LF AI & Data Technical Advisory Council (TAC)

Biweekly call - March 21, 2024

□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

LF AI & Data Useful Links

- › Web site: lfaidata.foundation
- › Wiki: wiki.lfaidata.foundation
- › GitHub: github.com/lfai
- › Landscape: https://landscape.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/lfai/artwork
- › PPT Template: https://drive.google.com/file/d/1eiDNJvXCqSZHT4Zk_-czASlz2GTBRZk2/view
- › Events: https://lfaidata.foundation/events/
- › Events Calendar https://wiki.lfaidata.foundation/pages/viewpage.action?pageId=12091544
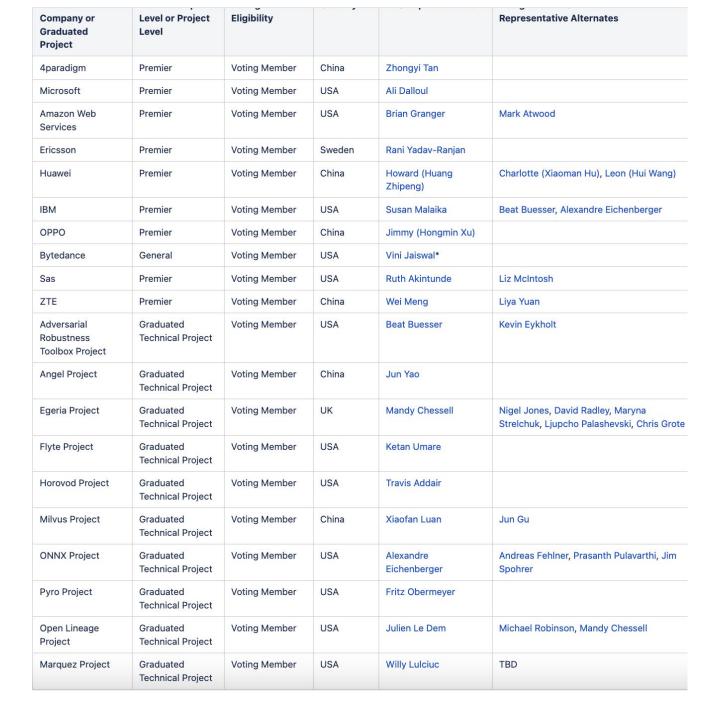- › Event Wiki https://wiki.lfaidata.foundation/pages/viewpage.action?pageId=10518553

LF AI & DATA

Agenda

› Roll Call (1 mins)

› Approval of Minutes from previous meeting (2 mins)

› Foundation Updates - Vini Jaiswal (10 mins)

› Open Source Initiative (OSI) Updates - Nick Vidal (15 mins)

› Generative AI Commons Updates - Matt White (5 mins)

› Open Discussion

TAC Voting Members

Note: we still need a few designated backups specified on [wiki](#)

Company or Graduated Project	Membership Level or Project Level	Voting Member Eligibility	Country	Representative	Designated Representative Alternates
4paradigm	Premier	Voting Member	China	Zhongyi Tan	
Microsoft	Premier	Voting Member	USA	Ali Dalloul	
Amazon Web Services	Premier	Voting Member	USA	Brian Granger	Mark Atwood
Ericsson	Premier	Voting Member	Sweden	Rani Yadav-Ranjan	
Huawei	Premier	Voting Member	China	Howard (Huang Zhipeng)	Charlotte (Xiaoman Hu), Leon (Hui Wang)
IBM	Premier	Voting Member	USA	Susan Malaika	Beat Buesser, Alexandre Eichenberger
OPPO	Premier	Voting Member	China	Jimmy (Hongmin Xu)	
Bytedance	General	Voting Member	USA	Vini Jaiswal*	
Sas	Premier	Voting Member	USA	Ruth Akintunde	Liz McIntosh
ZTE	Premier	Voting Member	China	Wei Meng	Liya Yuan
Adversarial Robustness Toolbox Project	Graduated Technical Project	Voting Member	USA	Beat Buesser	Kevin Eykholt
Angel Project	Graduated Technical Project	Voting Member	China	Jun Yao	
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	Andreas Fehlner, Prasanth Pulavarthi, Jim Spohrer
Pyro Project	Graduated Technical Project	Voting Member	USA	Fritz Obermeyer	
Open Lineage Project	Graduated Technical Project	Voting Member	USA	Julien Le Dem	Michael Robinson, Mandy Chessell
Marquez Project	Graduated Technical Project	Voting Member	USA	Willy Lulciuc	TBD

Minutes approval

Approval of March 7, 2024 Minutes

Draft minutes from the March 7, 2024, TAC call were previously distributed to the TAC members via the mailing list

Proposed Resolution:

› That the minutes of the March 7 2024, meeting of the Technical Advisory Council of the LF AI & Data Foundation are hereby approved.

LF AI & DATA

Foundation Updates

Election and Team Updates

General Member Board Representatives

Anni Lai
(FutureWei)

Junping Du
(DataStrato)

Nancy Rausch

Sr. Program Manager for Data Science at LF

Generative AI Commons Vice Chair

Arnaud Le Hors
(IBM)

Outreach Committee Chair

Richard Bian
(Ant Group)

Outreach Committee Kick off

› We are kicking off the Outreach Committee (OC) with the election of Richard Bian as the new Chair.
› The OC is responsible for marketing, communications, events, promotion, social media, etc. of LF AI & Data.

› An email invitation will go out to the LF AI & Data community announcing the call, which is open to anyone who wish to participate in this committee.

Action item:

Premier Members please provide the name/email of your voting representative to the LF AI & Data staff to be invited for the kick off call.

AI_Dev Events

19-20 June | Paris:
https://hubs.la/Q02l67Yg0

14-15 Oct | Seattle:
https://hubs.la/Q02l63_w0

Save the date:
23 Aug | Hong Kong
28-29 Oct | Tokyo

If you're interested in Sponsoring the event please reach out and we can work to announce and launch the event

Open Source Summit

- Main Conference - April 16-18
- Seattle, WA, US
- AI and Data mini Summit - April 15

Process for proposing a project for hosting in LF AI & Data

1. Contact ibrahim@linuxfoundation.org

2. Decide on a date to present to the TAC and request incubation

3. Ensure that your project implements these recommendations

4. Submit a formal request to incubate the project via a GH PR

5. Prepare deck and share with ED about 10 days prior to the presentation

6. Present to the TAC and get approval

7. Onboard the project with the LF AI & Data team and integrate the project with our services

8. Announce the project becoming hosted in LF AI & Data

LF AI & DATA

Open Source Initiative
The Open Source AI Definition

LF AI & DATA

The Open Source AI Definition

› The Open Source Initiative is driving a multi-stakeholder process to define an "Open Source AI" since 2022.

› The Open Source principles have demonstrated that massive benefits accrue to everyone when you remove the barriers to learning, using, sharing and improving software systems.

› We need essential freedoms to enable users to build and deploy AI systems that are reliable and transparent.

DEEP DIVE : AI

FINAL REPORT

open source initiative®

LF AI & DATA

The Four Essential Freedoms

› The freedom to run the program as you wish, for any purpose (freedom 0).

› The freedom to study how the program works, and change it so it does your computing as you wish (freedom 1). Access to the source code is a precondition for this.

› The freedom to redistribute copies so you can help others (freedom 2).

› The freedom to distribute copies of your modified versions to others (freedom 3). By doing this you can give the whole community a chance to benefit from your changes.

LF AI & DATA

GNU

ABOUT GNU **PHILOSOPHY** »

/ Philosophy / Essays & articles / About free software / Principles /

What is Free Software?

"Free software" means software that respects users' freedom and community. Roughly, it means that **the users have the freedom to run, copy, distribute, study, change and improve the software**. Thus, "free software" is a matter of liberty, not price. To understand the concept, you should think of "free" as in "free speech," not as in "free beer." We sometimes call it "libre software," borrowing the French or Spanish word for "free" as in freedom, to show we do not mean the software is gratis.

The Four Essential Freedoms applied to AI

To be Open Source, an AI system needs to be available under legal terms that grant the freedoms to:

› Use the system for any purpose and without having to ask for permission.

› Study how the system works and inspect its components.

› Modify the system to change its recommendations, predictions or decisions to adapt to your needs.

› Share the system with or without modifications, for any purpose.

◰LF AI & DATA

version 0.0.3

Leave comments for this text

About   Programs   Licenses   Open Source

stating the intentions of this document; the Definition of Open Source AI itself; and a checklist to evaluate licenses.
We follow the definition of AI adopted by UNESCO:

An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.

Preamble

Why we need Open Source Artificial Intelligence (AI)

Open Source has demonstrated that massive benefits accrue to everyone when you remove the barriers to learning, using, sharing and improving software systems. These benefits are the result of using licenses that adhere to the Open Source Definition. The benefits can be distilled to autonomy, transparency, and collaborative improvement.

Everyone needs these benefits in AI. We need essential freedoms to enable users to build and deploy AI systems that are reliable and transparent.

How we can get the benefits of Open Source AI

A precondition for a system to be Open Source software is that developers must have unrestricted access to the "preferred form to make modifications to the work".
For AI systems, the preferred form to make modifications to the work depends on the specific kind of AI.

    [Provide an example, based on machine learning?]

Out of scope issues

The Open Source AI Definition doesn't say how to develop and deploy an AI system that is ethical or responsible, although it doesn't prevent it. What makes an AI system ethical or responsible is a separate discussion.

What is Open Source AI

To be Open Source, an AI system needs to make its components available under licenses that individually grant the freedoms to:

• **Study** how the system works and inspect its components.

• **Use** the system for any purpose and without having to ask for permission.

• **Modify** the system to change its recommendations, predictions or decisions to adapt to your needs.

• **Share** the system with or without modifications, for any purpose.

    [Provide an example, based on machine learning?]

Checklist to evaluate licenses

TODO

Leave comments for this text

Getting the specifications

AI systems	List of components	Legal frameworks	Legal documents	**Checklist**
As defined by the OECD.	What elements are necessary to: - use - study - modify - share an AI system?	For each artifact, evaluate which laws apply. Some will be under "Intellectual Property" regimes, some will be under other regimes.	We'll match the components and the identified legal frameworks with the terms of the legal documents already in use, where available.	After repeating this exercise enough times, we'll be able to generalize the outcomes and write the specs to evaluate the freedoms granted.

The Model Openness Framework (MOF)

The Model Openness Framework: Promoting Completeness and Openness for Reproducibility, Transparency and Usability in AI
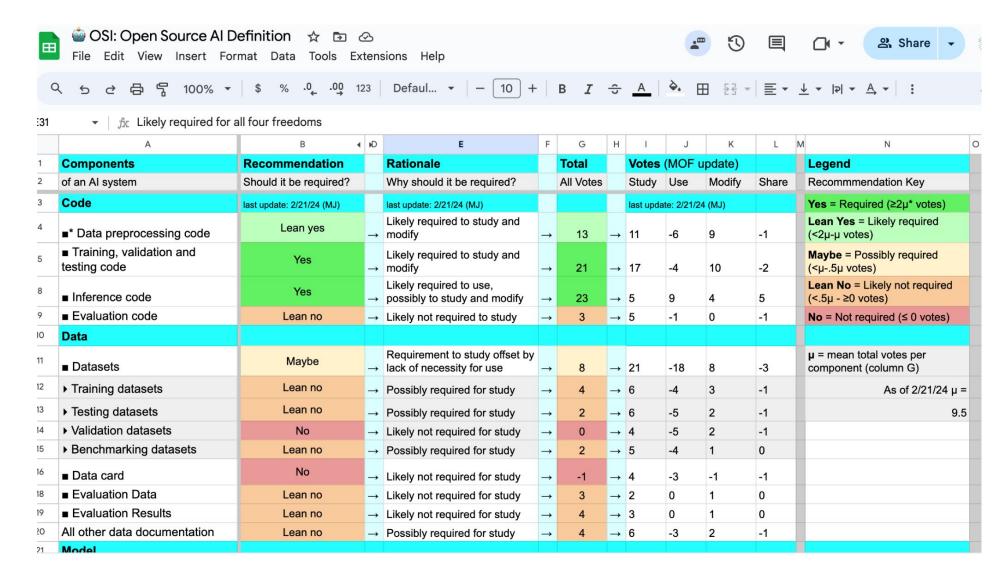
Matt White[1,2], Ibrahim Haddad[2], Cailean Osborne[2,3],
Xiao-Yang (Yanglet) Liu[1,4], Ahmed Abdelmonsef[1,5], Sachin Varghese[1]

[1] LF AI & Data - Generative AI Commons, [2] Linux Foundation,
[3] University of Oxford, [4] Columbia University, [5] IBM

matt.white@berkeley.edu, ibrahim@linuxfoundation.org,
cailean.osborne@oii.ox.ac.uk, xl2427@columbia.edu,
{ahmed.abdelmonsef,sachin.varghese}@genaicommons.org

LF AI & DATA

The Four Essential Freedoms
Evaluating each MOF Component

OSI: Open Source AI Definition

File Edit View Insert Format Data Tools Extensions Help

fx | Likely required for all four freedoms

	Components	Recommendation	Rationale		Total	Votes (MOF update)					Legend
	of an AI system	Should it be required?	Why should it be required?		All Votes	Study	Use	Modify	Share		Recommmendation Key
Code		last update: 2/21/24 (MJ)	last update: 2/21/24 (MJ)			last update: 2/21/24 (MJ)					**Yes** = Required (≥2μ* votes)
	▪* Data preprocessing code	Lean yes	→	Likely required to study and modify	→	13	→	11	-6	9	-1	**Lean Yes** = Likely required (<2μ-μ votes)
	▪ Training, validation and testing code	Yes	→	Likely required to study and modify	→	21	→	17	-4	10	-2	**Maybe** = Possibly required (<μ-.5μ votes)
	▪ Inference code	Yes	→	Likely required to use, possibly to study and modify	→	23	→	5	9	4	5	**Lean No** = Likely not required (<.5μ - ≥0 votes)
	▪ Evaluation code	Lean no	→	Likely not required to study	→	3	→	5	-1	0	-1	**No** = Not required (≤ 0 votes)
Data											
	▪ Datasets	Maybe	→	Requirement to study offset by lack of necessity for use	→	8	→	21	-18	8	-3	μ = mean total votes per component (column G)
	▸ Training datasets	Lean no	→	Possibly required for study	→	4	→	6	-4	3	-1	
	▸ Testing datasets	Lean no	→	Possibly required for study	→	2	→	6	-5	2	-1	As of 2/21/24 μ =
	▸ Validation datasets	No	→	Likely not required for study	→	0	→	4	-5	2	-1	9.5
	▸ Benchmarking datasets	Lean no	→	Possibly required for study	→	2	→	5	-4	1	0	
	▪ Data card	No	→	Likely not required for study	→	-1	→	4	-3	-1	-1	
	▪ Evaluation Data	Lean no	→	Likely not required for study	→	3	→	2	0	1	0	
	▪ Evaluation Results	Lean no	→	Likely not required for study	→	4	→	3	0	1	0	
	All other data documentation	Lean no	→	Possibly required for study	→	4	→	6	-3	2	-1	
Model											

The Four Essential Freedoms
Evaluating each MOF Component

- **Required**
  - Training, validation and testing code
  - Inference code
  - Model architecture
  - Model parameters
  - Supporting libraries and tools
- **Likely Required**
  - Data preprocessing code
- **Maybe Required**
  - Datasets
  - Usage documentation

- **Likely Not Required**
  - Evaluation code
  - Evaluation data
  - Evaluation results
  - All other data documentation
  - Model metadata
  - Model card
  - Research paper
  - Technical report
- **Not Required**
  - Data card
  - Sample model outputs

LF AI & DATA

Code Recommendations

Components	Recommendation		Rationale		Total		Votes (MOF update)			
of an AI system	Should it be required?		Why should it be required?		All Votes		Study	Use	Modify	Share
Code	last update: 2/21/24 (MJ)		last update: 2/21/24 (MJ)				last update: 2/21/24 (MJ)			
■* Data preprocessing code	Lean yes	→	Likely required to study and modify	→	13	→	11	-6	9	-1
■ Training, validation and testing code	Yes	→	Likely required to study and modify	→	21	→	17	-4	10	-2
■ Inference code	Yes	→	Likely required to use, possibly to study and modify	→	23	→	5	9	4	5
■ Evaluation code	Lean no	→	Likely not required to study	→	3	→	5	-1	0	-1

Data Recommendations

Components	Recommendation		Rationale		Total		Votes (MOF update)			
of an AI system	Should it be required?		Why should it be required?		All Votes		Study	Use	Modify	Share
Data										
■ Datasets	Maybe	→	Requirement to study offset by lack of necessity for use	→	8	→	21	-18	8	-3
▸ Training datasets	Lean no	→	Possibly required for study	→	4	→	6	-4	3	-1
▸ Testing datasets	Lean no	→	Possibly required for study	→	2	→	6	-5	2	-1
▸ Validation datasets	No	→	Likely not required for study	→	0	→	4	-5	2	-1
▸ Benchmarking datasets	Lean no	→	Possibly required for study	→	2	→	5	-4	1	0
■ Data card	No	→	Likely not required for study	→	-1	→	4	-3	-1	-1
■ Evaluation Data	Lean no	→	Likely not required for study	→	3	→	2	0	1	0
■ Evaluation Results	Lean no	→	Likely not required for study	→	4	→	3	0	1	0
All other data documentation	Lean no	→	Possibly required for study	→	4	→	6	-3	2	-1

Model Recommendations

Components	Recommendation		Rationale		Total		Votes (MOF update)			
of an AI system	Should it be required?		Why should it be required?		All Votes		Study	Use	Modify	Share
Model										
■ Model architecture	Yes	→	Possibly required to study and modify	→	20	→	9	0	9	2
■ Model parameters	Yes	→	Possibly required for all four freedoms	→	29	→	8	7	9	5
■ Model Metadata	Lean no	→	Likely not required for study	→	1	→	1	0	0	0
■ Model card	Lean no	→	Likely not required for study	→	1	→	2	0	0	-1
■ Sample model outputs	No	→	Likely not required for study	→	-3	→	2	-4	0	-1

Other Recommendations

Components	Recommendation		Rationale		Total		Votes (MOF update)			
of an AI system	Should it be required?		Why should it be required?		All Votes		Study	Use	Modify	Share
Other										
■ Research paper	Lean no	→	Possibly required for study	→	1	→	5	-3	0	-1
Usage documentation	Maybe	→	Likely not required for all four freedoms	→	9	→	2	2	3	2
■ Technical report	Lean no	→	Likely not required for study	→	3	→	2	0	1	0
■ Supporting [Libraries and*] Tools	Yes	→	Likely required for all four freedoms	→	50	→	10	16	13	11
* ■ = Model Openness Framework (MOF) components (as of 2/14/24)	Average (μ)			→	9.5					

2024 timeline

System testing work stream

Stakeholder consultation work stream

Release schedule

	February	March	April	May	June …	… October
System testing	Call For Volunteers + Activity Feedback and Revision	Virtual System Review Meetings Begin	Virtual System Review Meetings Continue	Virtual System Review Meetings END	Feedback Informs Content of OSI In-Person Stakeholder Meeting	Monthly Virtual Meetings
Stakeholder consultation	Bi-Weekly Virtual Public Townhalls	Bi-Weekly Virtual Public Townhalls	Bi-Weekly Virtual Public Townhalls	Bi-Weekly Virtual Public Townhalls	Townhall + **OSI In-Person Stakeholder Meeting** (date + place TBD)	**Release version 1.0**
Release schedule	Draft 0.0.5	Draft 0.0.6	Draft 0.0.7	Draft 0.0.8	RC1	v. 1.0

How can the LF AI & Data help?

› We need people who can commit time to follow the evolution of the draft

› We need help from partners to promote discussions online and at in person events so that we can reach a consensus on what constitutes "Open Source AI"

› We're looking for partners to co-sign the announcement and be ready between June and October to say "We like this, we support this definition"

Partnership with the LF AI & Data

Panels, workshops, talks, interviews and blog posts:

› [Panel: Does AI Change Everything? What is Open? Liability, Ethics, Values?](#)

› [Panel: Why Open Source AI Matters: The EU Community & Policy Perspective](#)

› [Panel: Why a Universal Definition of 'Open Source AI' is Essential for Humanity](#)

› [Panel: Why Open Source AI Matters: The Community & Policy Perspective](#)

› [Blog: Adapting the Definition of Open Source to AI: the Quest for the Holy Grail](#)

Join discuss.opensource.org

Draft v. 0.0.6 at opensource.org/deepdive

open source initiative®

About   Programs   Licenses   Blog   Open Source Definition   Join   Q

Home / Deep Dive / Drafts of the Open Source AI Definition / The Open Source AI Definition – draft v. 0.0.6

The Open Source AI Definition – draft v. 0.0.6

version 0.0.6

Leave comments for this text

Note: This document is made of three parts: A preamble,
stating the intentions of this document; the Definition of
Open Source AI itself; and a checklist to evaluate
licenses.

LF AI & DATA

Generative AI Commons Updates
Matt White (LF AI & Data)

Generative AI Commons Workstreams

Models and Data

- Hosting models and weights
- Hosting ethical datasets
- Data processing tools.
- Benchmarks and Reports
- Training and inference code

Applications

- Hosting AI Application frameworks
- Databases (vector stores)
- Agent frameworks
- Interface and metadata standards

Frameworks

- Generative AI adoption
- Reference architectures
- Model Openness Framework
- Best practices and guidelines
- Compliance in code

Education & Outreach

- Education and training
- Thought leadership
- Educational outreach
- Legislative representation
- State of Open-Source

Responsible AI

- Responsible AI
- Security, Privacy and Safety
- Informing Policy
- Copyright Issues
- Model and Data Lineage

Generative AI Commons Updates

› Generative AI Commons membership now at over 200 active participants and 80 member companies

› Joined NIST AI Safety Consortium with representation in all 5 working groups.

› Responsible AI workstream kicked off

› Model Openness Framework published to arXiv, and setup placeholder website http://isitopen.ai

› Assisting RWKV with World Data Set project and training of 10B RNN

› Launching Generative AI Glossary and generative AI developer survey

› Planning to work with Mozilla and OSI on open-source and open AI initiatives

› Several talks planned at AI + Data Forum at OSS NA as well as meetup and booth

› Working on Responsible AI Framework to publish as standard

› CFP closed for AI_Dev EU in Paris. Reviewing 130+ submissions.

Open Discussion

Upcoming TAC Meetings

➜ April 4, 2024

➜ April 18, 2024

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

LF AI & DATA

TAC Meeting Details

TAC Biweekly Meeting LF AI & Data

Ways to join meeting:

1. Join from PC, Mac, iPad, or Android

https://zoom-lfx.platform.linuxfoundation.org/meeting/95332329356?password=c708f2ee-fb78-4a12-91a3-47daa19b708f

2. Join via audio

One tap mobile:

US: +12532158782,,95332329356# or +13462487799,,95332329356

Or dial:

US: +1 253 215 8782 or +1 346 248 7799 or +1 669 900 6833 or +1 301 715 8592 or +1 312 626 6799 or +1 646 374 8656 or 877 369 0926 (Toll Free) or 855 880 1246 (Toll Free)

Canada: +1 647 374 4685 or +1 647 558 0588 or +1 778 907 2071 or +1 204 272 7920 or +1 438 809 7799 or +1 587 328 1099 or 855 703 8985 (Toll Free)

Meeting ID: 95332329356

Meeting Passcode: 040721

Legal Notice

LF AI & DATA

07MAR2024