



Georgia Tech Open Source AI TechXchange Dev Day and Meetup

Brad Topol, Ph.D.
IBM Distinguished Engineer, Director of Open Technologies
LF AI & Data Governing Board Chair
Georgia Tech CoC Advisory Board Member

Agenda

The agenda for today

Open source software and AI

The LF AI and Data Foundation

What the Georgia Tech students did at IBM
in 2025

Calls to Action



TechXChange Dev Day in Atlanta - Open Source AI in Atlanta

Georgia Tech | Klaus Building | 266 Ferst Drive | Room 1116

August 21, 10:00am Networking - 10:45am Sessions Start | 5:00pm Sessions End - Networking until 5:45pm



Hosts include: Jeff Young, Cherry Liu, Keerthy Kishore, Kiersten Stokes

Teaching Assistants include: Kiersten Stokes, Benjamin Hoover

10:00am - Networking

10:45am - Kick-Off Georgia Tech – Jeffrey White

11:00am - Kick-Off Open Source AI & TechXChange Dev Day – Brad Topol

11:45am - Lunch



Breaks will be taken as appropriate through the afternoon:

12:30pm - Granite Open Source Model - Hands-On Workshop – BJ Hargrave

2:00pm - Docling for document processing and parsing diverse formats - Hands-On Workshop – Ming Zhao

3:30pm – Generative programming with the Mellea Project - Hands-On Tutorial – Nathan Fulton

4:30pm – Lightning Talks about student projects:

- BeeAI and Docling Root Cause Analysis – Kenneth (Alex) Jenkins | Pytorch Foundation Models - Niraj Kamal | Data Prep Kit UI - Dobromir Iliev | Generics with Granite - Rakshit (Rocket) Nemakallu | A New Era for Associative Memories – Benjamin Hoover

5:00pm Networking

<https://www.ibm.com/events/reg/flow/ibm/9UH8DRMB/landing/page/landing>



Open Source AI in Atlanta – Georgia Tech | Klaus Building | 266 Ferst Drive | Room 1116
August 21, 5:00pm Networking | 5:45pm Talks Start | 7:20pm Talks End
Networking with Pizza until 8:00pm

Hosts include: Jeff Young, Cherry Liu, Keerthy Kishore, Kiersten Stokes

5:00pm Networking

5:45pm Opening Remarks – followed by Lightning Talks

- 5:50pm Georgia Tech AI Activities
- 6:00pm Granite open source AI model – BJ Hargrave
- 6:10pm Docling for document processing, parsing diverse formats for seamless integrations with gen AI ecosystem – Ming Zhao
- 6:20 Generative programming with the Mellea Project – Nathan Fulton
- 6:30 Responsible Technology By Design – Neera Mathur
- 6:40 Agentic AI with BeeAI - Madeline Lutwyche, AJETHAA Arumugam MURUGESAPANDIAN
- 6:50 Speaker Panel for Q&A

7:20pm Networking

8:00pm Close



Register: <https://lu.ma/zmfligoi>



Technology leaders do more than just consume OSS

Long IBM history of actively fostering balanced community participation



“For more than 20 years, IBM and Red Hat have paved the way for open communities to power innovative IT solutions.”
– Red Hat

IBM Seeks to Harmonize Open Source Consumption & Contribution

The Harmonic Cycle – Meaningful Contributions foster community trust, idea generation and encourages partner Contributions. This leads to continuous code improvement & feature acceleration which can be Consumed back by the enterprise.

Consume Internally

Today, many IBM solutions are built on Open Source software, some as much as 75%



Contribute Externally

IBM strives to be a “good neighbor”, helping where needed and delivering innovation back to the community (Over 4500 Active Contributors and 600 Maintainers)

Why does open source matter to software developers?

Open Source = Skills, Support, and Stature

1

Easy access to key open technologies accelerates the speed at which developers can do their jobs and learn from others

2

The open source community helps developers grow through an ecosystem of support

3

Developers get the chance to “show off” their skills and build their eminence. Open source communities are the premiere social communities for developers

Open source is free – as in puppies



Open source is free, free as in puppies. You might get a free puppy from the shelter, but it is now your responsibility.

Similarly, with open source, you need to:

- ❖ Ensure currency with upstream
- ❖ Remediate defects and vulnerabilities as they are discovered
- ❖ Beware of creating your own fork
- ❖ Ideally, contribute fixes and improvements upstream

Open is great but, there's risk – *choose wisely*



Is the project under open governance?

Is the license really an open source license?

Does the code even **have** a license?

Check for OpenSSF Best Practices Badge ([link](#))

Check the OpenSSF Audit Program if the project is under LF governance ([link](#))

Vet the project using a tool such as OpenSSF's [scorecard](#)

Are there languishing PRs or Issues?

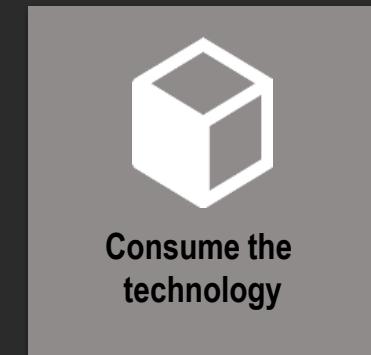
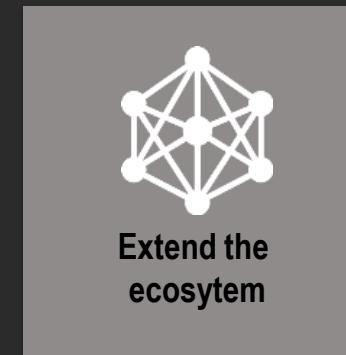
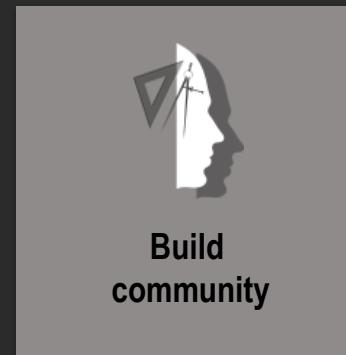
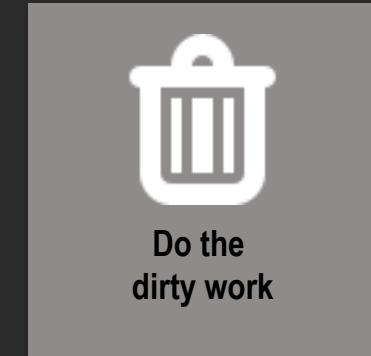
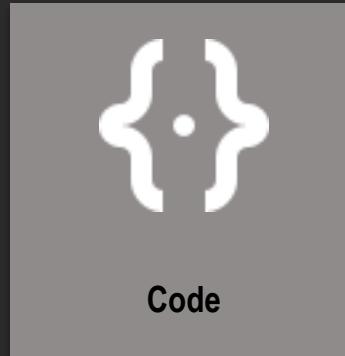
Does it have regular releases?

Does it have a large ecosystem of usage?

Does it have multiple vendors contributing to it?

How We Teach IBMers to Behave in Open Source Communities

- Be nice and courteous
- Don't be a talker – be a doer and good observer first
- Don't flaunt your title or education
- Get a mentor in the community. Let them put you to work
- Don't be a drive-by committer
- Start small – build trust



Some Key Open Source AI Projects

Recent projects:

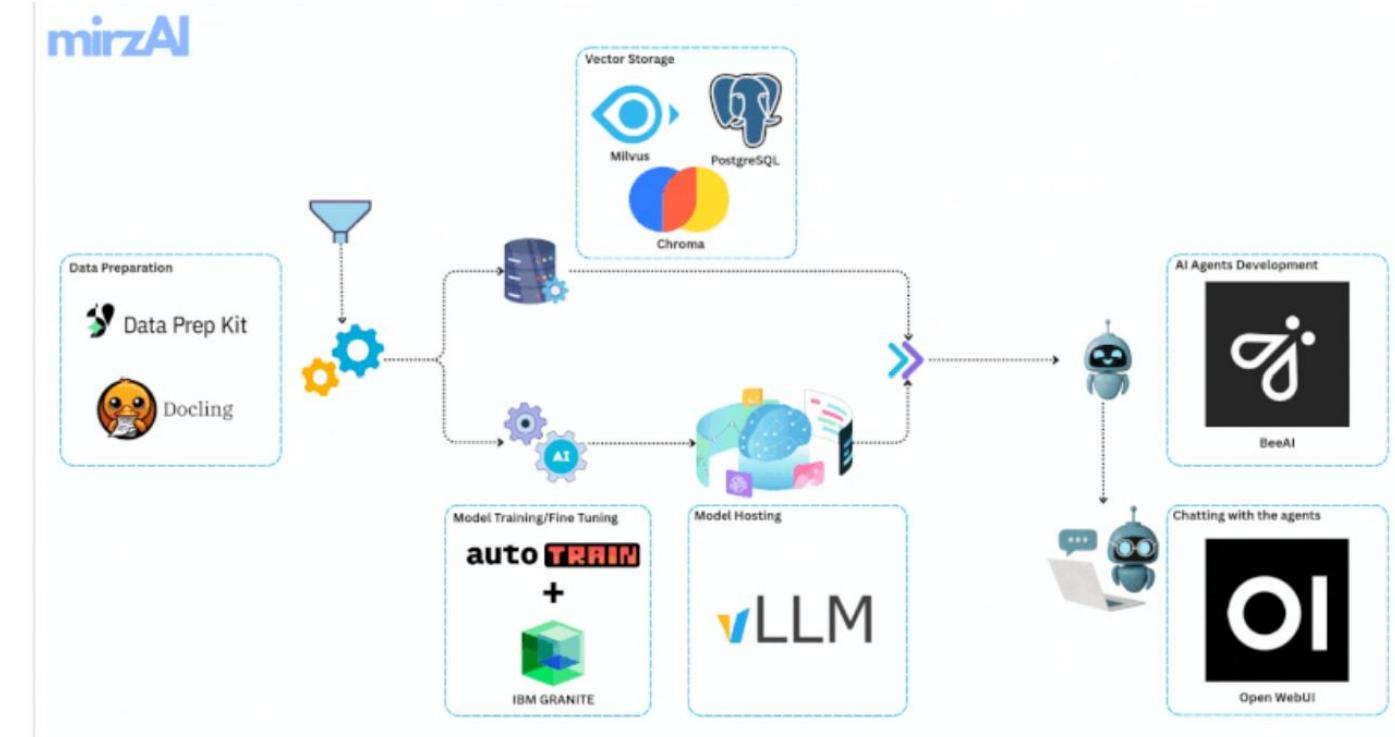
- Granite
- Mellea

At the PyTorch Foundation:

- PyTorch
- vLLM

At the LFAI and Data Foundation:

- Docling
- Data Prep Kit
- BeeAI



What is LF AI & Data Foundation?

LF AI & Data is a global not-for-profit foundation that **hosts critical components of the global AI & Data technology infrastructure**. It brings together the world's top developers, end users, and vendors to **identify and contribute to the projects and initiatives that address industry challenges** for the benefit of all participants.

Evolution & Mission



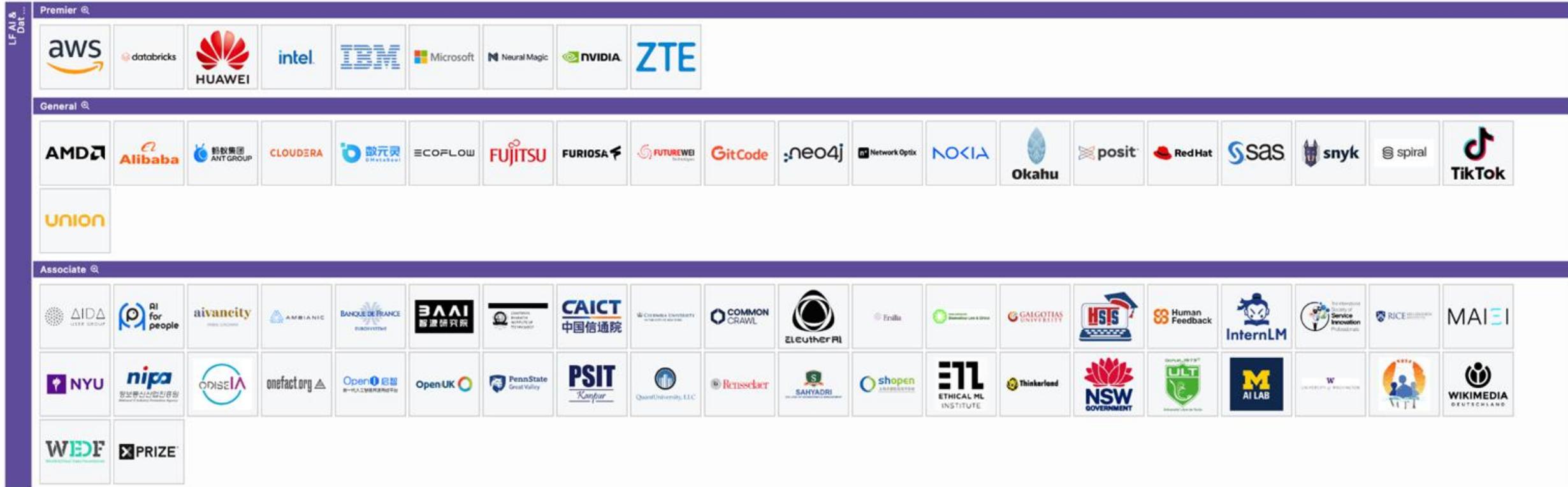
The mission of the LFAI & Data Foundation is to advance two critical pillars of modern technology: **accelerating the creation and deployment of open source-based AI applications and AI Agent frameworks**, while simultaneously **establishing data infrastructure standards**. As part of this mission, LFAI & Data Foundation also focuses on delivering common model formats, and standard protocols to empower organizations to **build and run scalable and efficient AI applications**.

LF AI & Data is a neutral facilitator that brings experience and credibility to its projects and initiatives.

We host critical assets for AI & Data technology infrastructure and bring together the world's top developers, organizations, end users, and vendors to identify and contribute to the projects and initiatives that address industry challenges for the benefit of all participants.

Members

July 2025



LF AI & Data - Top Ten Most Active Projects



ONNX

Created by Facebook and Microsoft, **ONNX** is an open format for representing deep learning models, allowing AI developers to easily move models between tools and choose the best combination for their needs.

GITHUB STARS: 43,575 - NEW STARS IN PAST YEAR: 4,623
TOTAL NUMBER OF CONTRIBUTORS: 8152



Milvus

Milvus is an open-source vector database built for GenAI applications. Install with pip, perform high-speed searches, and scale to tens of billions of vectors with minimal performance loss.

GITHUB STARS: 36,402 - NEW STARS IN PAST YEAR: 7,682
TOTAL NUMBER OF CONTRIBUTORS: 1,271



DELTA LAKE

Created and contributed by Databricks, **Delta Lake** is an open-source storage framework that enables building a format agnostic Lakehouse architecture with compute engines including Spark, PrestoDB, Flink, Trino, Hive, Snowflake, Google BigQuery, Athena, Redshift, Databricks, Azure Fabric and APIs for Scala, Java, Rust, and Python.

GITHUB STARS: 13,206 - NEW STARS IN PAST YEAR: 2,092
TOTAL NUMBER OF CONTRIBUTORS: 1,115



kedro

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

GITHUB STARS: 11,842 - NEW STARS IN PAST YEAR: 1,779
TOTAL NUMBER OF CONTRIBUTORS: 1,713



docling

Created and donated to LF AI by IBM, Docling is a powerful library which simplifies document processing, parsing diverse formats - including advanced PDF understanding - and providing seamless integrations with the gen AI ecosystem.

GITHUB STARS: 8,841 - NEW STARS IN PAST YEAR: 8,841
TOTAL NUMBER OF CONTRIBUTORS: 894



Flyte

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

GITHUB STARS: 7,399 - NEW STARS IN PAST YEAR: 1,224
TOTAL NUMBER OF CONTRIBUTORS: 520



Adversarial Robustness Toolbox

Adversarial Robustness Toolbox provides a set of tools that enable developers and researchers to evaluate, defend, certify and verify Machine Learning models and applications against the adversarial threats.

GITHUB STARS: 5,556 - NEW STARS IN PAST YEAR: 898
TOTAL NUMBER OF CONTRIBUTORS: 71



KServe

KServe provides a Kubernetes Custom Resource Definition for serving machine learning (ML) models on arbitrary frameworks. It aims to solve production model serving use cases by providing performant, high abstraction interfaces for common ML frameworks like Tensorflow, XGBoost, ScikitLearn, PyTorch, and ONNX

GITHUB STARS: 5,012 - NEW STARS IN PAST YEAR: 1,112
TOTAL NUMBER OF CONTRIBUTORS: 500



IREE

IREE (Intermediate Representation Execution Environment) is an MLIR-based end-to-end compiler and runtime that lowers Machine Learning models to a unified IR that scales up to meet the needs of the datacenter and down to satisfy the constraints and special considerations of mobile and edge.

GITHUB STARS: 3,647 - NEW STARS IN PAST YEAR: 853
TOTAL NUMBER OF CONTRIBUTORS: 1,015



NNStreamer

NNStreamer is a set of streamer plugins that support ease and efficiency for streamer developers adopting neural network models and neural network developers managing neural network pipelines and their filters.

GITHUB STARS: 1,240 - NEW STARS IN PAST YEAR: 112
TOTAL NUMBER OF CONTRIBUTORS: 439

Docling:

Prepare your Docs for use in AI Applications

- 📁 Parsing of multiple document formats incl. PDF, DOCX, XLSX, HTML, images, and more
- 📝 Advanced PDF understanding incl. page layout, reading order, table structure, code, formulas, image classification, ...
- 🔗 Unified, expressive DoclingDocument representation format
- 🔄 Various export formats (Markdown, HTML, JSON)
- 🔒 Local execution for sensitive data and air-gapped environments
- 🔌 Many plug-and-play ecosystem integrations
- 🔍 Extensive OCR support for scanned PDFs and images
- 💻 Simple and convenient CLI

github.com/docling-project



```
pip install docling

# a single document to markdown
docling https://arxiv.org/pdf/2408.09869.pdf

# a folder of documents to markdown and json
docling --to json --to md ./inputs/
```

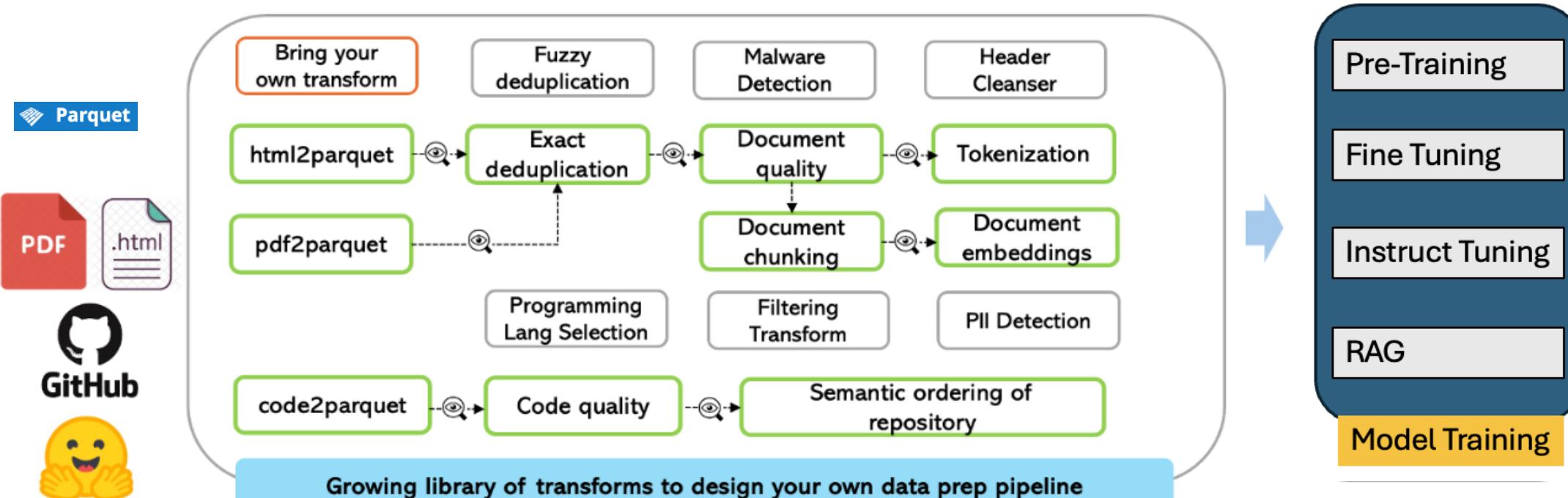




Data Prep Kit

<https://github.com/data-prep-kit/data-prep-kit>

DLF AI & DATA



Laptop to cluster scale flexibility





BeeAI helps enterprise developers turn ideas into production-ready agents with shareable interfaces - no need to build orchestration or frontends from scratch.



BeeAI is **fully open source** under Linux Foundation governance. Contribute, share ideas, and help shape the future of AI agents.



Two complementary components:
A **framework** for building reliable AI agents and a **platform** for creating shareable interfaces.



Hack with us on Github



[Start Here](#)

Get started building production-ready AI agents with enterprise-grade reliability, built-in caching, memory optimization, resource management, and real-time monitoring.



Get started creating a custom web interface with any AI agent in minutes.

<https://github.com/i-am-bee>

Thought Leadership

Members of the LF AI & Data can network and help shape the open source AI and Data market.

As a member you can participate in:



Governing Board Participation

Participate in elections to serve on the Governing Board to oversee the vision of LF AI & Data and work with the TAC.



Engagement with other Foundations and Industry Consortia

Participate in establishing collaborations with other umbrella foundations under the LF and broadly with other industry consortia organizations.



Technical Advisory Committee Leadership

LF AI & Data TAC provides technical leadership to the open source AI and Data community, accepts projects into incubation, graduates projects and launches new efforts and collaborations.

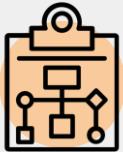


Interactive Landscape Placement

Our landscape is a comprehensive view of all open source AI and Data critical projects in the ecosystem. Anyone looking to adopt open source AI and Data projects comes here to review what technologies they should assess and adopt.

Community Engagement

LF AI & Data is a constellation of open source projects. Our members leverage many efforts to engage with our project's ecosystems and share their stories. As a member you can participate in:



ML/LLM Ops Work Group

Design a reference architecture for end-to-end ML Workflow, provide a reference implementation using hosted projects, and facilitate integration and interop across projects.



Generative AI Commons

The Generative AI Commons is a community-driven open membership initiative with elected chairs and represented by non-profits, academia and industry in a neutral forum. We embrace responsible and trustworthy AI, foster collaboration, promote AI literacy and advocate for our community.



BI & AI Work Group

Integrate the power of AI and BI to make it CI (Cognitive Intelligence) by combining the speed machines accelerate (AI) with the direction intuited by human insight (BI).



Security & Compliance Work Group

The LF AI & Data Security and Compliance Work Group is dedicated to formulating interconnected security use cases, threat models, and policies that can be leveraged to create a comprehensive security and compliance strategy for AI-enabled applications throughout their lifecycle.

Model Openness Framework

MOF Components



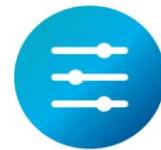
Datasets



Preprocessing
Code



Model
Architecture



Model Weights &
Parameters



Training Code



Inference Code



Evaluation
Code



Evaluation Data



Evaluation
Results



Technical
Report



Model Metadata



Model Card



Data Card



Research



Sample Model
Outputs



Configuration
file

2025 Conferences



- LF Member Summit: **March 18-20**, Napa, California
- Open Source Summit NA: **June 23-25**, Denver, CO
- Open Source Summit India: **August 5**, Hyderabad, India
- Open Source Summit Europe: **August 25-27**, Amsterdam, Netherlands
- AI Dev Europe: **August 28-29**, Amsterdam, Netherlands
- Open Source Summit Korea: **November 5**, Seoul, Korea
- AI Dev Japan: **December 8-10**, Tokyo, Japan

What the Georgia Tech Students did at IBM in 2025

Some Georgia Tech student interns:

- **GUI Wrapper for the DPK (Data Prep Kit) Planning Agent** – Student: Dobromir Iliev - Mentor: Shahrokh Daijavad
- **Root Cause Analysis with BeeAI and Docling** – Student: Alex (Kenneth) Jenkins – Mentor: Cynthia Unwin - GraniteRCA (Root Cause Analysis) - includes BeeAI , Docling and Granite
- **Generics with Granite : Do sharks attack beach goers?** – Student: Rakshit (Rocket) Nemakallu – Mentor: Alessandra Pascale
- **Showcase Granite with Python FIM prompts** – Student: Daniil Tatarinov – Mentor BJ Hargrave
- **KubeStellar** – Students: Gauri Sharma, Oorja Singh, Oliver Teng – Mentor Andy Anderson
- **Pytorch Foundation Models** – Student: Niraj Kamal - Mentor: Sahdev Zala
- **Model Openness Framework**– Students : Vishank Raghavan, Seungreol Tao - Mentors: Arnaud Le Hors – in collaboration with LF-AI and Data

In addition, Benjamin Hoover, an IBM employee, and also a PhD student at Georgia Tech, participated in the TechXChange Dev Day – topic : **A New Era for Associative Memories**



Call to Action

Join us at TechXChange in Orlando

- Community Day – Monday October 7
- Student Day – Tuesday October 8

Contribute to some of the open source projects and LF-AI and Data activities described in the following slides

Community
Day Pass

Free*

Monday's pre-conference day features open source projects and User Group gatherings.

- Community Day Kickoff
- Open source projects like Apache Superset and BeeAI
- User Groups like IDUG, COMMON, SHARE, MaximoGroups.org and more

*Excludes evening entertainment

Student Dev
Day Pass
Free*

Tuesday only: Hands-on learning designed just for students.

- Exclusive student-only sessions
- Opening General Session Kickoff
- Play in the Sandbox 3.0
- Lunch and refreshments included

*Students 18+ with valid student ID & .edu email (US and Canada only). Excludes evening entertainment.

Community Day - Free

- October 6 – Orlando, Florida
- Register -

<https://www.ibm.com/community/ibm-techxchange-conference/>



Open Source at Community Day: Open for all*

Connect with the community
building AI's open future



Contribute to real projects

Level up by contributing to production AI and open source codebases used by millions.



Connect with the community

Meet fellow developers, core maintainers, and contributors who are building the future of AI and open source together. Find your tribe and collaborate on what's next.



Shape AI's open future

Help steer the direction of AI tools and frameworks you rely on. Give feedback directly to maintainers and influence features that solve real problems for the community.



No-strings-attached collaboration

Join any conversation that sparks your interest. No sales pitches, no vendor lock-in. Just developers sharing knowledge, solving problems, and building together.

Open source projects including

AI

- BeeAI
- Data Prep Kit
- Docling
- Granite
- Ferrari
- Langflow
- Llama
- OPEA
- Open Horizon ML
- PAIG
- Podman
- PyTorch
- vLLM

Data

- Apache Cassandra
- Apache Gluten
- Apache Iceberg
- Apache Pinot
- Apache Ranger
- Apache Superset
- Clickhouse
- EdgeLake
- LangFlow
- llm-d
- NodeRED
- OpenSearch
- Presto
- Presto C++

Hybrid Cloud

- KServe
- Kubeflow
- KubeRay
- Kueue
- Kubernetes
- OpenCost
- Ray

Mainframe

- Open Mainframe Project
- Gradle
- Grafana
- Zowe

Student Day - Free

- October 7 – Orlando, Florida
- Register -
<https://www.ibm.com/community/ibm-techxchange-conference/>



IBM TechXchange 2025

October 7 | Orlando, FL

IBM

The logo for IBM TechXchange 2025 features a grid of various icons related to technology and development, such as a rocket, a diamond, a bee, and a person. Below the grid, the word "GO/" is written in large white letters, followed by the tagline "Explore. Build. Launch." and a photo of two people at a booth.

Boost your technical skills,
expand your network and
jumpstart your career in tech.



AI skill development

Deep dive into Gen AI, large language models (LLMs), and other next-gen technologies transforming the tech industry.



Real-world experience

Gain practical experience through hands-on workshops and labs that directly prepare you for the tech workforce.



Professional headshot, badge & networking

Level up your LinkedIn with a pro headshot and digital badge. Connect with industry leaders, peers, and future employers.



Exclusive access to cutting-edge tech

Live code with IBM AI tools, experiment with a quantum computer, and build your first AI chatbot.

Free*

Student Dev Day Pass

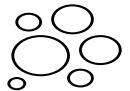
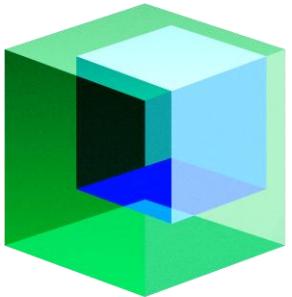
On-site access Tuesday with meals, swag and **Credly badge** to boost your resume.

October 7 Student Dev Day

- Opening General Session with the Senior Vice President of Software at IBM.
- Interactive sessions on generative AI and large language models.
- Lunch and learn panel discussion on AI ethics and AI-powered workforce development.
- Generative AI and LLMs BYOD workshops to help you build, solve real-world problems and learn to stay agile with the latest advances.
- Student Dev Showcase featuring student-led AI innovations in real-world applications.

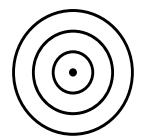
IBM Granite

A family of [open](#), [performant](#) and [trusted](#) AI models to accelerate enterprise AI adoption



Open

- Open sourced under Apache 2.0
- Transparency of data and training methods
- Customize with your data



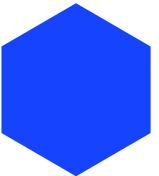
Performant

- Diverse range of fit-for-purpose models
- Designed for scalability
- Reasoning capabilities which uniquely maintain general performance
- Vision capabilities optimized for enterprise document understanding

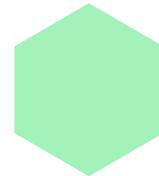


Trusted

- IP indemnification
- Responsible and safe AI
- Guardrails to secure data and mitigate risks
- Reasoning approach which uniquely preserves model safety



Large Language Models (LLMs) with Reasoning



Multimodal Models



Guardrail Models



Time Series Models



Embedding Models



Geospatial Models

Candidate Projects

- Create recipes
- Create mini-applications
- Adapt code recipes for latest models

Getting started with Granite models



ibm.biz/granite-community

ibm.com/watsonx/developer

huggingface.co/ibm-granite

contact: [BJ Hargrave](mailto:BJ.Hargrave@us.ibm.com)

email: hargrave@us.ibm.com

Mellea.ai

An [open-source library](#) for writing generative programs.



Tutorial

mellea.ai

GitHub

[generative-computing/mellea](https://github.com/generative-computing/mellea)

Contact

[Nathan Fulton](https://nathanfulton.com)

nathan@ibm.com

Star + Follow on GitHub



Local-first Capability

Generative programs can do Big Model Things™ without Big Model hardware.

Robust and Composable

Requirement-driven inference pipelines result in rock-solid libraries and deployable apps.

Ready to Scale

Swap out inference engines and models with ease. Massively scale inference in one line of code.

Open & Permissive

The open inference stack for building generative programs on open weight models.

PyTorch

- PyTorch is a leading open source framework for Machine Learning
- Project is hosted under the PyTorch Foundation
- Communication with the project community
 - [PyTorch Slack](#) - hosts a primary audience of moderate to experienced PyTorch users and developers for general chat, online discussions, collaboration. If you need a slack invite, fill this form:
<https://goo.gl/forms/PP1AGvNHpSaJP8to1>
 - [PyTorch Forum](#) - provides a primary medium for beginners looking for help

Skill

- ML, DL, Python, C++

Learning resources

- <https://pytorch.org/>

GitHub repo

- <https://github.com/pytorch/pytorch>



Contact

Sahdev Zala, spzala@us.ibm.com



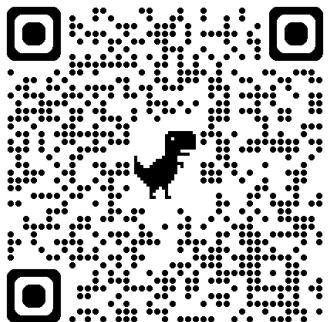
Data Prep Kit



<https://github.com/data-prep-kit/data-prep-kit>

Contact:
Santosh Borse ssborse@us.ibm.com
Shahrokh Daijavad shahrokh@us.ibm.com

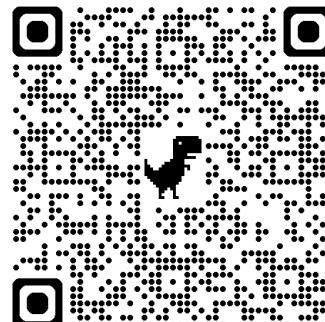
- **Powerful toolkit** for processing unstructured data for LLM training and downstream applications
- **38+ ready-to-use data transforms** for filtering, cleaning, and processing unstructured data.
- **Trusted by IBM Granite** - used to prepare data for the Granite family of open, high-performance AI models designed for enterprise adoption.
- **Effortless Scaling** - from single-machine development to large scale datacenter based processing.



Demo Notebook: GneissWeb 10T
token pretraining dataset recipe



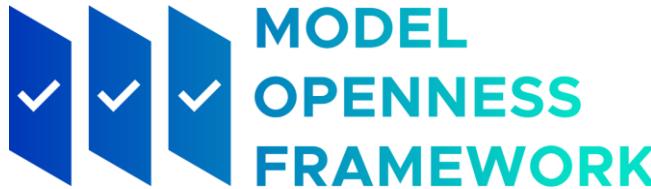
Demo Notebook: RAG



Demo Notebook: Data prep
for post training

Let's build it together: Open data. Open tools. Open collaboration.

- ✓ Build new transforms
- ✓ Improve Documentation
- ✓ Apply data prep kit in new use case
- ✓ Bug fixes
- ✓ Add support for new modalities
- ✓ Help benchmark and scale pipelines



Addressing the challenges in open AI

Challenges:

A lot of “open-washing”, lack of transparency, illegally converted licenses, inappropriate type of license, etc.

The Model Openness Framework (MOF):

A classification system to evaluate models based on their completeness and openness following open science principles.

The Model Openness Tool (MOT):

An online registry of models and their ranking against the MOF.

Contact: Arnaud Le Hors <lehors@us.ibm.com>

<https://isitopen.ai>



F k i l l 3E 4

Principal Authors



Dr. Haluk
Demirkan



Adel
Zaalouk



Dr. Suparna
Bhattacharya



Susan
Malaika

Notable contributors

Karen Bennet

Martin Foltin

Anni Lai

Michael Novak

Imani Carey

Adrian Gonzales

Kevin Lu

Ronald Petty

Oita Coleman

Ali Hashmi

Victor Lu

Jeff Polack

David Ellison

Ofer Hermoni

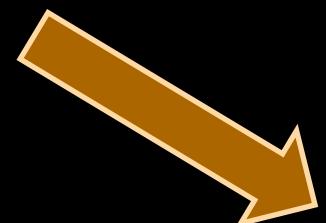
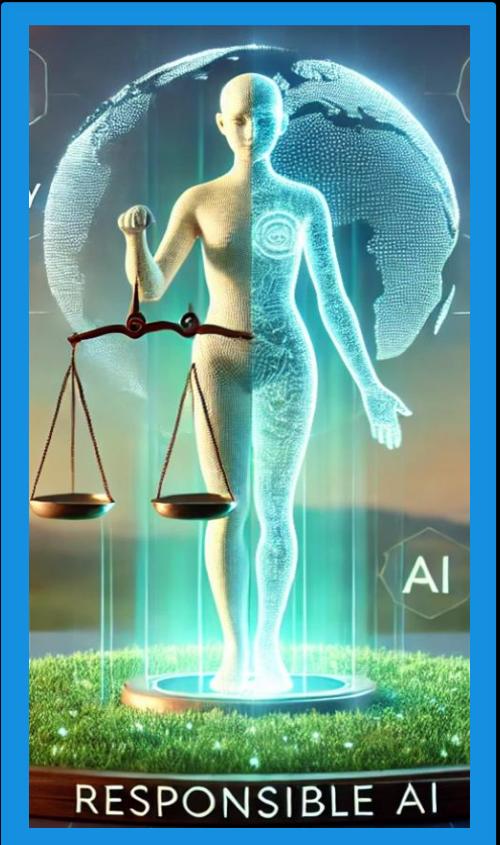
Rag Muthuregunathan

Annmary Roy

Andreas Fehlner Vini Jaiswal

Santhosh Sachindran

Join the
Responsible AI
Workstream at the
Generative AI
Commons to make
contributions, write
blogs, deliver
presentations and
workshops etc.



QLF AI & DATA

Thank you!

