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

November 3, 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)
- Docarray project review (40 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 (8 out of 16 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 | Jimmy (Hongmin Xu) | |
| SAS | Premier | Voting Member | USA | *Nancy Rausch | JP Trawinski |
| ZTE | Premier | Voting Member | China | Wei Meng | Liya Yuan |
| 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

# Minutes approval

LF AI & DATA

# Approval of October 20, 2022 Minutes

Draft minutes from the October 20 TAC call were previously distributed to the TAC members via the mailing list

**Proposed Resolution:**

- That the minutes of the October 20 meeting of the Technical Advisory Council of the LF AI & Data Foundation are hereby approved.



# DocArray @ LF AI

# CONTENTS

# Why donate DocArray to LF AI?

01

01

# Why donate DocArray to LF AI?

- Open our governance
- Make DocArray vendor-neutral

Motivation

# Open our governance

- Encourage collaboration between different vector databases.
- Be more transparent in decision making and broaden the circle of decision-making contributors.
- Speed up development of DocArray

Motivation

# Make DocArray vendor-neutral

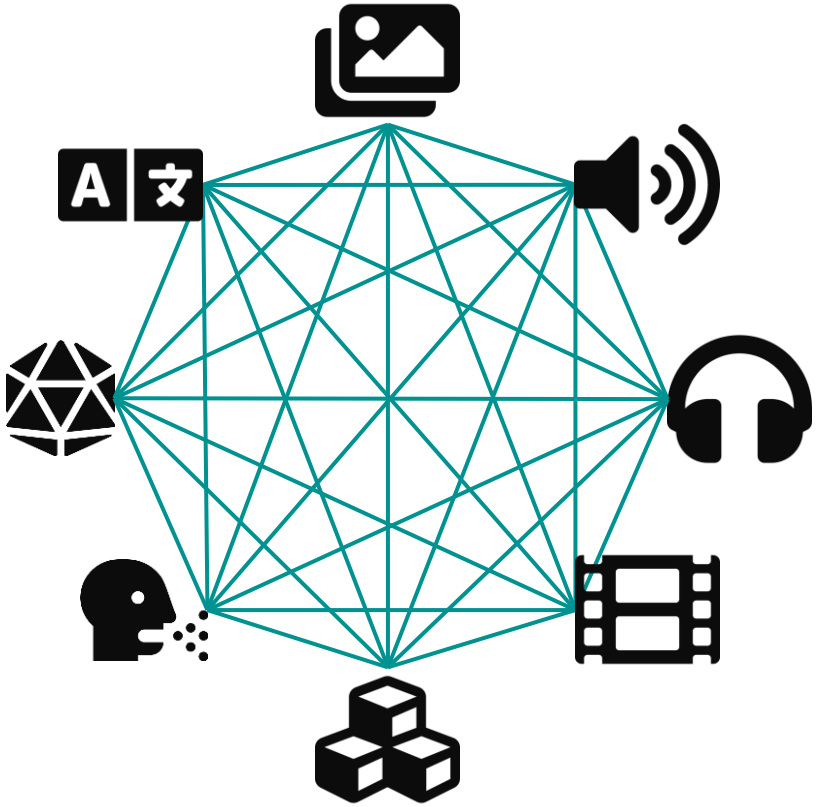
- Not for profit.
- Build synergy with other OSS projects.
- Increase community contributors.

# Problem

02

Background

# Multimodal is the Future of AI

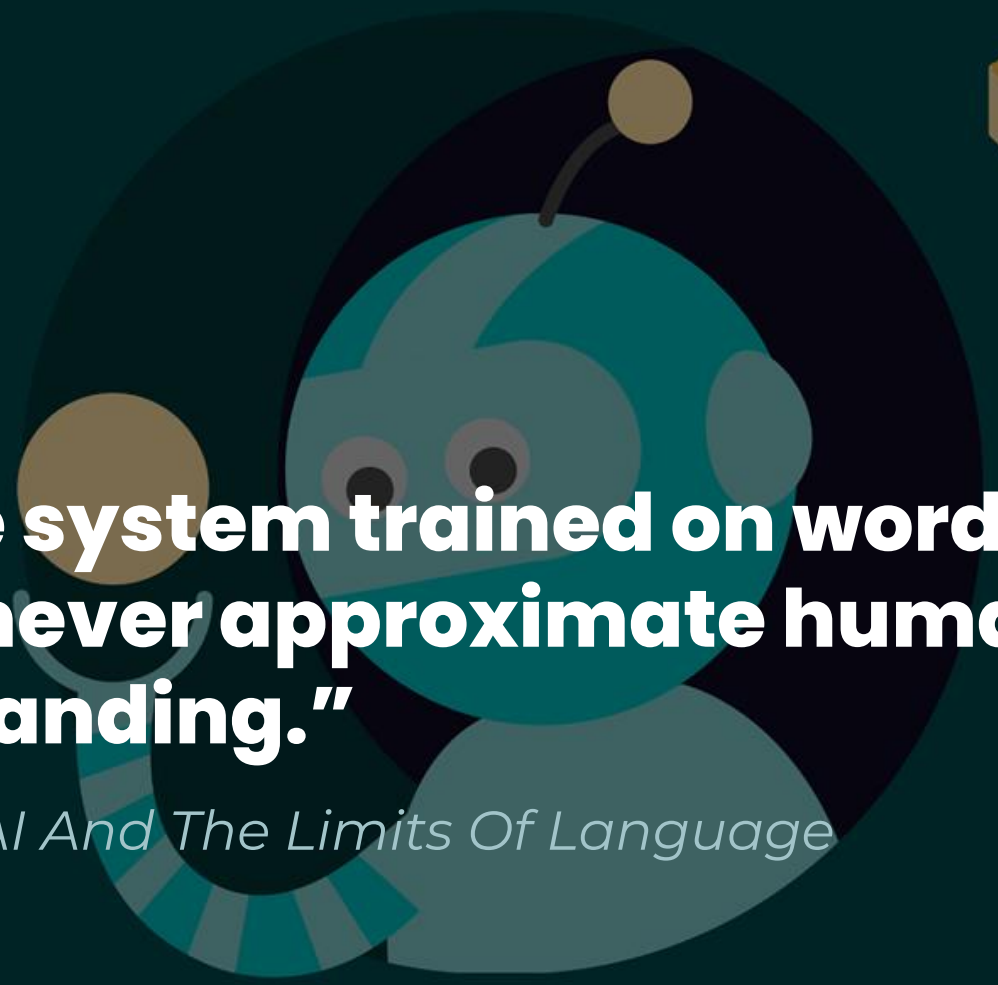


OpenAI: CLIP (2021), DallE 2 (2022)

DeepMind: Gato (2022)

Meta AI: Omnivore (2022), Data2Vec (2022)

Google Brain: Imagen (2022)



**“An artificial intelligence system trained on words and sentences alone will never approximate human understanding.”**

Y. Lecun in 2022 in *AI And The Limits Of Language*

# 02

- Multimodal is the future of AI, but the machine learning ecosystem is not (yet) suited for it

# Problem

- Working with multiple modalities simultaneously is complex.
- Explosion in numbers of vector databases but no universal client.

# Working with multiple modalities is complex

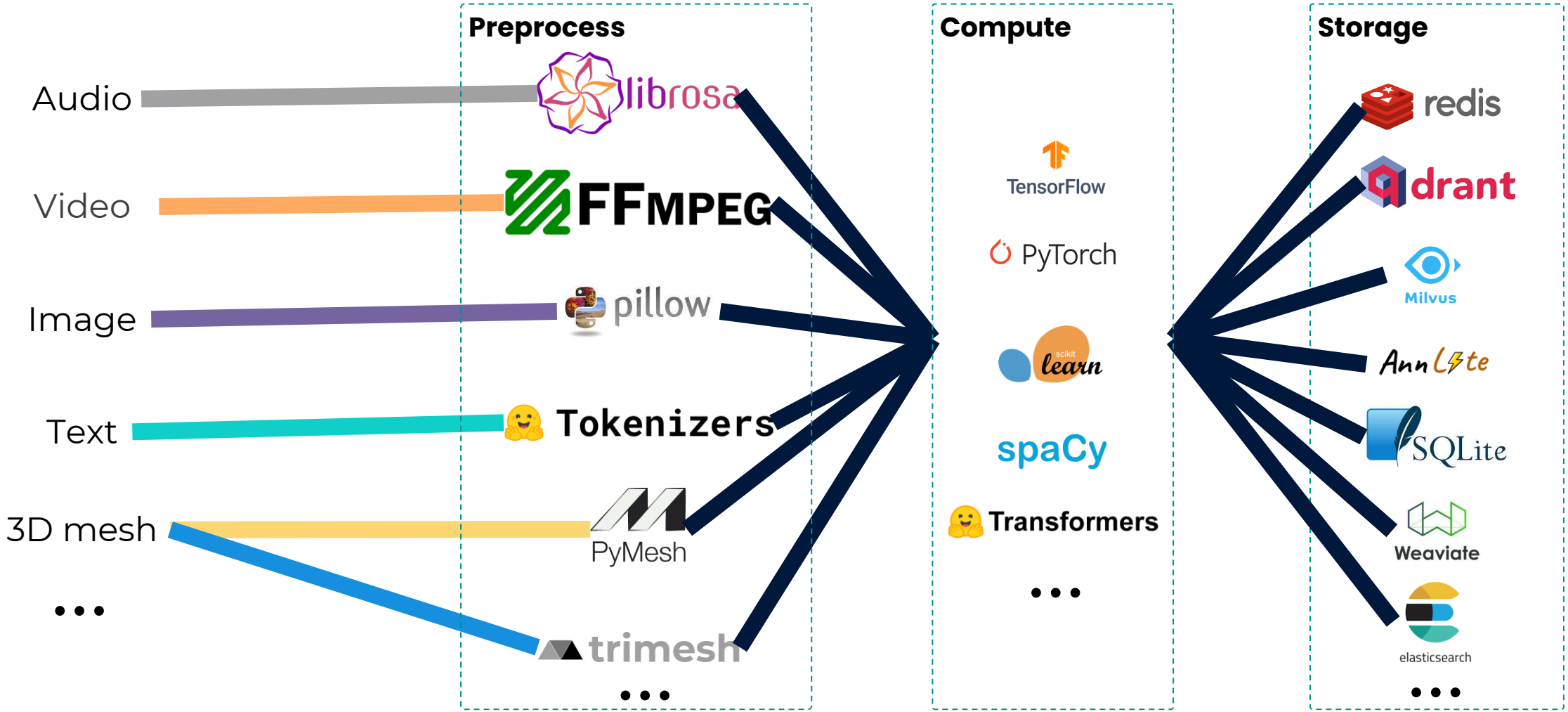
Problems

# Working with multimodal data is complex

- Lack of common interface for different modalities makes it difficult to work with multiple modalities at the same time.
- No easy way to represent unstructured and nested multimodal data.

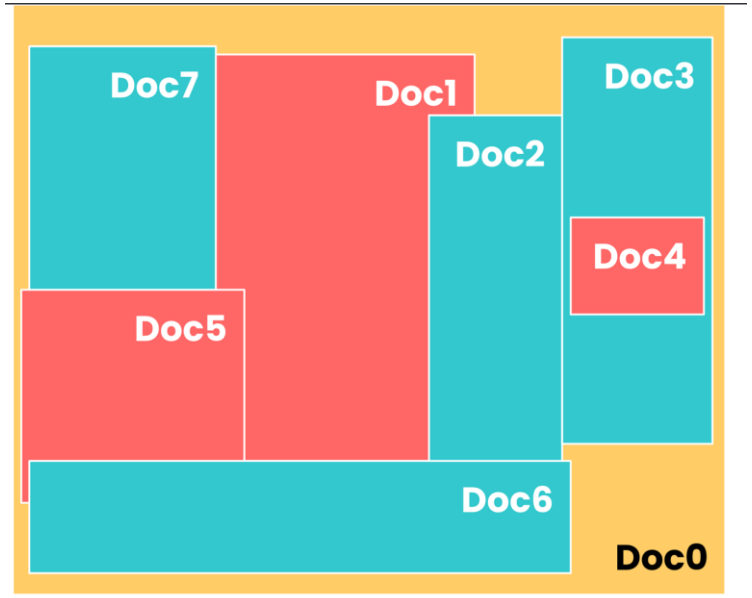
Problems

# Lack of common interface



Problems

# No easy way to represent unstructured nested multimodal data


- Unstructured document
- Nested content
- Different modalities (text and image)

.jina

# Storing nested data with databases is complicated

Problems

# Storing nested data with databases is complicated

- Complex and nested schema are not directly supported in databases
- Various (vector) databases with different APIs.

# DocArray

03

**DocArray** is a Python **library for nested, unstructured, multimodal data,** including text, image, audio, video, 3D mesh, etc

# What is DocArray?

🧑‍🔬 **Data science powerhouse**: Greatly accelerate data scientists' work with **multimodal data** and embeddings.

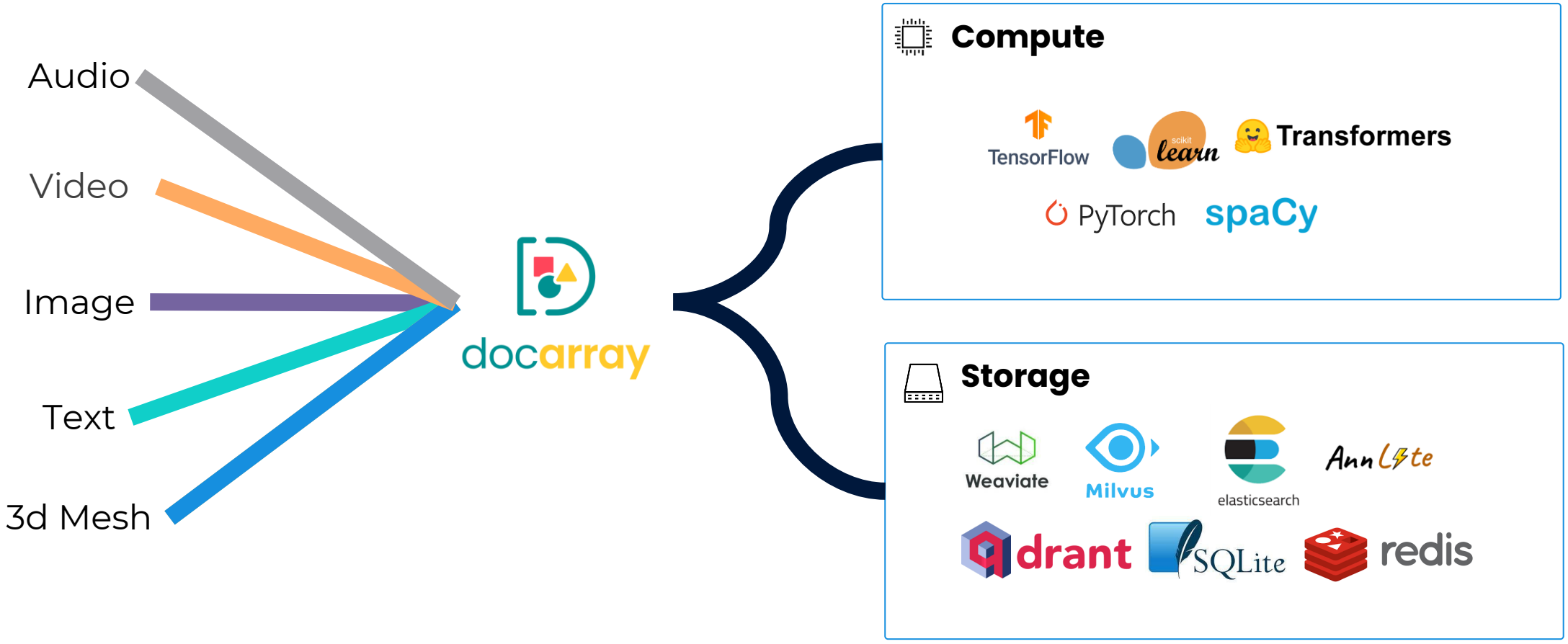
🔎 **Vector search**: Unified and consistent API for mainstream vector databases that allows nearest neighbour search including Elasticsearch, Redis, ANNLite, Milvus, Qdrant, Weaviate.

🚎 **Data in transit**: Optimized for network communication, ready-to-wire at any time with fast and compressed serialization in Protobuf, bytes. Perfect for streaming and out-of-memory data.

DocArray

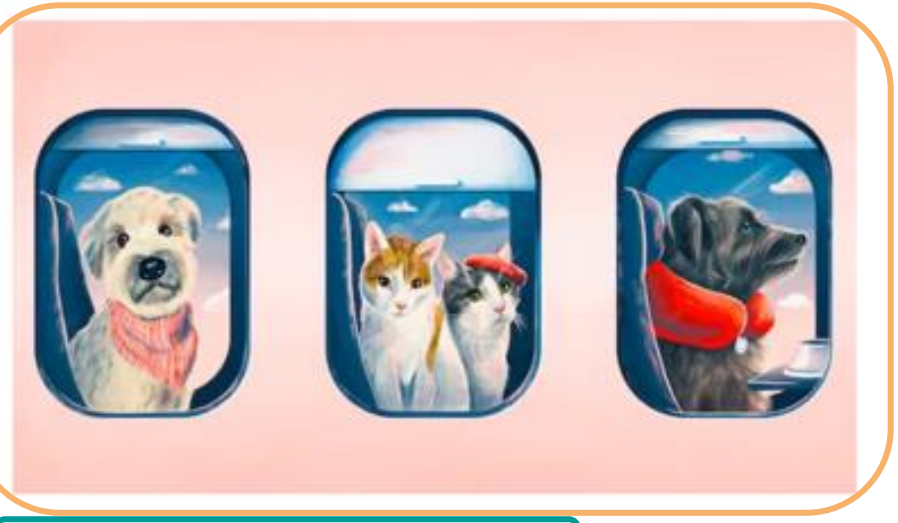
# Data science powerhouse

- Provide a common interface for different modalities



DocArray

# Data science powerhouse



**By the Way** A Post Travel Destination

**Everything to know about flying with pets, from picking your seat to keeping your animal calm**

By Nathan Diller

- Support to unstructured nested multimodal data
- Flexible ways to represent data. DocArray adapts to your data rather than forcing you to adapt your data to DocArray.

```
@dataclass
class WPArticle:
    banner: Image
    headline: Text
    meta: JSON
```

DocArray

# Vector Search

- Translate nested data schema to each database
- Provide consistent APIs for vector search across all storage backends


docarray

**Storage**

Weaviate

Milvus

elasticsearch

AnnLite

qdrant

SQLite

redis

What is DocArray?

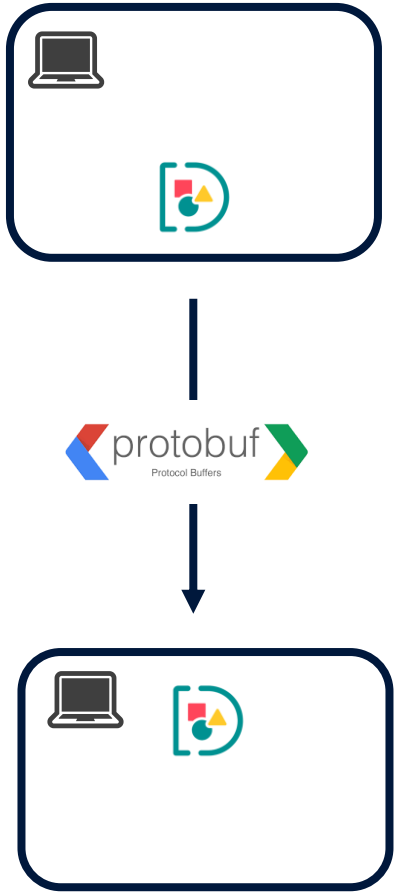
# Data in transit

DocArray is designed to be sent over the wire.

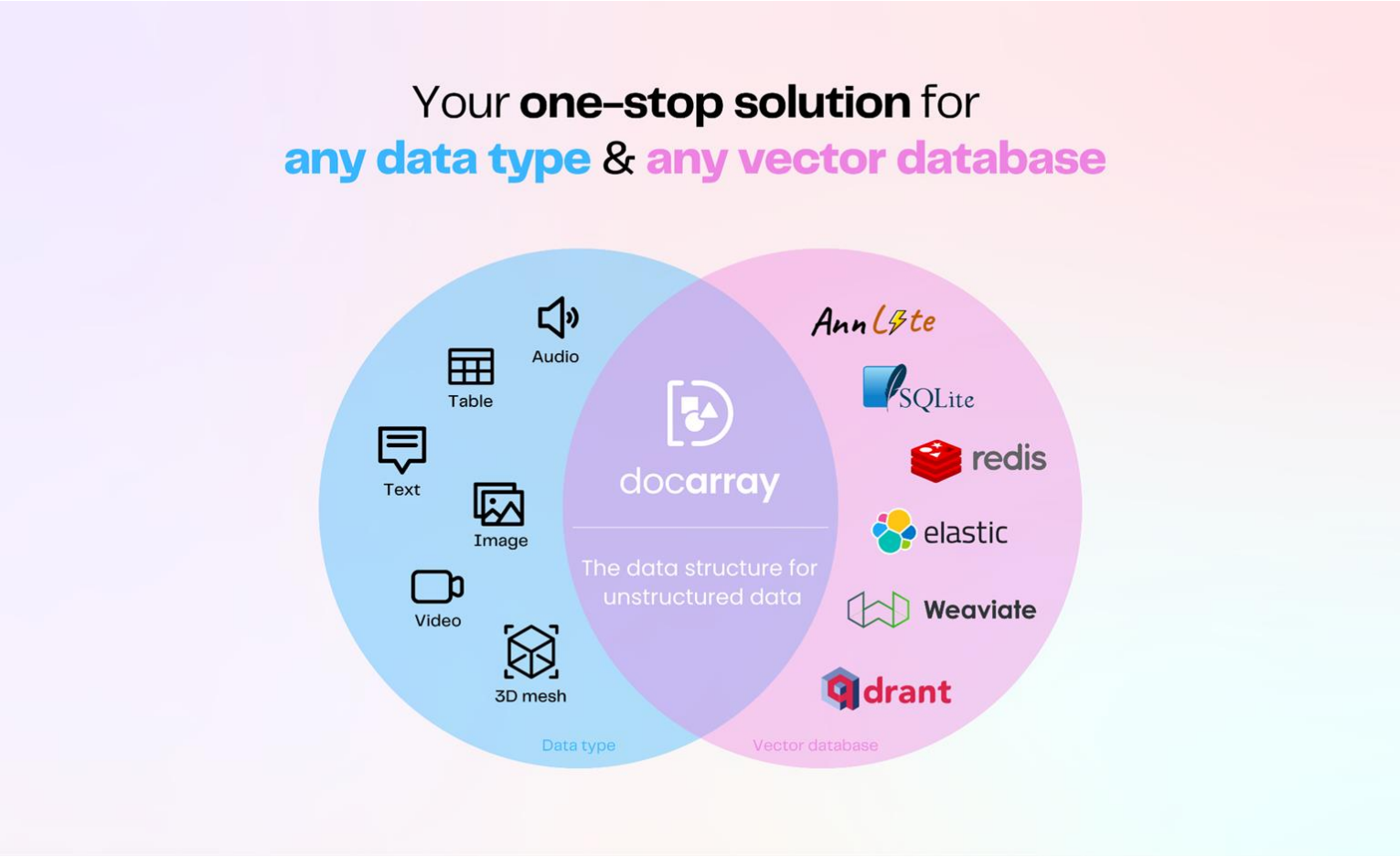
It can serve as a messaging format between microservices.

The underlying technologies are:

- **Protobuf** for fast programming language agnostic serialisation
- **gRPC** to send data over the wire



# DocArray in a nutshell



# Impact

04

Impact

# Who uses DocArray?

**1,200+**
GitHub stars

**40+**
Contributors

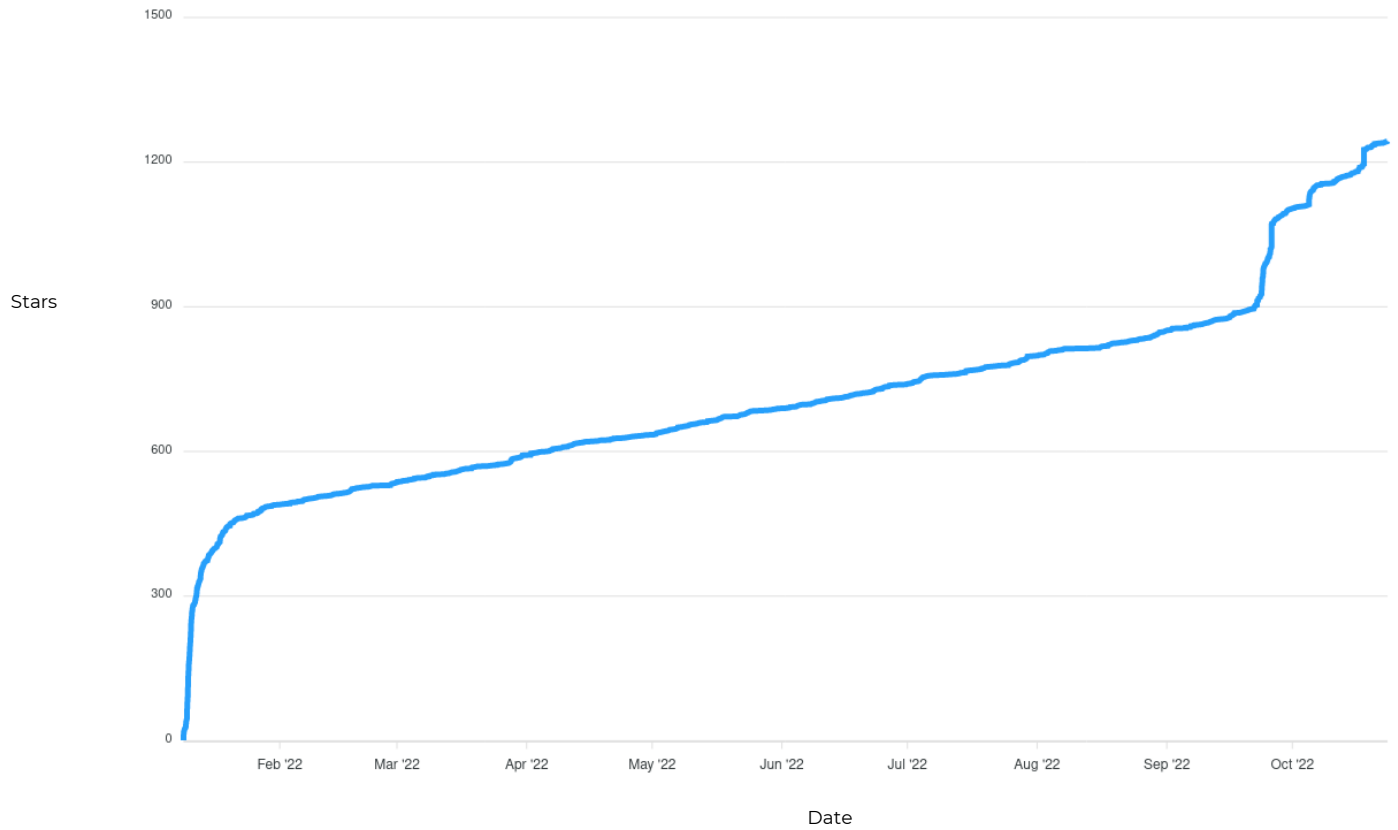
**2,000+**
Monthly active docs users

**100,000+**
Monthly downloads

**Trusted by companies of all sizes and needs**

Impact

# ☆ Star growth



LF AI & DATA

# Recent Contributions from the community

- Add Redis as a new storage backend (AnneYang)
- Improve SQLlite storage backend (linkerlin)
- Improve ElasticSearch storage backend (alphinside)
- Improve Weaviate storage backend (etiennedi, jjlatval)
- Improve Qdrant storage backend (generall)
- Fix the issue #393 (lhr0909)
- Fix documentation (k-zehnder)

LF AI & DATA

# Potential Collaborations in LF AI & DATA

- Add Milvus as a new storage backend
- Integrate with ONNX
- Integrate with Feast
- Integrate with Ludwig




docarray

LF AI & DATA

**Looking forward to your support in welcoming DocArray as an incubation project in the Sandbox Stage!**

# Q & A

# Proposal

**Proposed Resolution:**

- Proposal to add Docarray as a project as a new incubation project at the sandbox level to the LF AI&Data Foundation.

# TAC Open Discussion

LF AI & DATA

# Upcoming TAC Meetings

LF AI & DATA

# Upcoming TAC Meetings

- November 17, 2022 – Claimed, new sandbox proposal
- December 1, 2022 – Xtreme1, new sandbox proposal

Please note we are always open to special topics as well.

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