

LF AI & Data Foundation Technical Advisory Council (TAC) Meeting Minutes July 13, 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: 7 voting members were eligible to vote and 7 voting members were in attendance. Quorum was met for this meeting.

TAC Voting Member	Voting Representative	7/13/2023
Quorum	Attendance/No. Eligible Votes	7 of 7
4paradigm	Zhongyi Tan	Jerry Tan
Baidu	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	
	Alternate: Jonne Soininen	
OPPO	Jimmy (Hongmin Xu)	
SAS	*Nancy Rausch	*Nancy Rausch
	Alternate: JP Trawinski	
ZTE	Wei Meng	
	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

- Ibrahim Haddad, Executive Director, LF AI & Data
- Lucy Hyde, Program Manager, Linux Foundation
- Reden Martinez, Project Coordinator, Linux Foundation
- Nathan Southern, Project Coordinator, Linux Foundation

Invited Guests/Presenters from Oppo

- Qiang Qiu - Principal Engineering Manager - Oppo
- Pengzhouhu - Oppo

Call to Order

Nancy Rausch (NR) called the meeting to order at 6:03 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)
- ShaderNN Sandbox Project Proposal (40 mins)
- Open Discussion

Approval of Minutes

Nancy Rausch (NR) presented a resolution to approve the minutes of the June 29, 2023 TAC meeting.

Proposed Resolution:

That the minutes of the June 29, 2023 meeting of the Technical Advisory Council of the LF AI & Data Foundation are hereby approved.

Mandy Chessel made the first motion to approve the minutes, Jerry Tan seconded the motion.

APPROVED - By vote of the TAC, the minutes of the June 29, 2023 meeting of the Technical Advisory Council were approved.

ShaderNN Sandbox Project Proposal

Nancy Rausch (NR) extended a warm welcome to Qiang Qiu (QQ) and Peng Zhouhu (PZ). Both individuals are representing Oppo and are set to present the ShaderNN Project Sandbox Project Proposal.

ShaderNN is a lightweight and efficient inference engine for mobile GPUs. It is an innovative inference engine developed by the OPPO Computing & Graphics Research Institute. It addresses the challenges of mobile inference by leveraging the power of GPU shaders for efficient and lightweight deep learning inference on mobile devices.

The need for a lightweight and efficient inference engine arises from the challenges faced in mobile inference. These challenges include limited computational capacity, low power budget, model compatibility, customizable and lightweight implementation, and being deeply coupled with image/graphic applications. ShaderNN aims to overcome these challenges and provide a solution tailored specifically for mobile GPUs.

One of the key advantages of ShaderNN is its utilization of GPU shaders for optimal performance. By implementing core operators using GPU shaders, ShaderNN takes advantage of parallel computing capabilities, resulting in high-performance inference. Additionally, ShaderNN optimizes the inference process by utilizing a pre-built static computation graph, which incorporates constant folding and operator fusion to accelerate forward operation speed.

Portability and versatility are also key features of ShaderNN. It ensures independence from external libraries, reducing overhead and simplifying integration by avoiding third-party library dependencies. ShaderNN is optimized specifically for mobile platforms, enabling effortless

portability, deployment, and upgrades. It supports popular deep learning frameworks like TensorFlow, PyTorch, and ONNX, allowing compatibility with a wide range of models. Moreover, ShaderNN offers the flexibility of user-defined operators, making it easy to implement new models.

The architecture of ShaderNN is based on the fragment shader pipeline and is well-suited for real-time graphics applications. It offers efficient integration with the graphics rendering pipeline or image processing applications through texture-based input/output. This integration eliminates the need for expensive data transfers and format conversions between the CPU and GPU.

ShaderNN has demonstrated impressive performance and power consumption savings in comparison to other inference engines. It outperforms TensorFlow Lite and MNN on selected target processor chipsets, delivering significant performance gains and energy savings for various tasks such as spatial denoise, ESPCN, Resnet18, and YOLO v3 tiny.

The open-source nature of ShaderNN enables easy access to its source code, making it highly customizable and extendable. The project provides a developer guide, tools for model conversion, a demo app, and a model zoo containing common CNN models. Additionally, ShaderNN has a roadmap for further development, including support for Vulkan backend, new operator and model support, optimization of convolution and matrix multiplication, and engagement with more users.

OPPO seeks to collaborate with LF AI & Data as a Sandbox Project to host ShaderNN. The collaboration aims to integrate ShaderNN with other LF AI & Data projects, such as data lineage with ONNX and OpenBytes, middleware plugin integration with Adlik and DeepRec, and sharing optimization points with BeyondML and Acumos AI.

In conclusion, ShaderNN offers a shader-based lightweight and efficient inference engine for mobile GPUs. Its utilization of GPU shaders, portability, versatility, and impressive performance make it a promising solution for accelerating deep learning inference on mobile devices.

Open Discussion

Proposed Resolution:

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

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

APPROVED - By vote of the TAC, the ShaderNN project was approved as a Sandbox project.

Before closing the meeting, Susan Malaika (SM) suggested informing the LFEEdge Team in advance about this project, as it could be relevant to their AI-related work. Nancy Rausch concurred with the suggestion.

Upcoming TAC Meetings

July 27 - DocArray proposal to move from Sandbox to Incubation - Tentative Project Review

August 10 - LF Edge Presentation and DeepCausality Sandbox Project Proposal

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 6:57 AM Pacific.