Status of GraphNeT 2.0

Rasmus Ørsøe / 18th of August 2025, Copenhagen

GraphNeT

Overview of the talk

Presentation is broken into four parts

Origin of GraphNeT

GraphNeT 2.0

GraphNeT Today

Future of GraphNeT

Origin of GraphNeT

Neutrino Telescopes

Observations Preluding GraphNeT

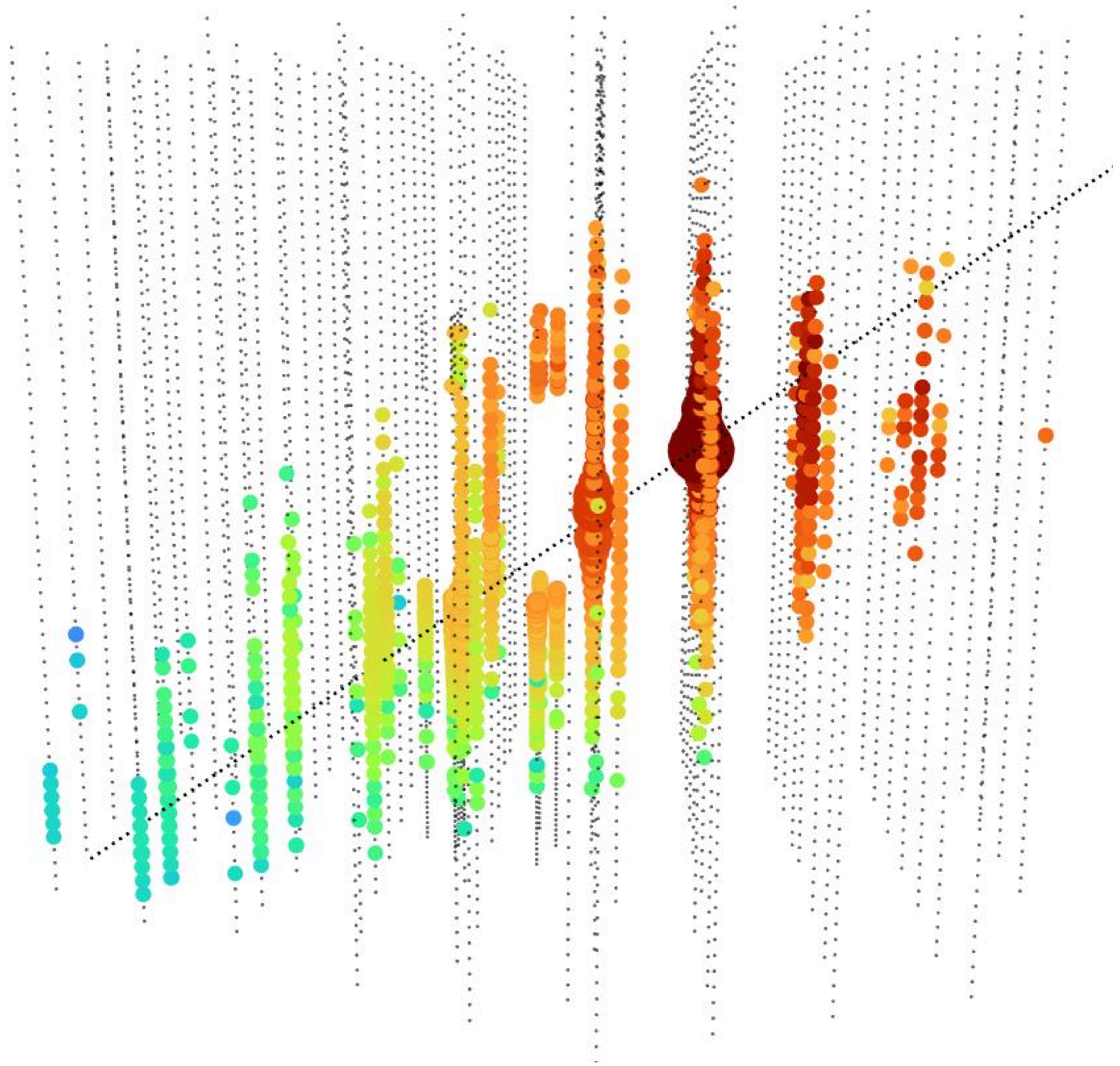
Low-level observations are largely identical in structure across experiments

Reconstruction needs are very similar across experiments

Deep-learning methods are universal function approximators, depending primarily on data structure

The adoption of deep-learning techniques is increasing - much remains unknown

Model development is largely silo´ed efforts, leading to duplicated work and comparison challenges



Illustrations of simulated 71 TeV track
Courtesy of Jorge Prado

Preliminary question

In 2020:

“Why are we not working together?”

“Why are we not working together?”

Incompatible codebases

Perfectly detector-agnostic methods are developed assuming a particular experiment, data representation and a narrow range of problems.

Lack of open-source datasets

Large-scale datasets suitable for training models are locked behind closed-source policies, making cross-experimental collaboration difficult.

"Why are we not working together?"

Incompatible codebases

Perfectly detector-agnostic methods are developed assuming a particular experiment, data representation and a narrow range of problems.

We can solve this with a python library that houses boilerplate code, models, etc. of common interest

Lack of open-source datasets

Large-scale datasets suitable for training models are locked behind closed-source policies, making cross-experimental collaboration difficult.

“Why are we not working together?”

Incompatible codebases

Perfectly detector-agnostic methods are developed assuming a particular experiment, data representation and a narrow range of problems.

We can solve this with a python library that houses boilerplate code, models, etc. of common interest

Lack of open-source datasets

Large-scale datasets suitable for training models are locked behind closed-source policies, making cross-experimental collaboration difficult.

We can write Santa 

Problem statement in GraphNeT

“We want a library where it is easy to:

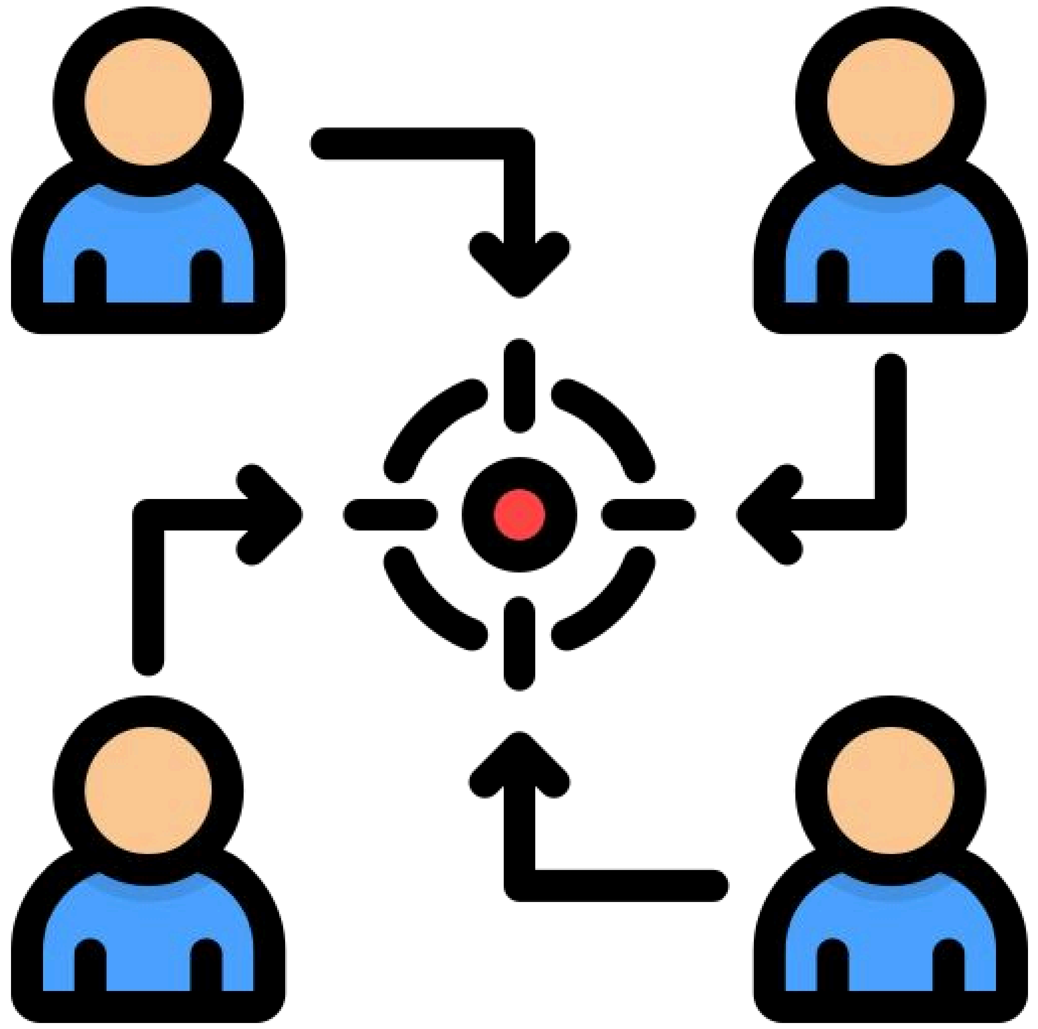
a) use models from one experiment in another
b) adjust models to perform new tasks
c) contribute with new, relevant techniques

that also contains boiler-plate code to reduce redundant efforts.”

Boiler-plate

What defines boiler-plate?

Functionality that everyone needs, and where the benefit of reproducing the code independently is very low.



For example:

- Typical training and inference loops
- Dataloading code
- Loss functions

Contributing with new, relevant tecniques

By making the library open-source, using well-known autodifferentiation frameworks and imposing mindful structure, the library would be accessible to a wider audience

Contribution guide

Outlining conventions and expectations on contributions.

- *“What is considered a relevant contribution?”*
- *“When is the contribution meeting expectations?”*

Pull request review

Each contribution is reviewed for quality and relevance to ensure a homogenized code base and consistent user experience.

GitHub

Model re-usability

What makes reusing models across experiments hard?

The contents of a model can be broken down into categories:

Experiment-specific details

Assumptions, code, conventions, needs that are specific to a single experiment or detector

Data Representations

The way raw observations are presented as input data to the model

Model Architecture

The part that maps input data to latent representations

Reconstruction Task(s)

Specific ways of mapping latent representations to final predictions

Model re-usability

What makes reusing models across experiments hard?

Boundaries between categories are often ill-defined

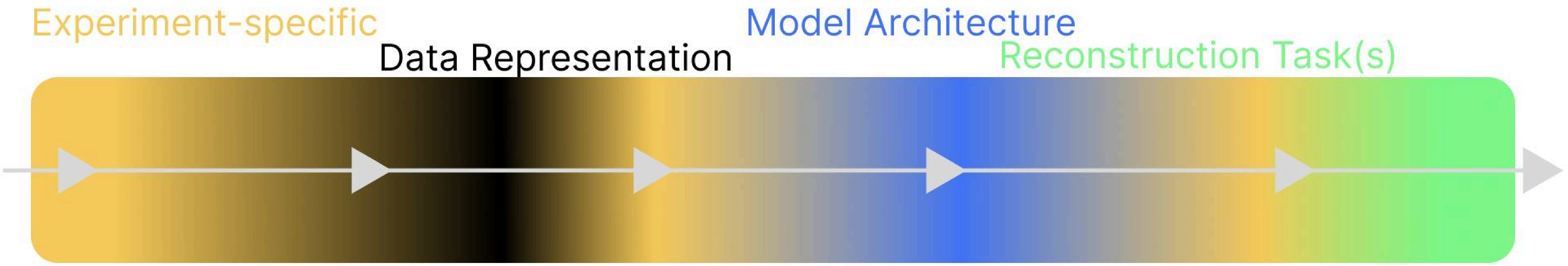


Diagram of model with liquid boundaries

Model re-usability

What makes reusing models across experiments hard?

Boundaries between categories are often ill-defined

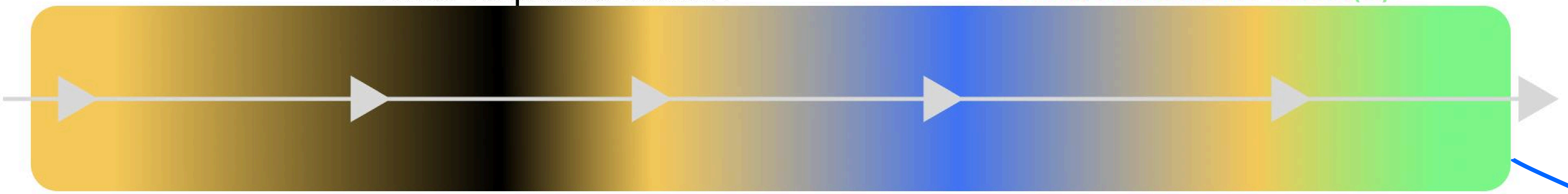


Diagram of model with liquid boundaries

Solved with conventions

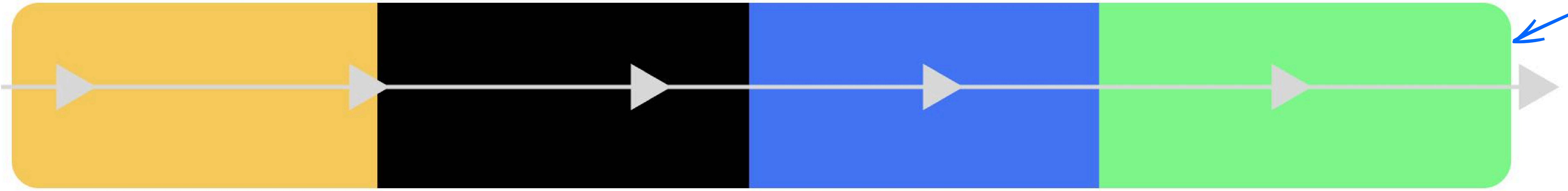


Diagram of model with strict boundaries

Model re-usability

Graphs provide an abstract, detector-agnostic representation

GNNs were very “in” prior to LLM explosion in 2021
(considered generalized CNNs)

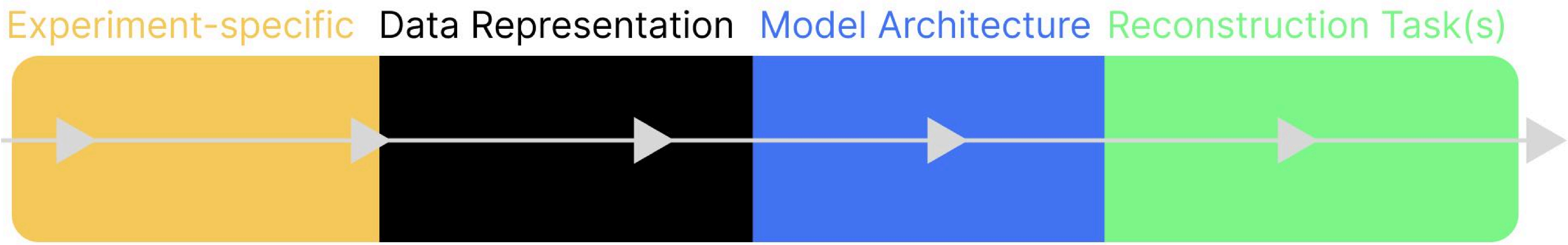


Diagram of model with strict boundaries

Model re-usability

Graphs provide an abstract, detector-agnostic representation

GNNs were very “in” prior to LLM explosion in 2021
(considered generalized CNNs)

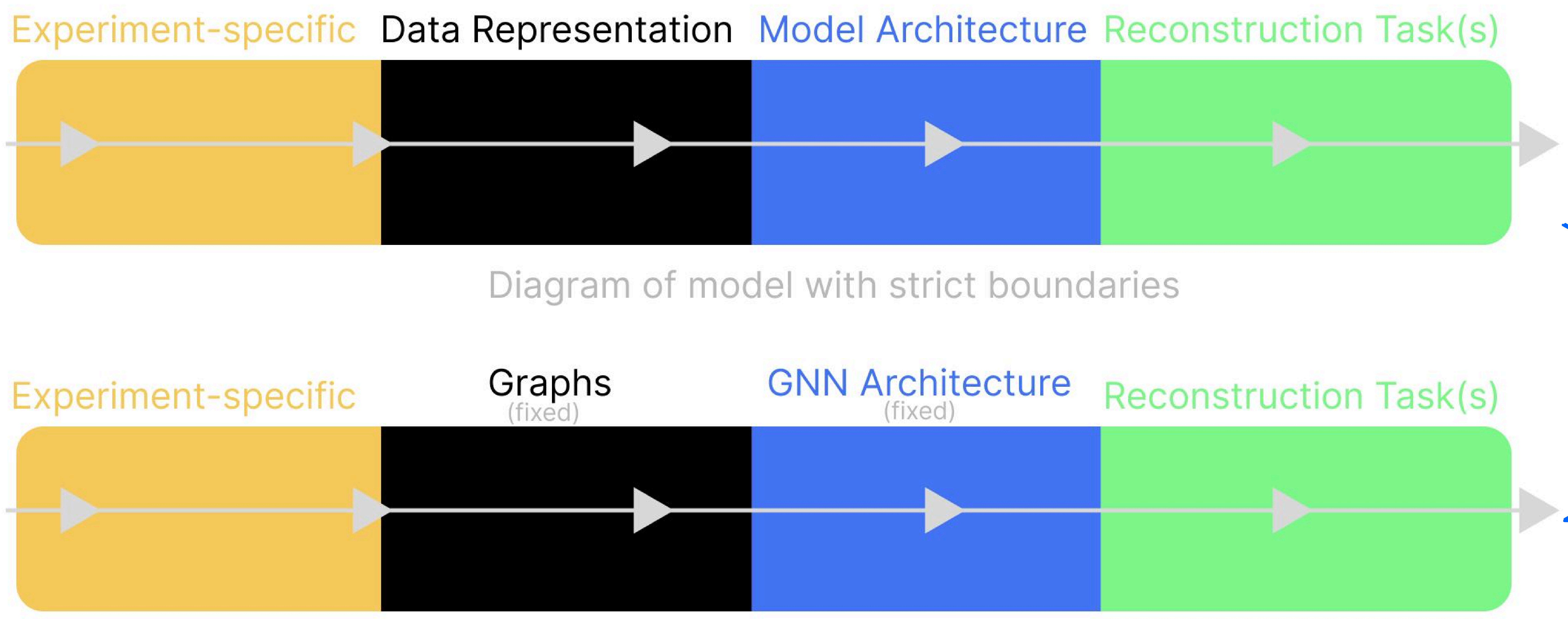


Diagram of model with strict boundaries

Diagram of GNN-specific model with strict boundaries

Technical challenges simplified by assuming graph representation and GNNs

GraphNeT 1.0

These reflections and decisions formed the essence of the Marie Skłodowska-Curie proposal by **Andreas Søgaard** in 2020:

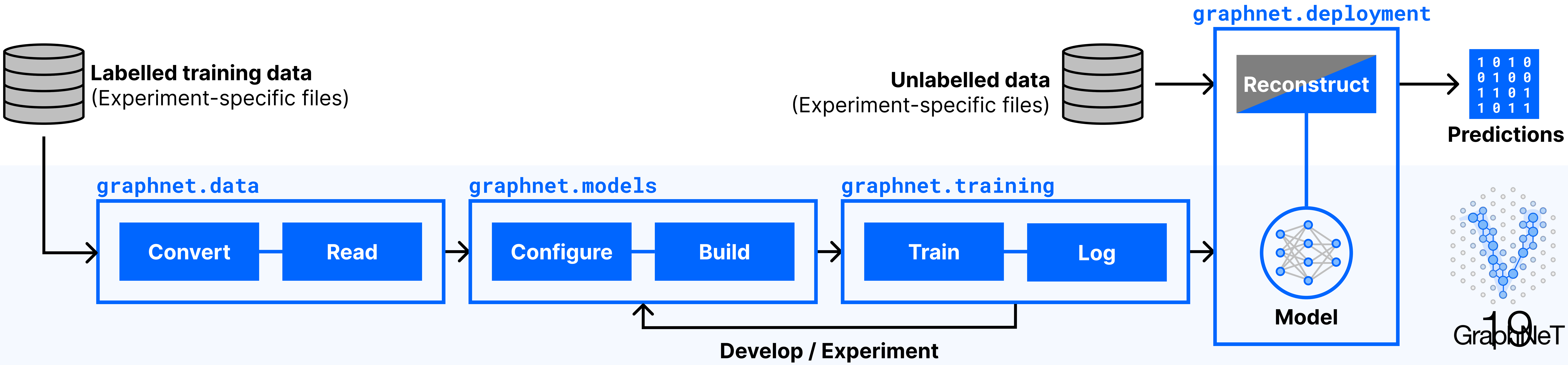
“**Graph** convolutional neural networks for **ne**utrino **t**elescopes”

Part of EU Horizon 2020, proposal here

He designed and lead the technical development of GraphNeT from September 2021 to May 2023 as a post-doc at NBI.

GraphNeT 1.0

An open-source library for neutrino telescopes



GraphNeT 1.0

These efforts culminated in our first “large” workshop in 2023 with the first stable release.

Participants helped define the next steps towards 2.0

GraphNeT 2.0

Takeaways from workshop in 2023

- **Implementing new experiments was hard**
  People struggle with converting their experiment-specific files to suitable formats and writing dataset classes

→ Data conversion is boiler-plate too

Takeaways from workshop in 2023

- **Implementing new experiments was hard**
  People struggle with converting their experiment-specific files to suitable formats and writing dataset classes

  → Data conversion is boiler-plate too

- **Growing interest in transformers**
  Following the IceCube open-data challenge, it had become clear that next-gen reconstruction techniques would likely rely on transformers in one way or another

  → Add support for major deep-learning paradigms

Takeaways from workshop in 2023

- **Implementing new experiments was hard**
  People struggle with converting their experiment-specific files to suitable formats and writing dataset classes

→ Data conversion is boiler-plate too

- **Growing interest in transformers**
  Following the IceCube open-data challenge, it had become clear that next-gen reconstruction techniques would likely rely on transformers in one way or another

→ Add support for major deep-learning paradigms

- **Good issues created but few followed up**
  Participants that assigned themselves to active issues did often not commit fully. We needed something to keep people engaged beyond the workshop.

→ More effort in community building needed - contribution procedure should be explained better - more workshops

Data Conversion is boiler-plate too

Solution:
Detector- and format-agnostic conversion code

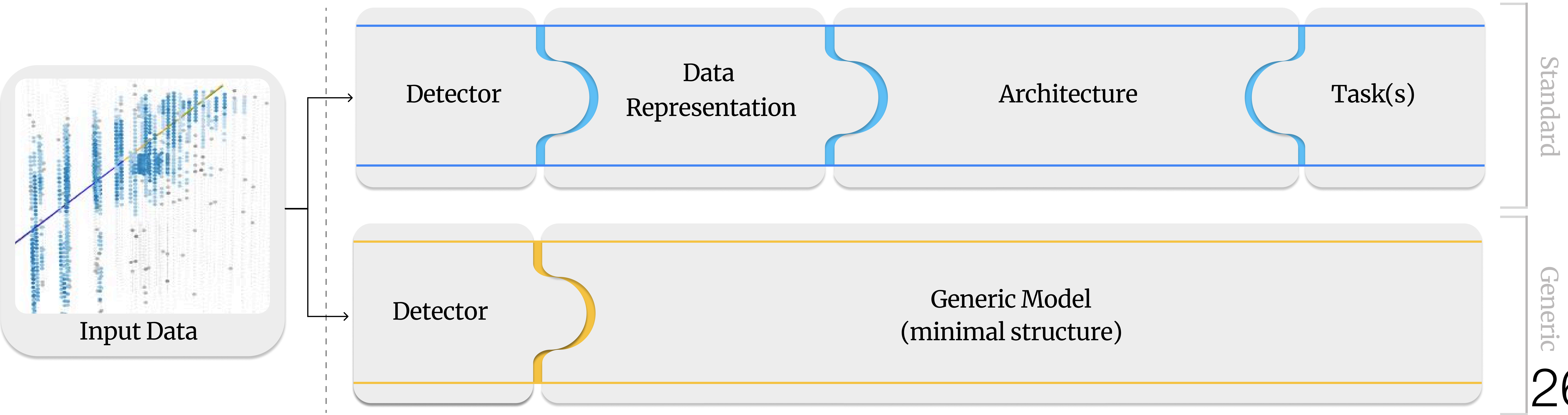
Can be extended to include new experiments and file formats

Significantly lowers the technical threshold of integrating a new experiment

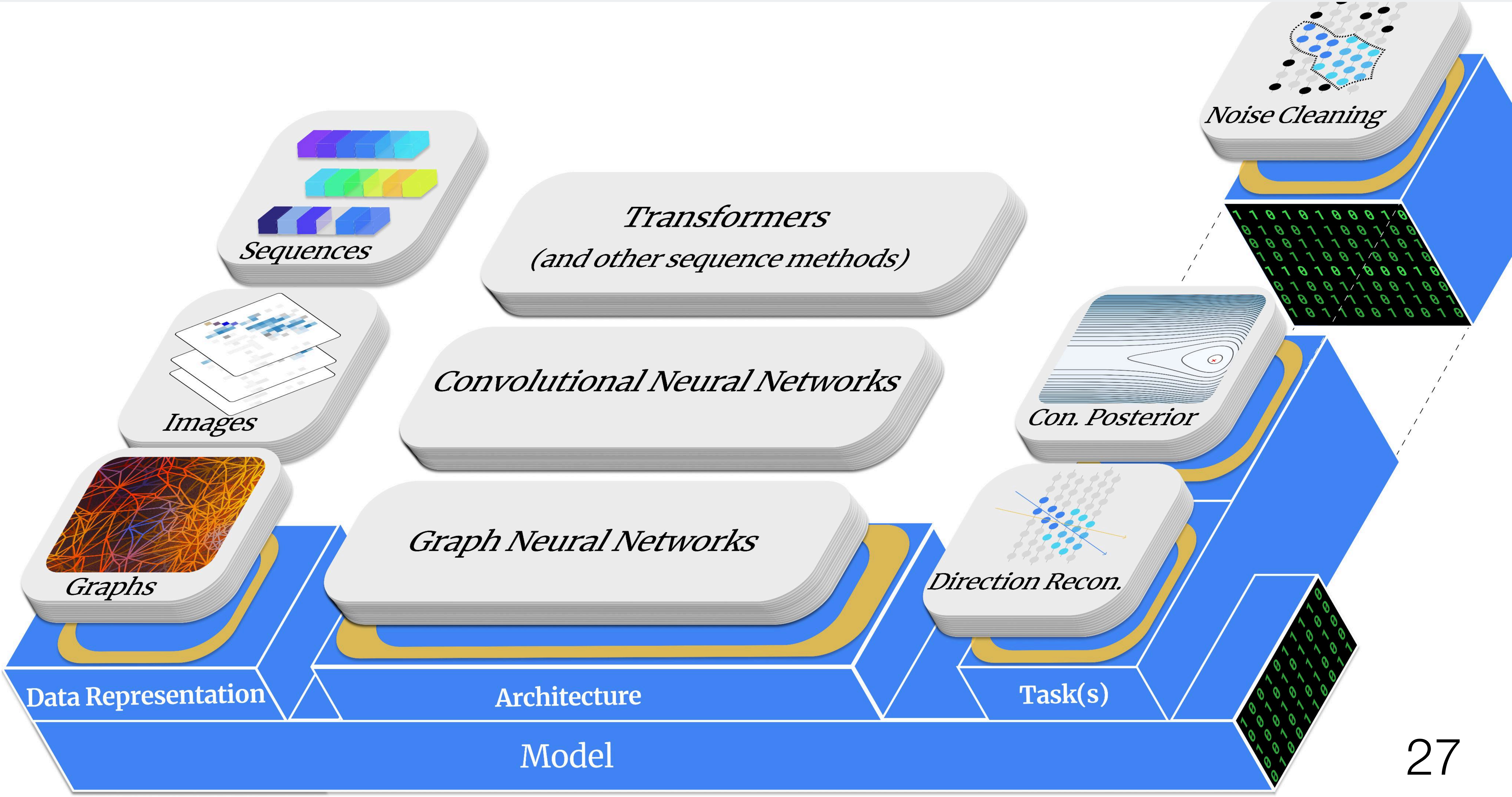
Support for major deep learning paradigms

Solution:
Introduce Data Representation as Model component (Graph, Sequence, Image)

Enable Architectures to be any of the major deep-learning paradigms (CNNs, GNNs, Sequence-models, etc)



Support for major deep learning paradigms

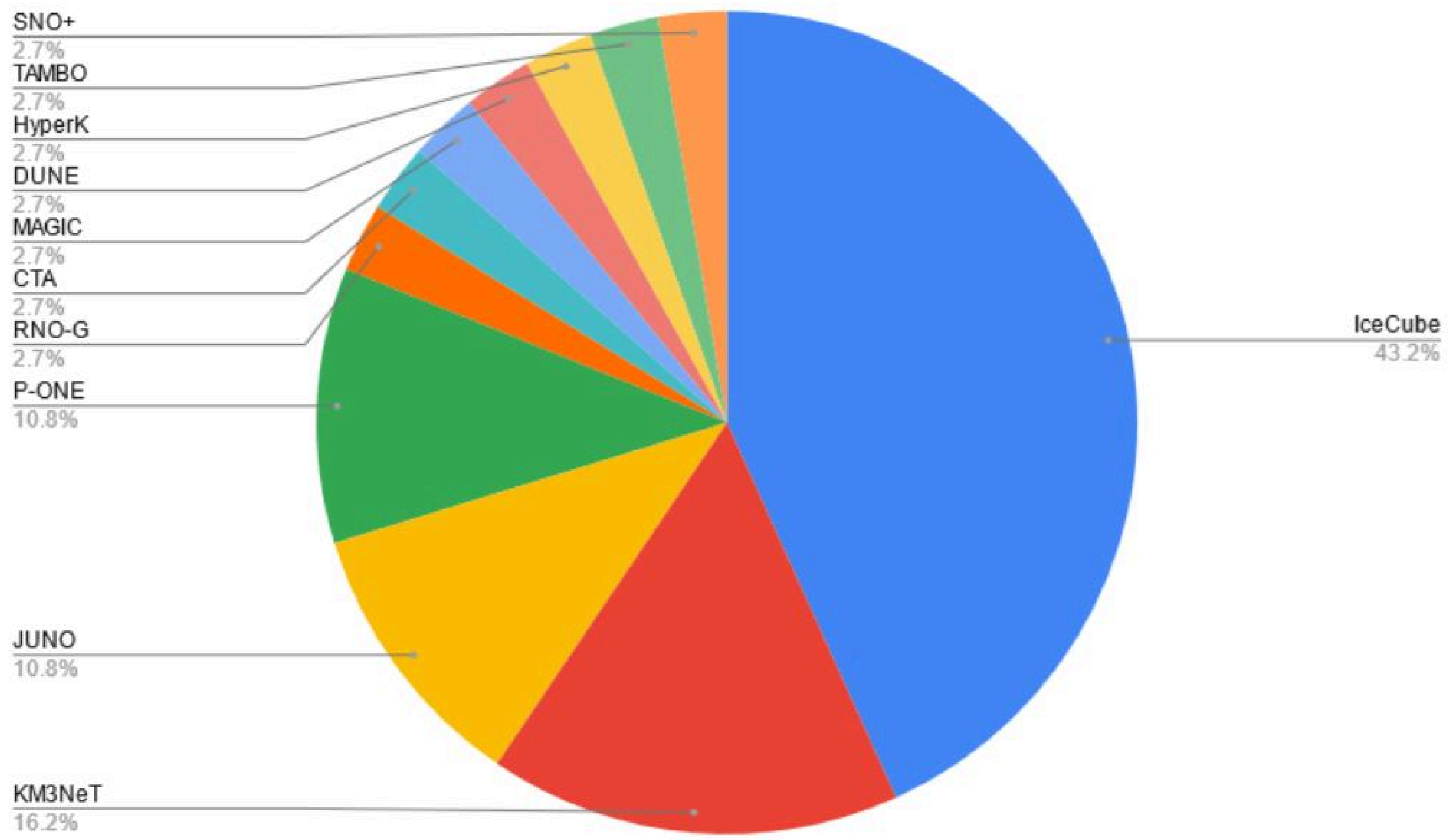


More community building

4th Workshop: “Graph Neural Networks and Beyond”

Focus:

- Rebranding from GNN-library to deep-learning library
- Broader representation of experiments
- Making connections to related fields such as jet-tagging
- Community Project to keep us engaged



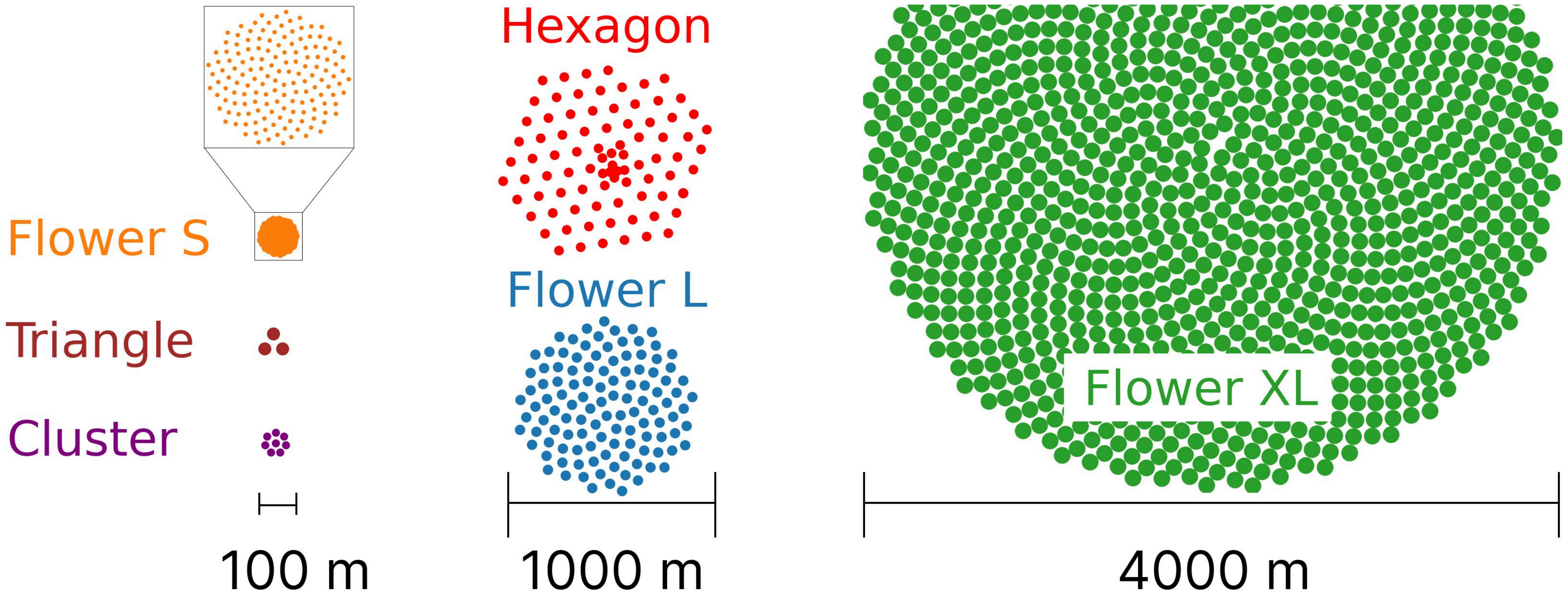
More community building

Community Project:

130 million simulated neutrino events in 6 different detector geometries with the prometheus team.

Goal: Release datasets and processing code for future comparisons/iterations. Publish paper.

Train and compare GNNs and transformer-based methods on 5 common reconstruction tasks.



GraphNeT Today

GraphNeT today in numbers

GitHub

Fork 104

Starred 102

Contributors 29

+ 15 contributors

Pull requests 9 ✓ 444 Closed

Issues 58 Closed 307

GraphNeT today in numbers

GitHub

Fork 104 | Starred 102

Contributors 29

+ 15 contributors

Pull requests 9 ✓ 444 Closed

Issues 58 Closed 307

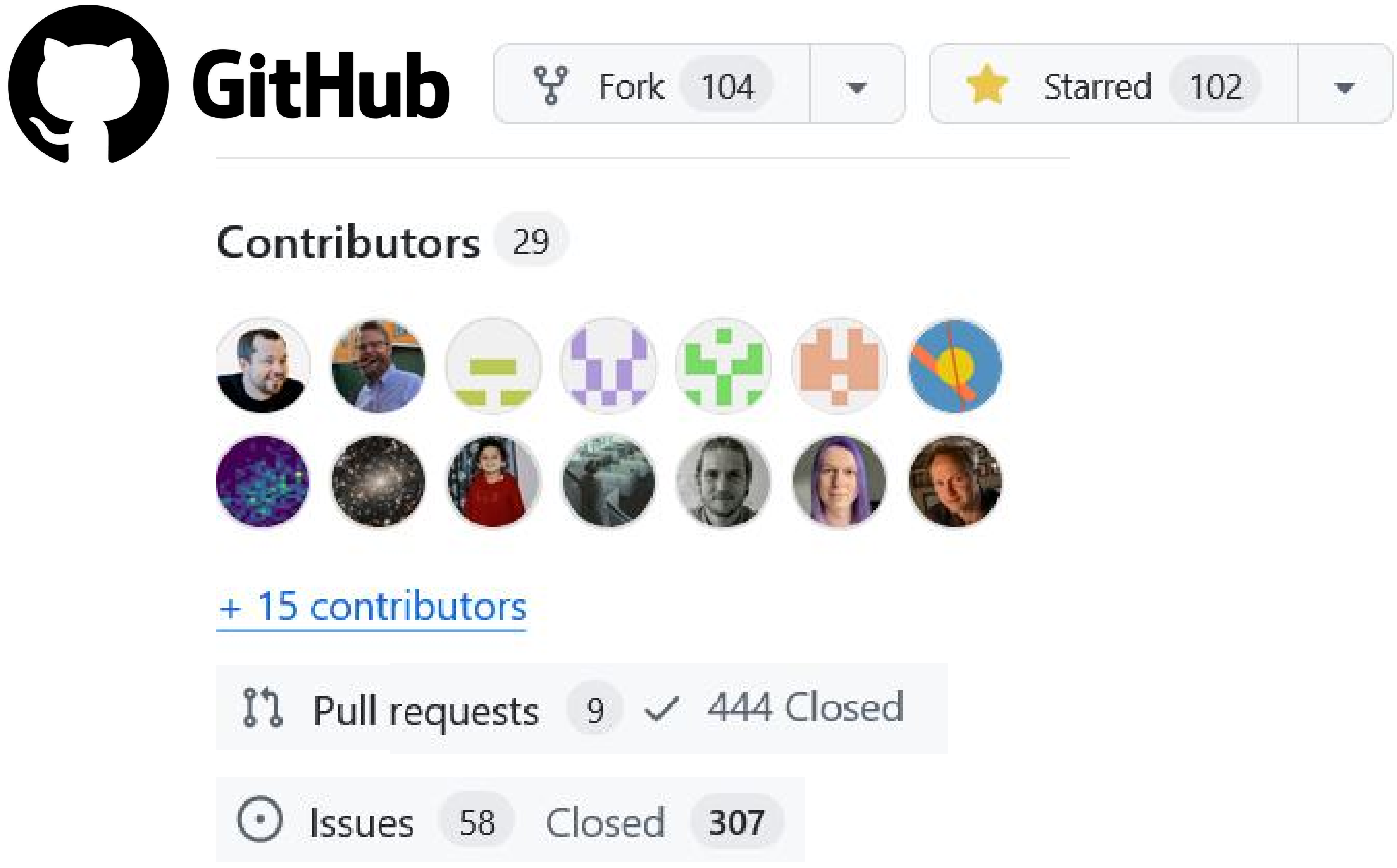
slack — 136 people on slack

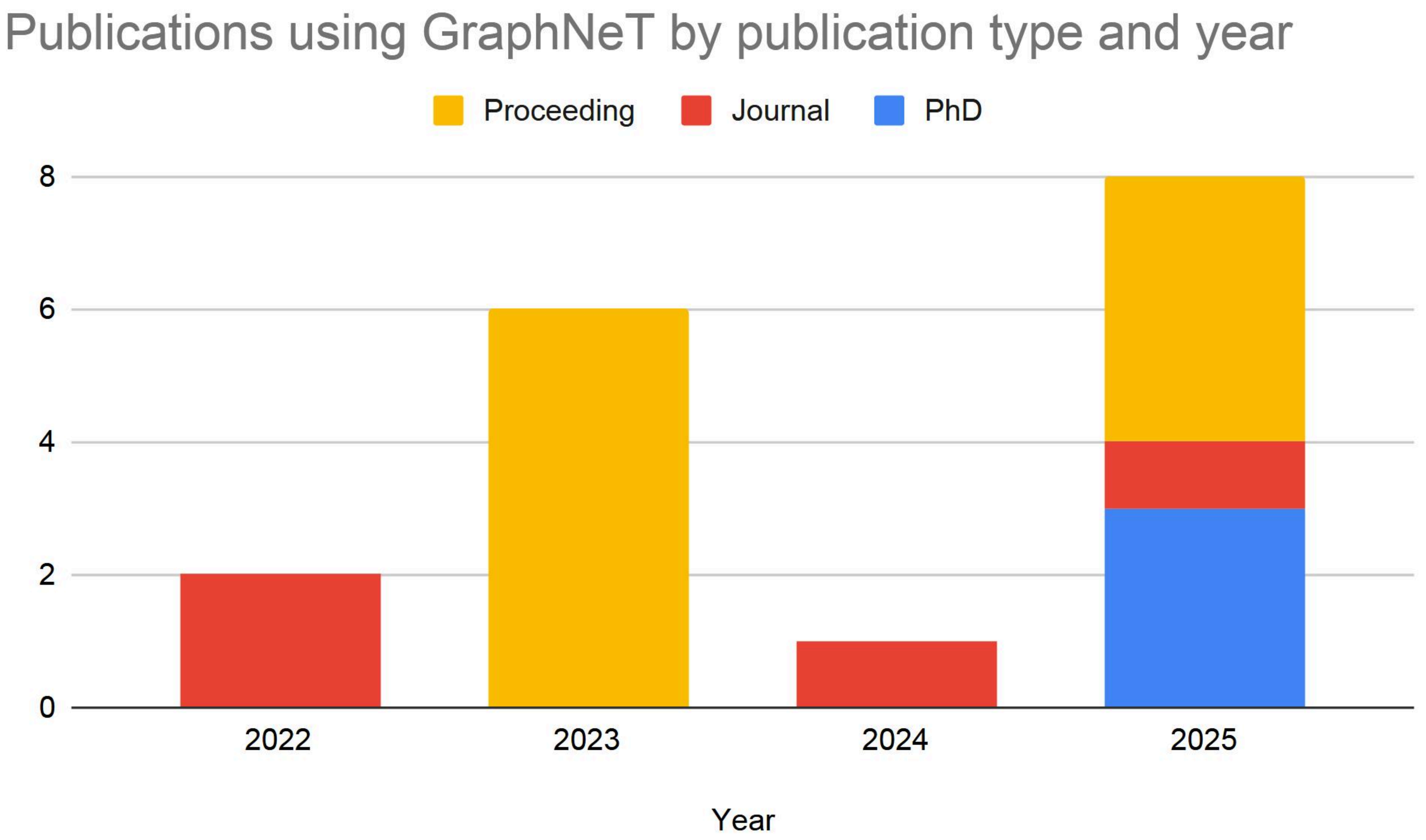
~225.000€ in direct funding

zoom — > 100 regular calls

5 workshops in 2 different countries

GraphNeT today in numbers





Around 17 in total

(I probably missed a few)

136 people on slack

~225.000€ in direct funding

> 100 regular calls

5 workshops in 2 different countries

GraphNeT today in numbers

7 models, three paradigms

Transformers:

TitoModel (1st place Kaggle)
IceMix (2nd place Kaggle)
ISeeCube

GNNs:

ParticleNeT/ORCANet (KM3NeT)
DynEdge (IceCube)
GRIT
ConvNet

Normalizing Flows:

Support for jammy_flows implemented

GraphNeT today in numbers

7 models, three paradigms

Transformers:

TitoModel (1st place Kaggle)
IceMix (2nd place Kaggle)

ISeeCube

GNNs:

ParticleNeT/ORCANet (KM3NeT)
DynEdge (IceCube)
GRIT
ConvNet

Normalizing Flows:

Support for jammy_flows implemented

Two integrated experiments

Three experiments being integrated

“Also applied in”

Experiments integrated today

Recent survey suggests around 40% of ML efforts in IceCube use GraphNeT in one way or another.

GraphNeT plays a central role in low-energy regime currently.

Experiments integrated today

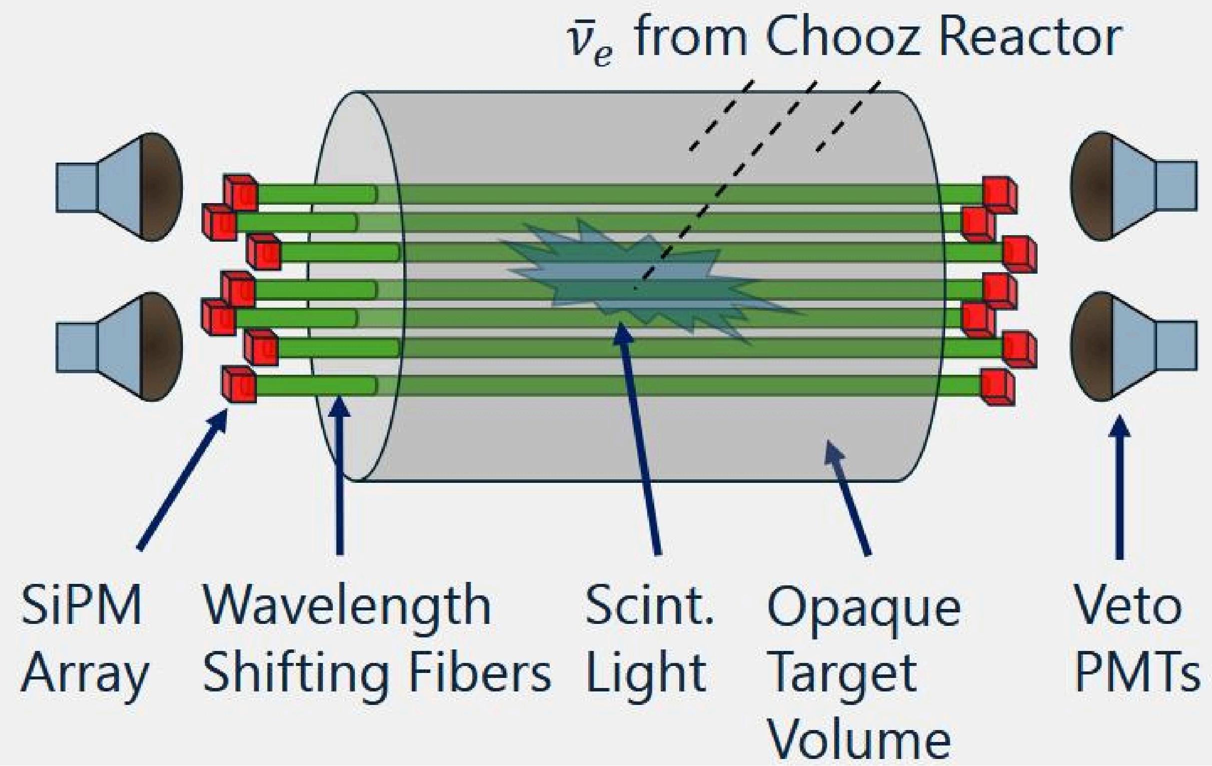
Recent survey suggests around 40% of ML efforts in IceCube use GraphNeT in one way or another.

GraphNeT plays a central role in low-energy regime currently.

Also known as CLOUD. A detector for reactor neutrinos. In prototyping stage. Actively using GraphNeT for various things.

Fig. 1: Diagram of the CLOUD Detector

35 m from the Chooz reactor, ~10000 fibers, 5–10-ton opaque target volume.
See poster by D. Navas



Experiments being integrated

KM3NeT

Integration lead by **Jorge Prado** and **Iván Mateo**

GraphNeT has been applied within KM3NeT quite a bit already.

Status: Integration work is completed, but the PR is pending internal review.

Experiments being integrated

KM3NeT

Integration lead by **Jorge Prado** and **Iván Mateo**

GraphNeT has been applied within KM3NeT quite a bit already.

Status: Integration work is completed, but the PR is pending internal review.

MAGIC

Integration lead by **Jarred Green**

Working local integration and first results looks promising.

Status: Refactoring of local integration pending (Rasmus has promised to help)

Experiments being integrated

P-ONE

Lead by Victoria Parish / Cristina Gualda / Rasmus Ørsøe

Different local integrations exists - GraphNeT has been applied quite a bit already for triggering, geometry and reconstruction studies.

P-ONE uses IceTray, so they are “somewhat” integrated already through IceCube.

Status: Formal integration is pending a finalized simulation chain from P-ONE.

Failed/Botched Integrations

ESS neutrino Super Beam plus (2023)

Lead by Kaare Iversen (Lund University)

Local integration successful.

The formal integration was attempted prior to the generalization of the dataconverter, which meant Kaare got stuck trying to write the conversion code from scratch.

Was only working on ESSnuSB part-time, eventually moved on to new adventures.

Status: Paper published, but integration inactive

Failed/Botched Integrations

ESS neutrino Super Beam plus (2023)

Lead by Kaare Iversen (Lund University)

Local integration successful.

The formal integration was attempted prior to the generalization of the dataconverter, which meant Kaare got stuck trying to write the conversion code from scratch.

Was only working on ESSnuSB part-time, eventually moved on to new adventures.

Status: Paper published, but integration inactive

SNO+ (2024)

Lead by Meng Lou (UPen)

Local integration successful. Applied GraphNeT for background rejection in dark matter searches with promising results months before finishing his PhD.

Formal integration delayed because Meng was finishing his PhD.

Status: Inactive

Models in “production” in IceCube

OscNext

A sample of low-energy neutrinos. Used primarily for studying neutrino oscillations and mass ordering. Analysis containing 11 years of data has unblinded. Uses models from GraphNeT for predictions on key variables.

Variables: Zenith, Energy, Track/Cascade

Leads:
Tom Stuttard

IceCube Upgrade / QUESO

The near-future extension of IceCube. Will significantly improve sensitivity to low-energetic neutrinos. Current simulation chain relies on GraphNeT for noise cleaning and reconstruction.

Variables: Noise cleaning, Zenith, Energy, Track/Cascade, Direction

Leads:
Kayla DeHolton
Jan Weldert
Rasmus Ørsøe

Simplified Canvas

Active Issue	Initialized	Has PR	PR Reviewed	Merged	
Generalize Data conversion	✓	✓	✓	✓	
Introduce DataRepresentation Component	✓	✓	✓	✓	**2.0**
Improve documentation, add installation matrix	✓	✓	✓	✓	
ImageRepresentations & CNNs	✓	✓	−		
KM3NeT Integration	✓	✓	−		
MAGIC Integration	✓	−			**>2.0**
SequenceRepresentations & Transformers at scale	✓	−			

Future of GraphNeT

Future of GraphNeT

We are near the end of the first feedback-cycle

What's next?

Future of GraphNeT

We are near the end of the first feedback-cycle

What's next?

We begin to define this *together* today

Open questions on my mind

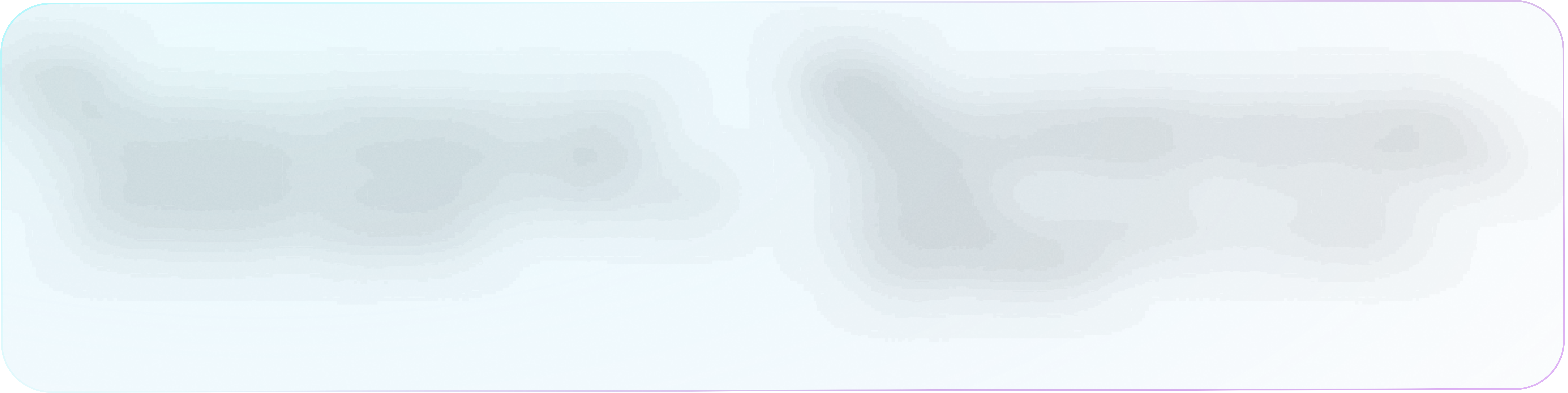
We want fast random access in our file formats.

SQLite offers this, but query speeds are linear in event size, IO intensive and it has no compression. This can be prohibitive.

Are there good alternatives/supplements?

Open questions on my mind

DataRepresentations are computed in real-time. Is that a problem?



Open questions on my mind

Are we missing important features?

Open questions on my mind

Are there flaws in the premise of a common library?

Open questions on my mind

As the library spreads to more experiments, and is used in more analyses, which unique problems might arise and how may we mitigate them now?

Open questions on my mind

As deep-learning adoption in analyses are increasing, what requirements should be set by collaborations on models and their deployment, and how may the library be helpful here?

Open questions on my mind

How do we best ensure the longevity of the library?

Thanks!