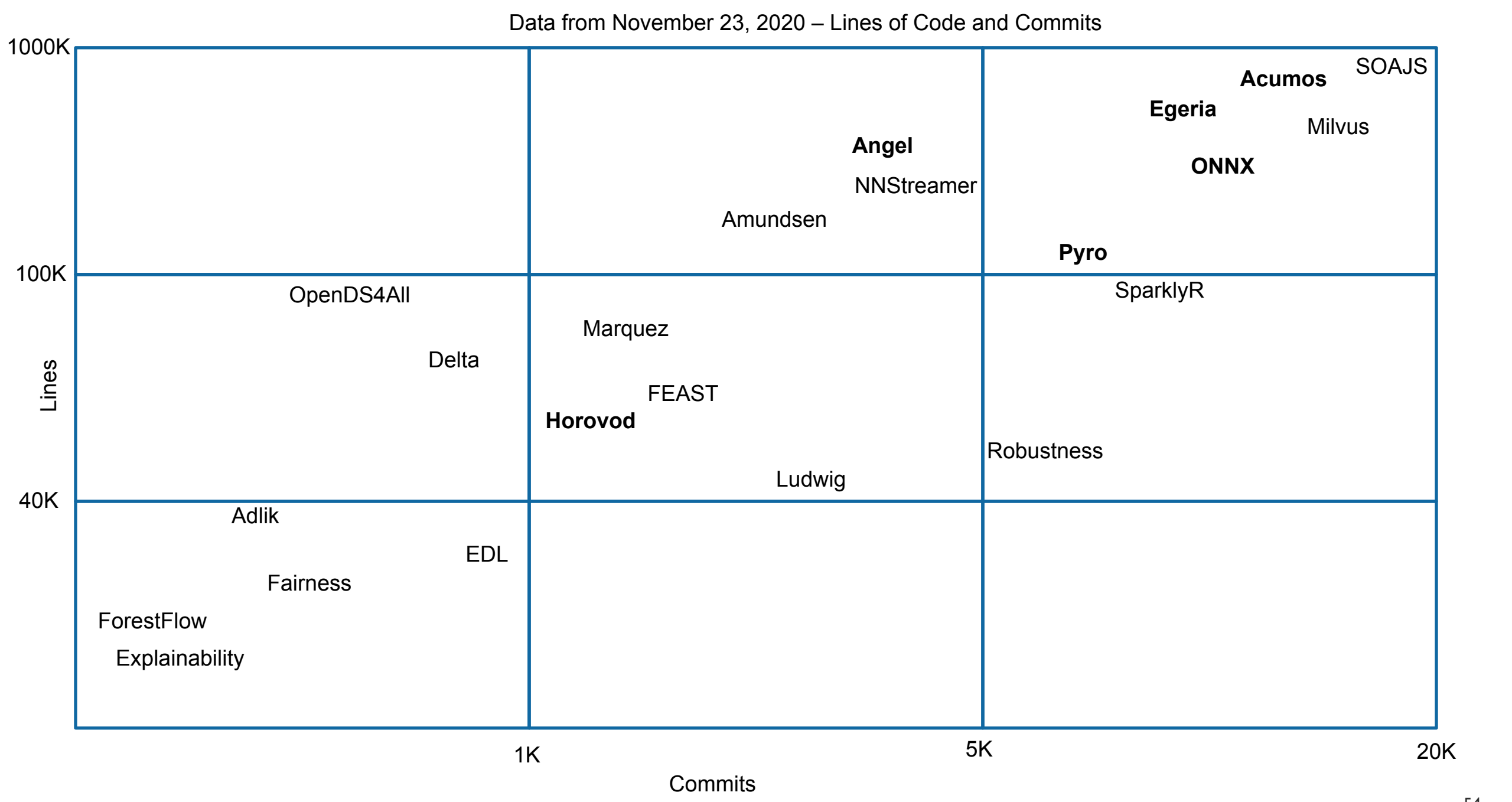
** **bold = graduate vote**

*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
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

# 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/

- **March 24, 2021 - ONNX Community Virtual Meetup**
  a. **Wednesday @ 5:00 pm - 8:00 pm PT USA**
     **Thursday @ 8:00am - 11am China Time**
     **LF AI Day: ONNX Community Virtual Meetup – March 2021**
     **(Virtual - Free - Asia-friendly time – Host Ti Zhou - Baidu)**

- **Sept 29 - Oct 1, 2021 - OSS Global**
  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)

›

› Mar 11: Sandbox project proposal - RosaeNLG

› Mar 25: Substra Foundation

› April 8: Adlik (ZTE)

› April 22: TBD

› May 6: All project updates

›

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

LF AI & DATA