Meeting of the LF AI & Data Technical Advisory Council (TAC)

June 30, 2022

LF AI & DATA

Antitrust Policy

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

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

Recording of Calls

Reminder:

TAC calls are recorded and available for viewing on the TAC Wiki

Reminder: LF AI & Data Useful Links

- Web site: lfaidata.foundation
- Wiki: wiki.lfaidata.foundation
- GitHub: github.com/lfaidata
- Landscape: https://landscape.lfaidata.foundation or https://l.lfaidata.foundation
- Mail Lists: https://lists.lfaidata.foundation
- Slack: https://slack.lfaidata.foundation
- Youtube: https://www.youtube.com/channel/UCfasaeqXJBCAJMNO9HcHfbA
- LF AI Logos: https://github.com/lfaidata/artwork/tree/master/lfaidata
- LF AI Presentation Template: https://drive.google.com/file/d/1eiDNJvXCqSZHT4Zk_-czASlz2GTBRZk2/view?usp=sharing

- Events Page on LF AI Website: https://lfaidata.foundation/events/
- Events Calendar on LF AI Wiki (subscribe available): https://wiki.lfaidata.foundation/pages/viewpage.action?pageId=12091544
- Event Wiki Pages: https://wiki.lfaidata.foundation/display/DL/LF+AI+Data+Foundation+Events

Agenda

- Roll Call (2 mins)
- Approval of Minutes from previous meeting (2 mins)
- FlagAI Sandbox Proposal – 20 minutes
- Outreach Committee update (tentative) – 5 minutes
- LF AI General Updates (2 min)
- Open Discussion (2 min)

TAC Voting Members - Please note

Please ensure that you do the following to facilitate smooth procedural quorum and voting processes:

- Change your Zoom display name to include your First/Last Name, Company/Project Represented
  - example: Nancy Rausch, SAS
- State your First/Last Name and Company/Project when submitting a motion
  - example: First motion, Nancy Rausch/SAS

TAC Voting Members

Note: we still need a few designated backups specified on wiki

Member Representatives (9 out of 17 required for quorum)

Member Company or Graduated Project	Membership Level or Project Level	Voting Eligibility	Country	TAC Representative	Designated TAC Representative Alternates
4paradigm	Premier	Voting Member	China	Zhongyi Tan	
Baidu	Premier	Voting Member	China	Ti Zhou	Daxiang Dong, Yanjun Ma
Ericsson	Premier	Voting Member	Sweden	Rani Yadav-Ranjan	
Huawei	Premier	Voting Member	China	Howard (Huang Zhipeng)	Charlotte (Xiaoman Hu) , Leon (Hui Wang)
Nokia	Premier	Voting Member	Finland	@ Michael Rooke	@ Jonne Soininen
OPPO	Premier	Voting Member	China	Jimin Jia	
SAS	Premier	Voting Member	USA	*Nancy Rausch	JP Trawinski
ZTE	Premier	Voting Member	China	Wei Meng	Liya Yuan
Acumos Project	Graduated Technical Project	Voting Member	USA	Amit Kumar	Prasanna Kulkarni
Adversarial Robustness Toolbox Project	Graduated Technical Project	Voting Member	USA	Beat Buesser	
Angel Project	Graduated Technical Project	Voting Member	China	Bruce Tao	Huaming Rao
Egeria Project	Graduated Technical Project	Voting Member	UK	Mandy Chessell	Nigel Jones, David Radley, Maryna Strelchuk, Ljupcho Palashevski, Chris Grote
Flyte Project	Graduated Technical Project	Voting Member	USA	Ketan Umare	
Horovod Project	Graduated Technical Project	Voting Member	USA	Travis Addair	
Milvus Project	Graduated Technical Project	Voting Member	China	Xiaofan Luan	Jun Gu
ONNX Project	Graduated Technical Project	Voting Member	USA	Alexandre Eichenberger	Prasanth Pulavarthi, Jim Spohrer
Pyro Project	Graduated Technical Project	Voting Member	USA	Fritz Obermeyer	

*Current TAC Chairperson

LF AI & DATA

Minutes approval

LF AI & DATA

Approval of June 16, 2022 Minutes

Draft minutes from the June 16th TAC call were previously distributed to the TAC members via the mailing list

Proposed Resolution:

- That the minutes of the June 16th meeting of the Technical Advisory Council of the LF AI & Data Foundation are hereby approved.

FlagAI

LF AI & DATA

30JUN2022

FlagAI: **F**ast **LA**rge-scale **G**eneral **AI** models

Presenter: Ledell Y. Wu (wuyu@baai.ac.cn)

Github: https://github.com/BAAI-Open/FlagAI

Proposal: https://github.com/lfai/proposing-projects/blob/master/proposals/flagai.adoc

Project Level Request to TAC: Sandbox

Background & Motivation

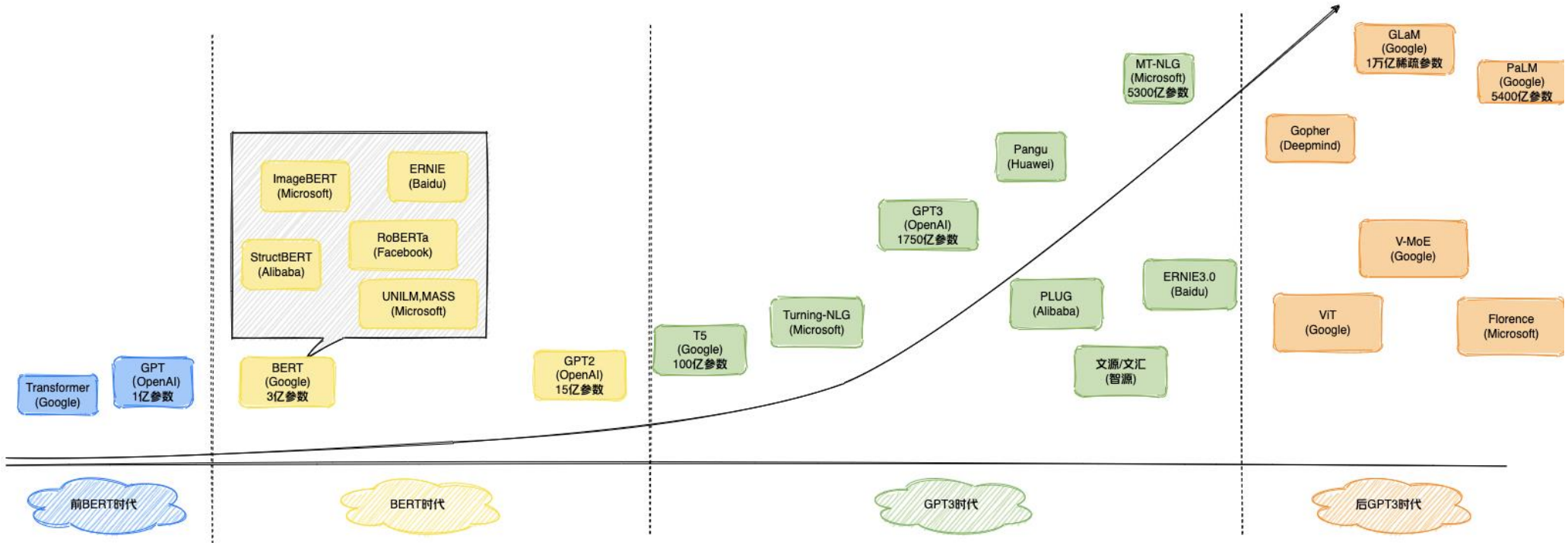


Photo credit: Jingang Wang

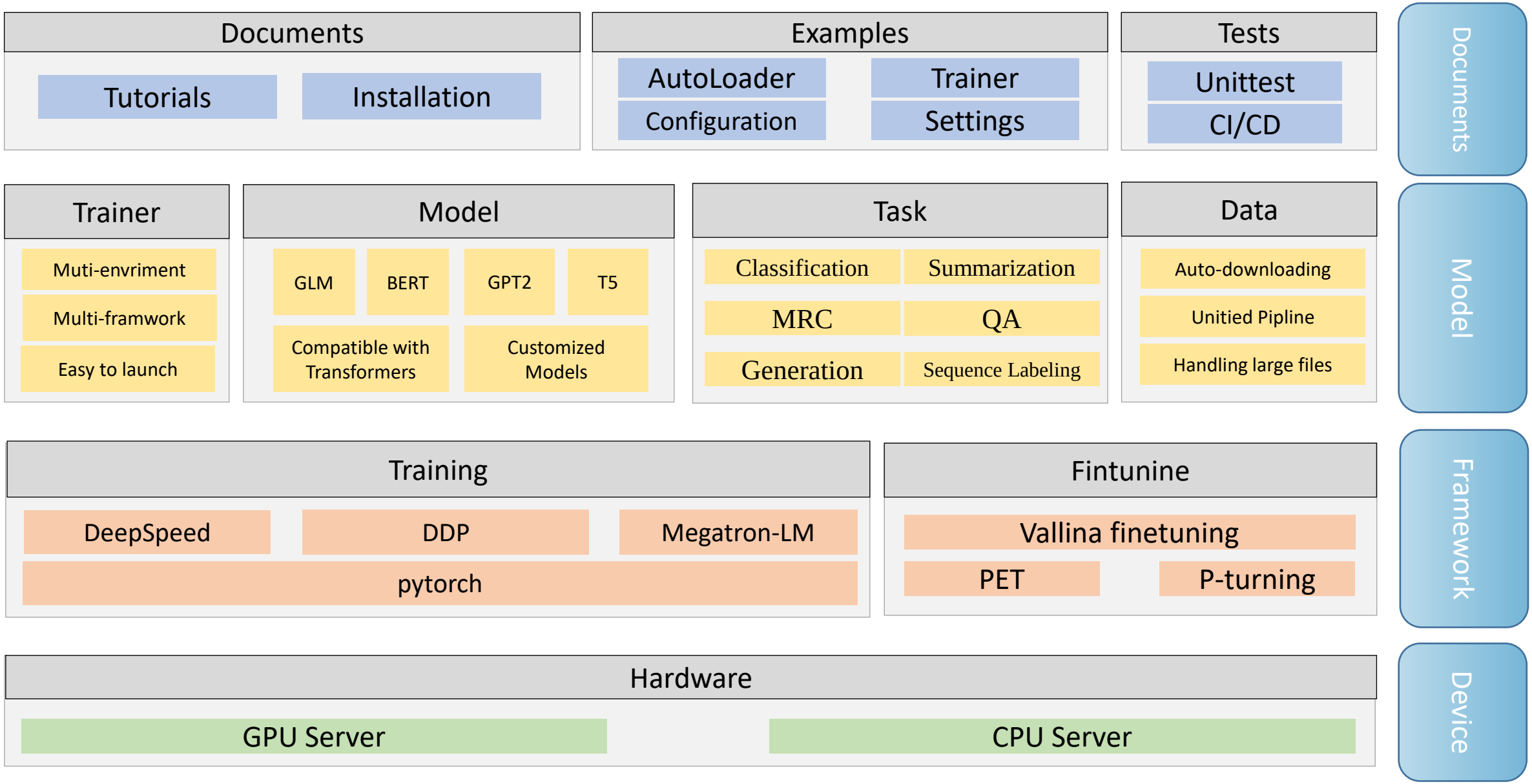
- Large-scale pre-training models become more and more important in deep learning;
- Need for tool than handles large-scale models properly while being easy-to-use
- Downstream tasks can benefit largely from large-scale models, but need specifically designed tools

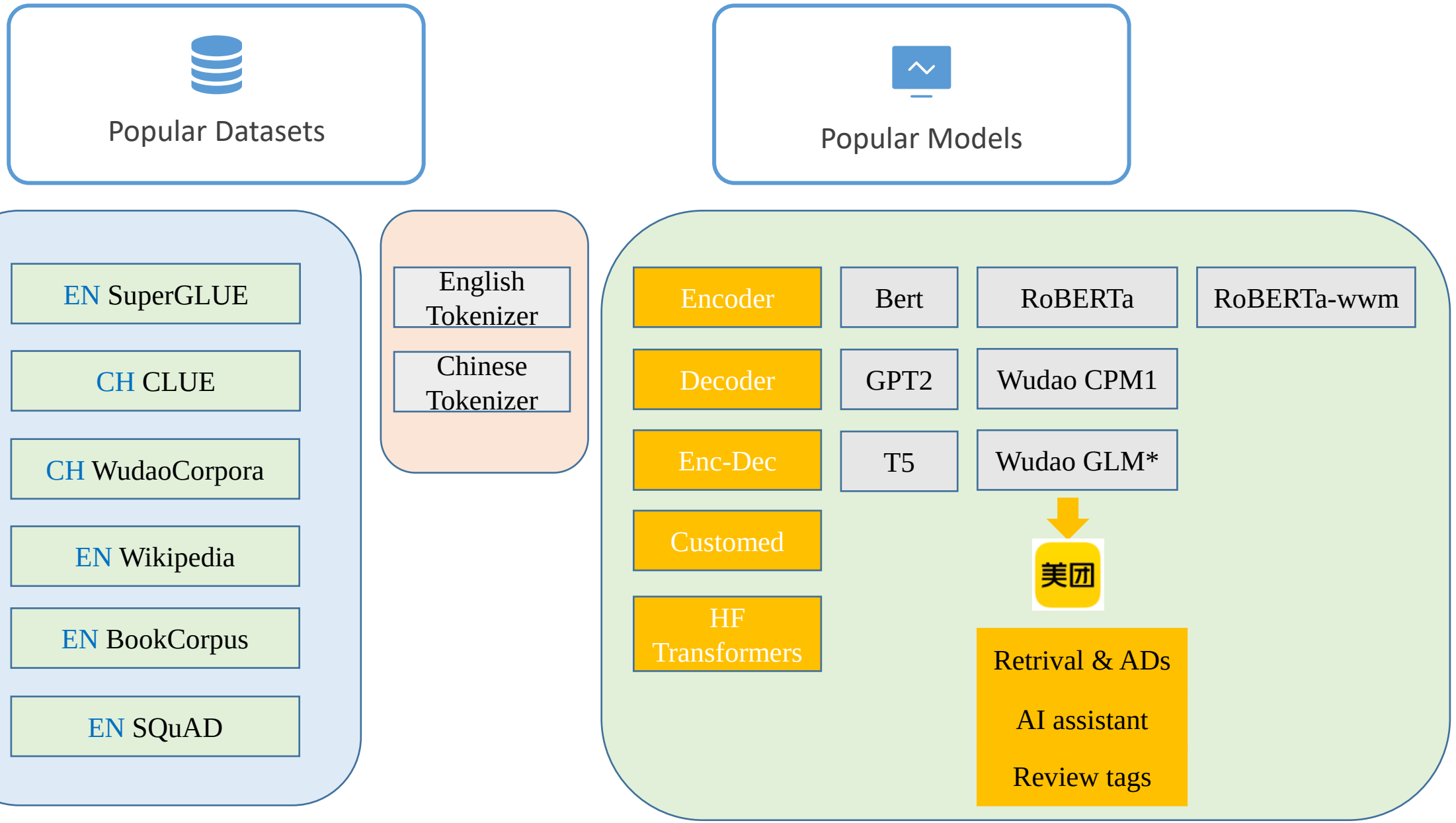
Overview

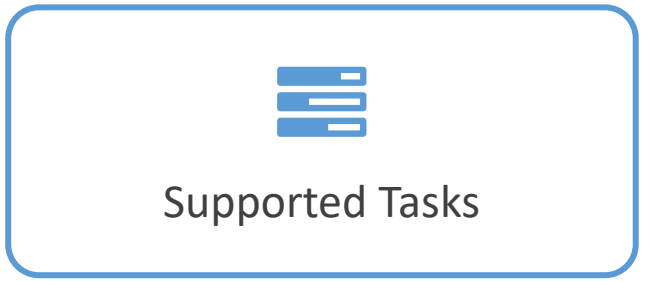
FlagAI (**F**ast **LA**rge-scale **G**eneral **AI** models) is an **fast, easy-to-use and extensible toolkit** for large-scale models. Our goal is to support training, fine-tuning, and deployment of large-scale models on various downstream tasks with multi-modality. Currently, we are focusing on NLP models and tasks.

- It supports **WuDao GLM** with a maximum of 10 billion parameters. It also supports **BERT**, **RoBERTa**, **GPT2**, **T5**, and models from Huggingface Transformers.
- It provides APIs to quickly download and use those pre-trained models on a given text, fine-tune them on widely-used datasets collected from SuperGLUE and CLUE benchmarks. It also provides prompt-learning toolkit for few shot tasks.
- These models can be applied to (Chinese/English) Text, for tasks like text classification, information extraction, question answering, summarization, and text generation.
- FlagAI is backed by the three most popular data/model parallel libraries — PyTorch/Deepspeed/Megatron-LM — with seamless integration between them. Users can parallel their training/testing process with less than ten lines of code.

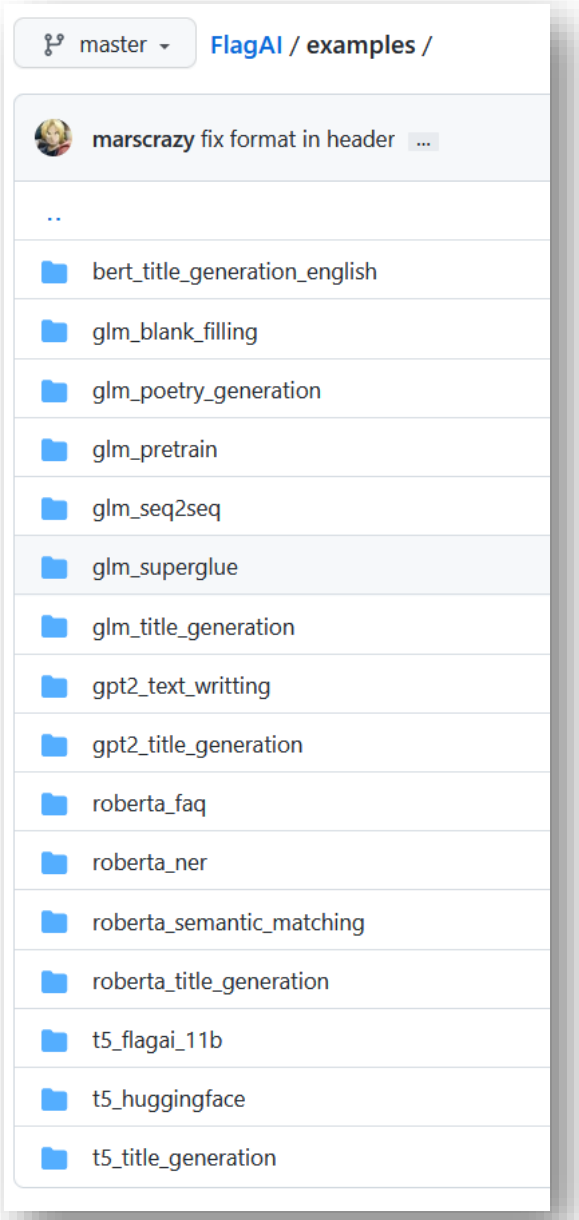
FlagAi飞智 Framework




Popular Datasets
Popular Models
EN SuperGLUE
CH CLUE
CH WudaoCorpora
EN Wikipedia
EN BookCorpus
EN SQuAD
English Tokenizer
Chinese Tokenizer
Encoder
Decoder
Enc-Dec
Customed
HF Transformers
Bert
GPT2
T5
RoBERTa
Wudao CPM1
Wudao GLM*
RoBERTa-wwm
美团
Retrival & ADs
AI assistant
Review tags



Semantic Matching	Search similar products	
Summarization	Title generation	
QA	MRC	FAQ
NER	KBQA	Task-oriented Dialogue
Classification	Sentiment Classification	Intent classification
Poetry Generation	Constrained generation	

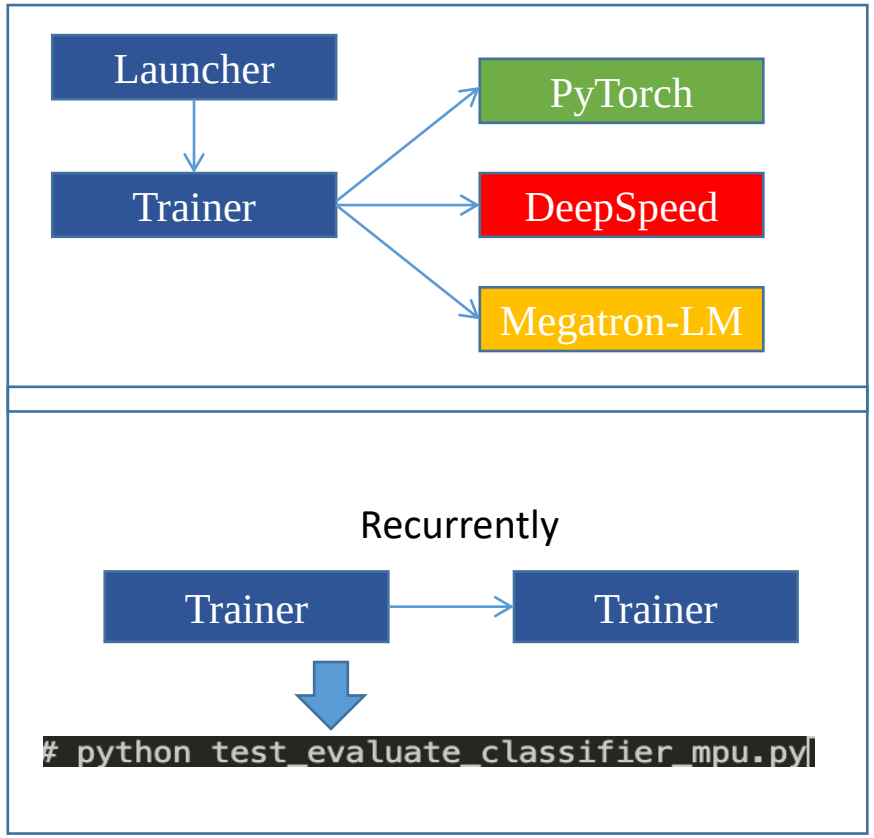


Multi-framework integration

Efficient Training

Combined

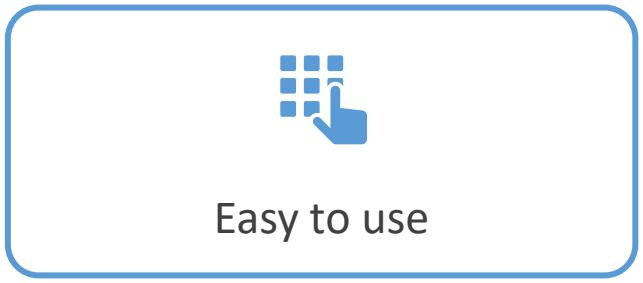
HuggingFace Transformers: extensiable
DeepSpeed, Megatron-LM: efficiency



Data Parallel	PyTorch DDP	Checkpointing
Model Parallel	DeepSpeed Stage1/2/3	Mixed Precision
Pipline Parallel	Megatron-LM	CPU Offload
Fused Parallel	BMTrain	Kernels

Related documents

- How to transform a customized model into a megatron-LM-style parallel model
- Using huggingface's t5-11b & tricks



AutoLoader | Predictor | Trainer

例：标题生成
输入：2007年，乔布斯向人们展示iPhone，并宣称它将会改变世界，还有人认为他在夸大其词。然而在8年后，以iPhone为代表的触屏智能手机已经席卷全球各个角落。未来智能手机将会成为真正的个人电脑。

输出：乔布斯宣布iphone 8年后将成为个人电脑

Detailed tutorials：

- Tutorial 1: How to construct and use Tokenizer
- Tutorial 2: Dataset Preprocessing Pipeline
- Tutorial 3: Major Function of Model Module
- Tutorial 4: Customize trainer for model and data-parallel training
- Tutorial 5: Simplify model and tokenizer Initialization by Using Autoloader
- Tutorial 6: Use off-the-shelf inference Algorithms with Predictor
- Tutorial 7: Use FlagAI prompt-learning tool-kit to improve performance on SuperGLUE
- Tutorial 8: Setup environment for training models with multi-machine
- Tutorial 9: Text generation with encoder/decoder/encoder-decoder models
- Tutorial 10: How to transform a customized model into a megatron-LM-style parallel model

Pretrained Models and examples

- Blank_Filling_QA with GLM
- Title Generation with GLM
- Poetry generation with GLM-large-ch
- Using huggingface's t5-11b & tricks
- Title Generation with RoBerta-WWM
- Semantic Matching with RoBerta-WWM
- NER with RoBerta-WWM
- Writing with GPT-2
- Title generation with T5
- Supported tasks

Tutorials

We provide a set of quick tutorials to get you started with the library:

- Tutorial 1: How to construct and use Tokenizer
- Tutorial 2: Dataset Preprocessing Pipeline
- Tutorial 3: Major Function of Model Module
- Tutorial 4: Customize trainer for model and data-parallel training
- Tutorial 5: Simplify model and tokenizer Initialization by Using Autoloader
- Tutorial 6: Use off-the-shelf inference Algorithms with Predictor
- Tutorial 7: Use FlagAI prompt-learning tool-kit to improve performance on SuperGLUE
- Tutorial 8: Setup environment for training models with multi-machine
- Tutorial 9: Text generation with encoder/decoder/encoder-decoder models
- Tutorial 10: How to transform a customized model into a megatron-LM-style parallel model

Example Usage

Python ▾ 自动换行 | 复制

```python
import torch
from flagai.auto_model.auto_loader import AutoLoader
from flagai.model.predictor.predictor import Predictor

loader = AutoLoader(task_name="lm",
                    model_name="GLM-large-ch",
                    )
model = loader.get_model()
tokenizer = loader.get_tokenizer()
device = torch.device("cuda" if torch.cuda.is_available() else "cpu")
model.to(device)

predictor = Predictor(model, tokenizer)

text = '问题：啤酒伤胃吗？回答：[gMASK]'
output=predictor.predict_generate_randomsample(text, top_k=50,
                                               repetition_penalty=4.0, top_p=1.
print(t,'\n',output)
```

QA example with GLM:
Question: Is drinking beer bad for your health?

Collaborators & Roadmap

Collaborators

- Industrial and Commercial Bank of China (ICBC):
  - Innovation on adapting large-scale models in financial banking scenarios
- Meituan:
  - Knowledge Distillation techniques on GLM models
- Tsinghua University NLP group:
  - Fast training and inference tools

Roadmap

- Integration of more modalities
  - Vision: ViT, Swin Transformers
  - Multi-modal: WenLan 2.0, CogView2
- Retrieval models / Model-based IR
- Fast training techniques:
  - BMTrain, FlashAttention
- Model Compression techniques:
  - Pruning, knowledge distillation, quantization

Potential Collaborations with LF AI&Data

- ONNX:
  inference and deploy with varies models supported by FlagAI
- RosaeNLG:
  to integrate Chinese NLG models and tasks in RosaeNLG
- Ludwig:
  integrate large-scale models and tasks from FlagAI to the declarative machine learning framework

Alignment with LF AI’s mission

Large-scale pre-trained model is a keypath to **build general AI**. It becomes the most powerful and famous technique lately.

- The transformer-based model can **understand natural language, category images and play video games as human beings;**
- Provide off-the-shelf tools for researchers/engineers to train/infer large-scale models, e.g., data-/model- parallel training algorithms.
- Build-in supports for most used model structures, e.g., encoder, decoder, and enc-dec structure.
- Build-in supports for many popular downstream tasks. Users can directly pull the finetuned model and run the model with a few lines of code.

TAC Vote on Project Proposal: FlagAI

Proposed Resolution:

The TAC approves the FlagAI Project as an incubation project at the sandbox level of the LF AI Foundation

MLOps Committee Proposal

- Jerry Tan
- 4Paradigm

LF AI & DATA

Purpose of MLOps Committee

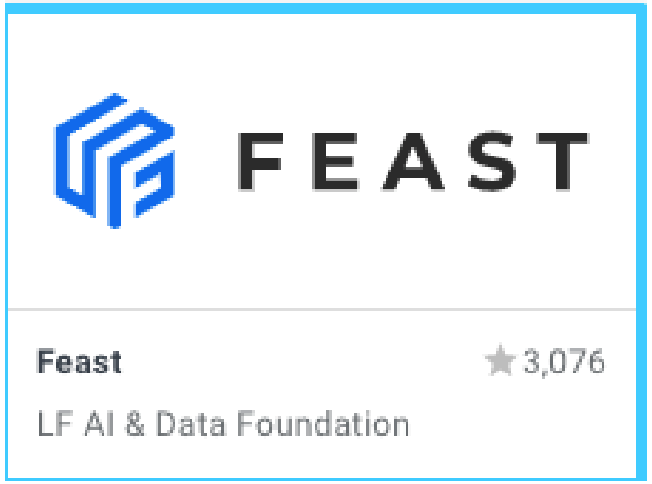
- We want to build one committee under LF AI & Data， named as MLOps Committee

- Initial Sponsors from TAC：ZTE and 4Paradigm
- Baidu，Tencent，huawei，DiDi agree to support this.

- MLOps’ Vision: To enable Data Engineers and System Engineers to build, test, run and update Machine Learning Application in the real world faster and easier.

- Committee’s Value Statement: To help MLOps related projects to get more recognization and adoption through cooperation by a passionate community of members.

The initial Projects



We will include more projects later.

Some Potential Projects:

- SQLFlow from Ants
- Meta database for AI (from Didi)

Benefits for LF AI & Data

- Enhance the branding of LF AI & Data in MLOps area
- Attract more projects MLOps related into LF AI & Data

What is the communication plan?

- Plan to have one online meeting once a month， 4Paradigm can organize it.
- Topics：
  - Identify projects that fit for MLOps Committee and their scenario
  - Exposure to industry about these projects
  - Provide opportunity for committee members to cooperate across different projects

- Possible outputs：
  - White Paper of MLOps for LFAI & Data
  - Best practice or Engineering Guide for MLOps

Plan for Deliverables Objective in this year

- A book named as 《MLOps Practice & Tools》

- This book will include：
  - What is MLOps? What is it for? And the guideline
  - Open Source Tools （Feast，OpenMLDB，Adlik)
  - MLOps in real business （Netease，xiaomi，ICBC，etc)

TAC Vote on Project Proposal: MLOps Committee

Proposed Resolution:

The TAC approves the MLOps committee as an committee of LF AI Foundation

Outreach Committee Update

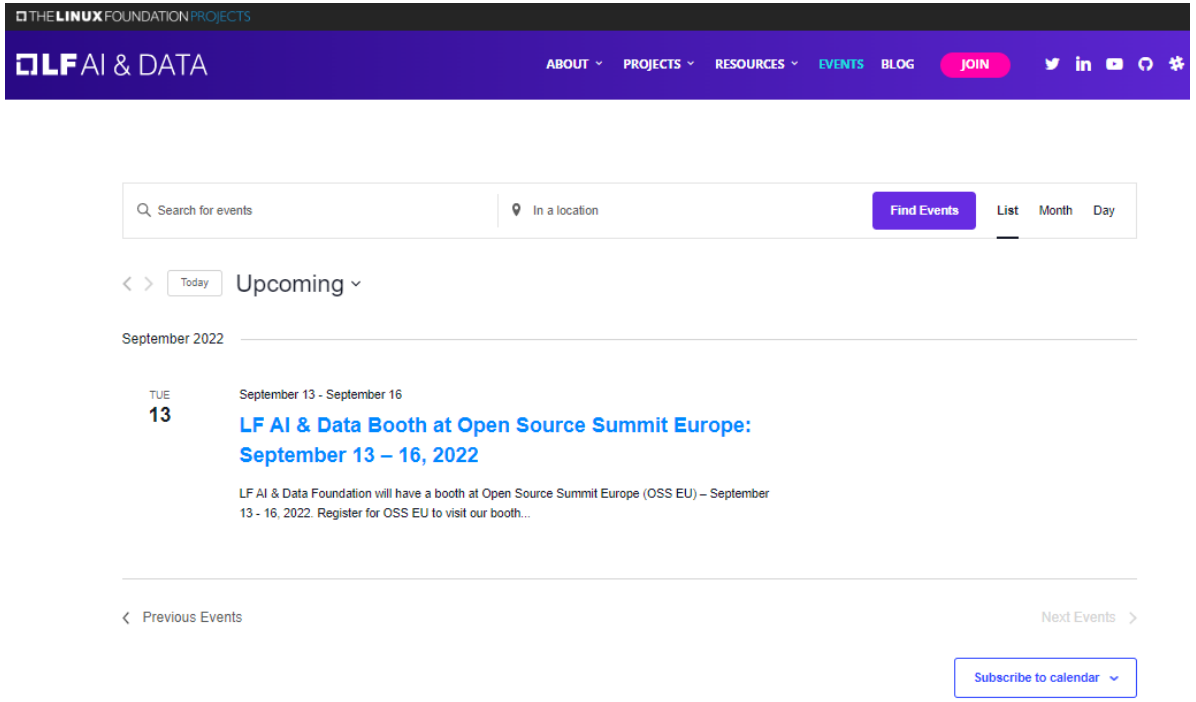
June 30, 2022

Hu Xiaoman (Charlotte)

THE LINUX FOUNDATION

LF AI & DATA

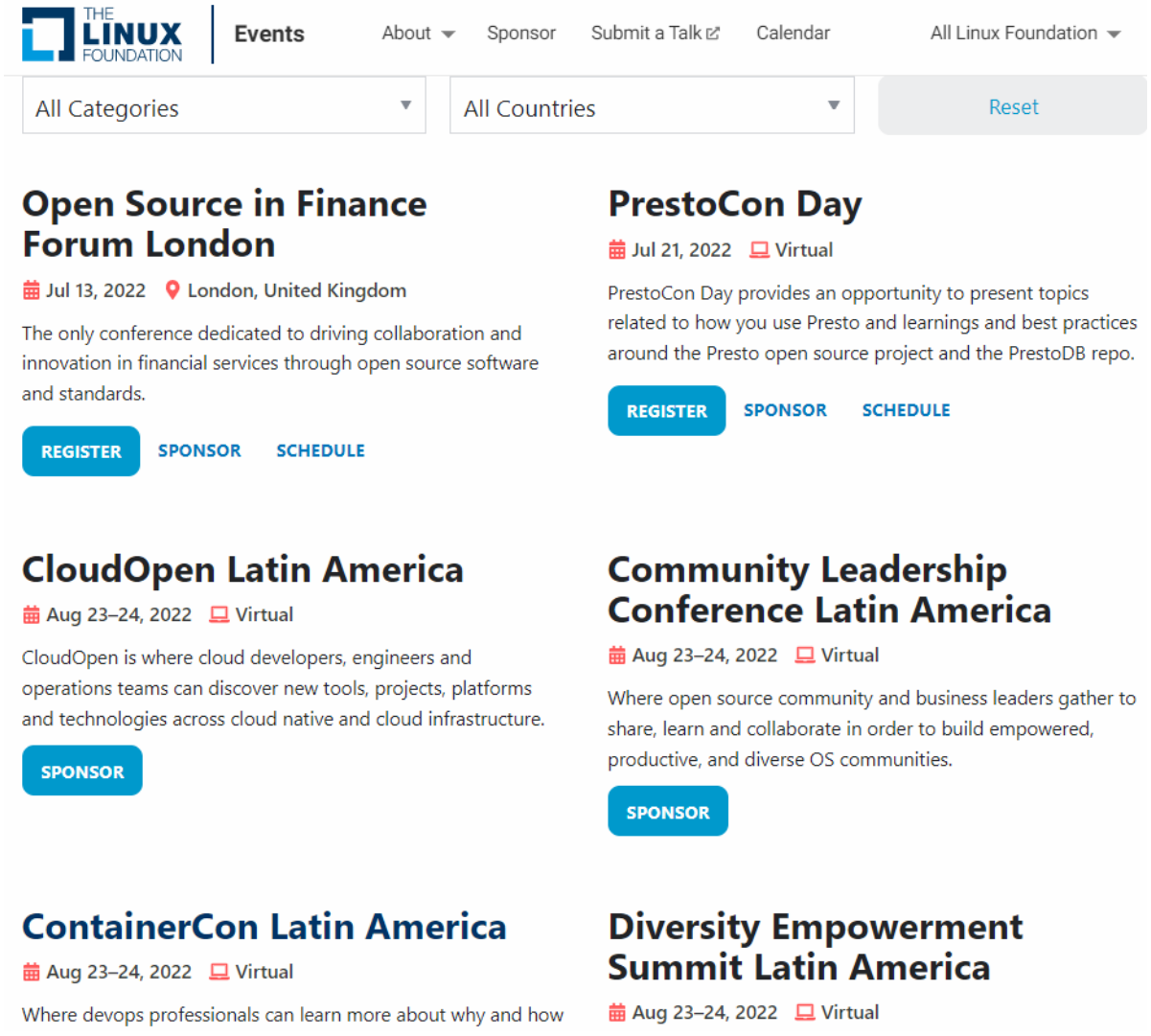
Upcoming Events



Visit the LF AI & Data Events Calendar or the LF AI & Data 2022 Events wiki for a list of all upcoming events

To discuss participation in an event or to host an event, please email events@lfaidata.foundation

Event Call For Proposals



All **Linux Foundation** events for 2022 are published here

Browse the events list and determine if you'd like to submit a proposal to present your project

Recent Announcements

Visit the LF AI & Data Blog for announcement details

Interested in contributing a guest blog post? Review our blog guidelines, submit a blog request, and we'll reach out to you soon. For questions, please email pr@live-lfprojects3.linuxfoundation.org. We look forward to hearing from you!

BLOG GUIDLINES | SUBMIT A BLOG

ONNX Community Election Announcement

Jun 22, 2022

Guest Author, Alexandre Eichenberger As an open format to represent machine learning and deep learning models that deploy and execute on diverse hardware platforms and clouds, ONNX is widely supported by leading hardware and software providers. The community continues to grow with new members, new contributors, new capabilities, and new...

Adlik Eagle Release (v0.5.0) Now Available!

Jun 21, 2022

We are thrilled to announce that Adlik, an LF AI & Data Foundation Incubation-Stage Project, released version 0.5.0 on June 21th, called Eagle. Adlik is a toolkit for accelerating deep learning inference, which provides an overall support for bringing trained models into production and eases the learning curves for different...

Release of Report: A Guide to Enterprise Open Source

Subscribe here to tac-general mailing list to get emails with recent announcements

Promoting Upcoming Project Releases

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

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

LF AI & Data PR/Comms

Please follow LF AI & Data on Twitter (@LFAIDataFdn) & LinkedIn and help amplify news via your social networks - Please retweet and share!

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

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

Upcoming TAC Meetings

LF AI & DATA

Upcoming TAC Meetings

- July 14: FATE from Webank; Open Dataology from BAAI
- July 28: Open

Please note we are requesting special topics for future meetings.

If you have a topic idea or agenda item, please send agenda topic requests to tac-general@lists.lfaidata.foundation

Open Discussion

LF AI & DATA

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

Legal Notice

- The 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