

LF AI & Data Foundation Technical Advisory Council (TAC) Meeting Minutes July 27, 2023 (6:00am - 7:00am PST) via Zoom

TAC Meetings are bi-weekly and open to everyone in the LF AI & Data community.

Voting Notes

Under the original charter rules we need 50% attendance from the voting members (represented by premier members + graduate projects) to meet quorum. To ensure quorum is met more consistently, the TAC implemented the following attendance and voting eligibility rule effective Dec 1, 2022: Voting members of the TAC who miss 2 consecutive meetings will not be allowed to vote and must then attend 2 consecutive meetings before their voting privileges are reinstated.

With the new attendance and voting rule adopted by the TAC, the number of eligible voting members varies per week and is tracked [here](#). Please see meeting minutes on the [TAC Wiki](#) for details on attendance and voting eligibility on a per meeting basis.

Please see the [TAC Wiki](#) for more information about voting.

Please see current voting members, including alternates, on the [TAC Wiki](#).

Voting details for today's meeting: 9 voting members were eligible to vote and 10 voting members were in attendance. Quorum was met for this meeting.

TAC Voting Member	Voting Representative	7/27/2023
Quorum	Attendance/No. Eligible Votes	9 of 10
4paradigm	Zhongyi Tan	Jerry Tan
Baidu	Jun Zhang	Jun Zhang
	Alternate: Daxiang Dong	
	Alternate: Yanjun Ma	
Ericsson	Rani Yadav-Ranjan	
Huawei	Howard (Huang Zhipeng)	Huang Zhipeng
	Alternate: Charlotte (Xiaoman Hu)	
	Alternate: Leon (Hui Wang)	
IBM	Susan Malaika	Susan Malaika

	Alternate: Manish Nagireddy	
Nokia	Michael Rooke	Michael Rooke
	Alternate: Jonne Soininen	
OPPO	Jimmy (Hongmin Xu)	
SAS	*Nancy Rausch	*Nancy Rausch
	Alternate: JP Trawinski	
ZTE	Wei Meng	
	Alternate: Liya Yuan	Alternate: Liya Yuan
Adversarial Robustness Toolbox Project	Beat Buesser	Beat Buesser
	Alternate: Kevin Eykholt	
Angel Project	Jun Yao	
Egeria Project	Mandy Chessell	Mandy Chessell
	Alternate: Nigel Jones	
	Alternate: David Radley	
	Alternate: Maryna Strelchuk	
	Alternate: Ljupcho Palashevski	
	Alternate: Chris Grote	
Flyte Project	Ketan Umare	
Horovod Project	Travis Addair	
Milvus Project	Xiaofan Luan	
	Alternate: Jun Gu	
ONNX Project	Alexandre Eichenberger	Alexandre Eichenberger
	Alternate: Jim Spohrer	Alternate: Jim Spohrer
	Alternate: Prasanth Pulavarthi	
	Alternate: Andreas Fehlner	
Pyro Project	Fritz Obermeyer	

Note for Voting Members

Please ensure you attend the bi-weekly TAC meetings to maintain voting eligibility. If you have not already provided an alternate representative, please email Nancy Rausch (Nancy.Rausch@sas.com), TAC Chair AND operations@lfaidata.foundation to designate an alternate representative. It is critical to meet quorum during the meetings especially when there are voting items on the agenda.

LF Attendees

Lucy Hyde, Program Manager, Linux Foundation
Reden Martinez, Project Coordinator, Linux Foundation
Nora Anwar, Marketing Manager, Linux Foundation
Nathan Southern, Project Coordinator, Linux Foundation

Invited Guests/Presenters

Sami Jaghouar - DocArray Project
Joan Fontanals Martinez - DocArray Project
Nan Wang - DocArray Project
Julien Le Dem - OpenLineage Project
Michael Robinson - OpenLineage Project

Jon Stine, Executive Director, Open Voice Network
Masahiro Fukuyori - Fujitsu
Jigisha Mavani - IBM
Ali Hasmi - IBM
Rafael Vasquez
Bj Hargrave
Suparna Bhattacharya

Call to Order

Nancy Rausch (NR) called the meeting to order at 6:01 am Pacific and Reden Martinez (RM) recorded the minutes.

NR reviewed the antitrust policy notice.

Agenda

Nancy Rausch (NR) reviewed the agenda for the meeting. There were no further changes or additional topics added.

- Roll Call (1 mins)
- Approval of Minutes from previous meetings (2 mins)
- DocArray Project : Promotion from Sandbox to Incubation
- OpenLineage Graduation Proposal
- Open Discussion

Approval of Minutes

Nancy Rausch (NR) presented a resolution to approve the minutes of the July 13, 2023 TAC meeting.

Proposed Resolution:

That the minutes of the July 13, 2023 meeting of the Technical Advisory Council of the LF AI & Data Foundation are hereby approved.

Liya Yuan made the first motion to approve the minutes, and Jerry Tan seconded the motion.

APPROVED - By vote of the TAC, the minutes of the July 13, 2023 meeting of the Technical Advisory Council were approved.

DocArray Project : Promotion from Sandbox to Incubation

Nancy Rausch (NR) welcomed Sami Jaghourar (SJ) from the DocArray project. SJ provided updates regarding the DocArray project.

DocArray is a Python library expertly crafted for the representation, transmission, storage, and retrieval of multimodal data. Tailored for the development of multimodal AI applications, its design guarantees seamless integration with the extensive Python and machine learning ecosystems.

The DocArray Project was launched in January 2022 and has experienced remarkable growth and success. After its donation to LF AI in November 2022, the project's popularity soared, garnering 2.4k GitHub stars (a 100% increase), attracting 60+ contributors (a 50% rise), and achieving 140k monthly downloads (a 40% surge). A major milestone was reached with the release of DocArray v2, which involved a complete rewrite of the codebase. These achievements exemplify the project's impact and effectiveness within the developer community.

SJ also shared the current users of the project and its contributors. SJ also presented the project overview, showing its stars' history, total commits, issues, forks, and PR creators.

DocArray benefits from active contributions from three organizations: Jina AI, Qdrant, and Weaviate. These organizations actively engage in fixing issues, publishing articles, and participating in TSC (Technical Steering Committee) meetings to shape the project's future. Additionally, other contributors, including Redis and IBM, contribute intermittently. Moreover, the project receives input from individual contributors, with ongoing collaborations through Google Summer of Code and dozens of others contributing to fixes, documentation, and other essential aspects of the project. This diverse and dedicated community plays a crucial role in the continued success and development of DocArray.

DocArray has made significant strides in integrating into the AI landscape, with several popular packages, including LangChain, LLama index, and GptCache, successfully incorporating it into their functionalities. Moreover, DocArray's adoption in the academic realm has been notable, with renowned figures like Andrew Ng utilizing it in courses at deeplearning.ai. The project's recognition extends to even more prominent platforms, such as its usage in the CVPR2023 Tutorial: Neural Search in Action. These achievements showcase the widespread acceptance and impact of DocArray within the AI community.

DocArray holds a Silver OpenSSF Best Practice Badge.

Currently, the project is applying for project promotion from the Sandbox to Incubation stage.

Proposed Resolution:

That the DocArray project is approved by the Technical Advisory Council (TAC) as an Incubation project of the LF AI & Data Foundation.

Mandy Chessell made the first motion to approve, and Susan Malaika seconded the motion.

APPROVED - By vote of the TAC, the DocArray project was approved as an Incubation project.

OpenLineage Project: Graduation Proposal

Nancy Rausch (NR) welcomed Julien Le Dem (JLD) from the OpenLineage project, and JLD provided updates regarding OpenLineage.

OpenLineage began in December 2020 when it was first released as open source. By May 2021, the project joined LFAI & Data as a Sandbox project and gained recognition and support within the AI community, evidenced by the acquisition of 300 GitHub stars. The collaboration with Marquez further strengthened its capabilities, solidifying its position in the developer landscape. By December 2022, it advanced to the Incubation stage and garnered 1.1K GitHub stars. In December 2022, The project's success attracted collaborations with organizations such as Microsoft, Snowflake, Airflow, Egeria, Northwestern Mutual, Amundsen, and GX, further enhancing its reach and impact.

Today, it has formed collaborations with DataGalaxy, Atlan, Manta, Keboola, Matillion, Microsoft, Snowflake, Airflow, Egeria, Northwestern Mutual, Amundsen, and GX, showcasing its widespread adoption and potential for further growth. Moreover, it has achieved 1.3K GitHub stars and garnered a community of 1K+ Slack members.

OpenLineage Progress Since 2021

Since 2021, the OpenLineage Project has achieved remarkable progress, making it a standout in the open-source community. The project has attracted 266 diverse and dedicated contributors, including organizations like DataGalaxy, Bloomberg, Microsoft, Astronomer, HSBC, UBS, Northwestern Mutual, and NTT Data Deutschland.

The emphasis on best practices has earned the project the CII Best Practices Gold Badge.

Development activity has remained consistently strong, with a total of 4.8K commits in the past 12 months, averaging around 41 per week.

Governance and procedures play a crucial role in the project's success, and OpenLineage has implemented essential structures, including a Technical Steering Committee (TSC), Contributor Guide, Charter, CoC., and Code Quality Assurance measures.

OpenLineage has also fostered collaborations with other projects under LFAI & Data, including Egeria, Marquez, and Amundsen.

Today, OpenLineage is applying for project promotion from the Incubation stage to Graduation stage.

Before the end of the presentation, Nancy Rausch (NR) inquired with Julien Le Dem (JLD) about the potential integration of OpenLineage with other projects outside the foundation, particularly ones gaining popularity like Open Metadata. Additionally, Nancy asked if there were any plans for collaboration with these projects.

JLD responded by clarifying that they have indeed had discussions with Open Metadata and DataHub. JLD expressed that while the idea of integration between the projects makes sense, each project has its own set of priorities and commitments. As a result, it seems that integrating with these projects may not be an immediate priority for them in the near future.

Following up on the previous question, NR asked JLD whether there would be any overlap between OpenLineage and the other projects or if they would be complimentary in terms of features.

JLD clarified that there might be a little bit of overlap between OpenLineage and the other projects. However, JLD also expressed the belief that despite this overlap, the projects will also have complementary features. This indicates that while there may be some shared functionalities, they will likely complement each other, offering distinct benefits and value to users and the broader community.

JLD also shared the 2023 Roadmap of the project. The OpenLineage project 2023 roadmap focused on enhancing capabilities and integration. Key milestones include Static Lineage support and OpenLineage 1.0.0 release, improving SQL parser, and enriching Flink integration. Collaboration goals involve contributing to Airflow, enabling real-time updates, and extending column lineage. Other plans include automating docs, adding dbt Cloud support, and supporting Apache Atlas. OpenLineage's dedication to improvement promises an even more powerful data lineage and metadata solution for users.

Proposed Resolution:

That the OpenLineage project is approved by the Technical Advisory Council (TAC) as a graduated project of the LF AI & Data Foundation.

Mandy Chessel made the first motion to approve, and Alexandre Eichenberger seconded the motion.

APPROVED - By vote of the TAC, the OpenLineage project was approved as a Graduated project.

Open Discussion

During the discussion, Susan Malaika (SM) raised the issue that the TrustedAI Committee needs to confirm before August 24, 2023, whether they can proceed with the scheduled presentation due to a schedule conflict. Nancy Rausch (NR) acknowledged and agreed with this concern.

In addition, SM inquired about the integration of Fujitsu's Intersectional Fairness code with AIFairness360. Masahiro Fukuyori responded, stating that the integration has not commenced yet, and they will need to have further discussions with the IBM team to move forward with the integration.

Upcoming TAC Meetings

August 10, 2023 - DeepCausality Sandbox Project Proposal, LFEEdge Presentation

August 24, 2023 - AutoMLI project from Fujitsu, Trusted AI Committee update

Please note the TAC is always open to agenda suggestions and guest presentations. If you have a topic you would like to request, please email tac-general@lists.lfaidata.foundation for review and coordination via the TAC Chair accordingly.

Closing

With no further business, the meeting was adjourned by NR at 6:42 AM Pacific.

Chat:

21:13:25 From Julien Le Dem (OpenLineage) to Everyone:

Very impressive!

21:13:40 From Michael Robinson to Everyone:

Reacted to "Very impressive!" with 

21:13:47 From Fukuyori, Fujitsu to Everyone:

Reacted to "Very impressive!" with 

21:13:58 From Beat Buesser (ART, voting member) to Everyone:

Reacted to "Very impressive!" with 

21:14:23 From Nora Anwar to Everyone:

Reacted to "Very impressive!" with 

21:14:59 From Julien Le Dem (OpenLineage) to Everyone:

Congrats!

21:15:03 From Nan Wang (DocArray & Jina AI, no-voting) to Everyone:

Many thanks to all for your support!

21:15:04 From Michael Robinson to Everyone:

Congratulations!

21:15:56 From susan malaika (IBM - voting) to Everyone:

welcome docarray - sorry no questions - but looks very interesting 😊 love data projects :)

21:30:35 From Michael Robinson to Everyone:

<https://github.com/Natural-Intelligence/openLineage-openMetadata-transporter>

21:33:47 From Michael Robinson to Everyone:

Yes, apologies for the oversight!

21:34:17 From susan malaika (IBM - voting) to Everyone:

Replying to "Yes, apologies for t..."

no problem at all - awesome that you have the connection

21:36:55 From Suparna Bhattacharya to Everyone:

Congratulations!

21:36:55 From Michael Robinson to Everyone:

Thank you!