

Neutrino Open Data Challenges: Community Report from NPML 2025

Roger Huang

9/29/2026

Where did this white paper come from?

NPML 2025 in Tokyo held an **unconference-style** satellite workshop to discuss open data challenges for neutrino physics in general, covering:

- What do we hope to achieve by running open data challenges?
- What should an open data challenge look like to achieve those goals?
- What ideas do people have for open data challenges in the field right now?
- What do we need to do to work towards setting up our first open data challenges?

This white paper is to summarize and publicize the outcomes of those discussions

What is the goal of this white paper?

Many people will agree that “open data is good”, but don’t have concrete ideas of what they should look like or how they should be used

Aim to be a statement endorsed by people working in ML and neutrino physics:

- Establish that we believe open data challenges are beneficial and worth pursuing for problems within neutrino physics
- Discuss the criteria we developed to determine what types of problems were most well-suited to the open data challenge format
- Describe various challenge proposals, brainstormed during the workshop or contributed later, and subject them to these criteria

Everyone is welcome to sign on

Authorship

Publicized form to NPML, DUNE ML, T2K ML, and a number of other collaboration mailing lists to garner signatures and comments

Currently around ~50 authors

Neutrino Physics and Machine Learning: Open Data Challenge Community Report

Manoj Adhikari¹, Saul Alonso-Monsalve², Omar Alterkait^{3,4}, Andrew Atta⁵, Muhammad Bilal Azam⁶, Marta Babicz⁷, Rishi Babu⁸, David Caratelli⁹, Bear Carlson¹⁰, Seokju Chung¹¹, Xuefeng Ding¹², Zelimir Djurcic¹³, Max Asa Albert Dornfest¹⁴, Francois Drielsma¹⁵, M. A. DuVernois¹⁶, Castaly Fan¹⁰, Terrance Figy¹⁷, Peter Gaemers¹⁵, Gabor Galgoczi¹⁸, K. M. Graczyk¹⁹, Pierre Granger²⁰, Gajendra Gurung^{21,20}, Nathan Higginbotham²², Aaron Higuera²³, Roger G. Huang²⁴, César Jesús-Valls²⁰, Daeun Jung²⁵, Christina Karagianni^{26,27}, Lukas Koch²⁸, Jiaoyang Li (J.-Y. Li)²⁹, Jiaxi Liu³⁰, Nived Puthumana Meleppattu³¹, Jun Nakane³², Subrahmanya Saicharan Pemmaraju³³, Nick W. Prouse³⁴, Geting Qin³⁵, Benjamin Quilain³⁶, Masooma Sarfraz³⁷, Hasung Song³⁸, Mohammed Sultan³⁹, Kazuhiro Terao¹⁵, Matt Troups²⁹, Ka Vang Tsang¹⁵, Hilary Utaegbulam⁴⁰, Ziao Wang⁴¹, Callum Wilkinson²⁴, Dowling Wong⁴², Junjie Xia¹⁵, Yikai Xu⁴³, Yongheng Xu⁴⁴, Samuel Young⁴⁵, and Larry Zhao³⁰

Outline

Contents

1	Introduction	2
2	NPML Workshop Format	4
3	Expectations for Open Data Challenges	6
3.1	Definition of Open Data Challenges	6
3.2	Required Resources	7
4	Criteria for Judging Open Data Challenge Proposals	7
5	Near-Term Proposals	8
6	Mid-Term Proposals	10
7	Long-Term Proposals	13
8	Conclusions	15

Plan to add an executive summary at the beginning

- Will include table on what we mean by near-term, mid-term, long-term proposals

Unconference Description

2 NPML Workshop Format

Includes description of unconference format and example of a quad chart

Lists all of the questions discussed in the various quad chart sessions we held

Scientific Challenges

- Providing sensible metrics/labels so the models are trying to learn interesting/useful/relevant information
 - Made more difficult by application of same models to multiple detectors?
- How to break out of the neutrino bubble?
- Getting people to compare against the same data (avoid space race mentality), even within the same experiment
- Interpretability and robustness to baked in physics assumptions – uncertainty quantification

What Datasets? How to obtain/maintain?

- Generic open-source datasets – how to ensure these are sophisticated enough?
 - How to ensure similar across detector types?
 - Experiments may also want to have “internal” datasets for internal benchmarking
- Finding shared resources everyone can access is challenging within a collaboration, also tough outside

Scopes (science/detectors/AI methods)

- Finding complementarity between research subfields
 - But maybe also good to not force it, and recognize that sometimes factorization is sensible
- Is it possible to find a level of abstraction for comparing multiple detector types? Or is that just FMs?

Increase Participation

- Providing data in a friendly format: flat ROOT trees might be okay, but custom ROOT classes a no-no
- Documentation of standards – consistent choices between open datasets
- If a challenge is so large that you need to download and process millions of large events, it might be difficult both from a local storage and processing point of view

Open Data Challenge Expectations

3 Expectations for Open Data Challenges

7

3.1 Definition of Open Data Challenges . . .

3.2 Required Resources

Section to cover what we require out of a open data challenge in order to realize the benefits

Suggestion received to add a 3rd requirement: documentation and examples of how to use the open data

3.1 Definition of Open Data Challenges

We define an open data challenge as any setup that contains both:

1. A set of open data, freely available to anyone who wishes to use it.
2. Standardized performance metrics on that data, which define a problem of interest and provide a means to compare different solutions. Such metrics should generally be accompanied by benchmark values to indicate what performance levels are currently achievable or are otherwise considered “good”.

Data Challenge Classification

4 Criteria for Judging Open Data Challenge Proposals

We judge every open data challenge proposal by asking:

1. Is the problem of broad interest to neutrino physicists?
2. How hard is it to generate the required dataset, and could it be reused for other challenges?
3. Are there clearly defined metrics for evaluating solutions?
4. What computational resources would be needed?

We then sort every proposal into **near-term**, **mid-term**, and **long-term** proposals based on these criteria, as a suggestion for how to prioritize effort

Near-Term Proposal Example

These proposals are feasible to do in the near future, as some of the first data challenges. A combination of wide interest and not too much development needed

Particle-Level Instance Segmentation This focuses on the low-level task of segmentation of detector-level hits, the results of which are used as input for many types of reconstruction algorithms. Improving performance on this task would improve the downstream performance of many other aspects of reconstruction as well.

- **Dataset:** Labeled simulated datasets suited for this challenge already exist, having been created to train the current generation of reconstruction algorithms. These datasets are generally specific to a particular detector technology, but do not have to be specific to a particular experiment.
- **Metrics:** Purity and Adjusted Rand Index (ARI) metrics are well-established metrics for evaluating current solutions to this task.
- **Other Notes:** This challenge seems particularly well-suited for attracting outside experts, given the problem requires little knowledge of the associated physics and has straightforward metrics.

Mid-term Proposal Example

Challenges with clear interest, and a basic idea of what needs to be done, but which will require some effort to work out

Supernova Data Compression Instead of performing supernova pointing directly from the raw detector data, this challenge would ask participants to compress the raw data while retaining enough information to perform effective pointing. This is of particular interest for underground experiments with heavy computational and bandwidth constraints at their detector site. If supernova data could be efficiently compressed, they could be quickly sent out to remote clusters, where the initial pointing could be done by more computationally intensive advanced algorithms.

- **Dataset:** Raw detector data from a supernova burst would be part of this dataset, like in the previous challenge. However, the storage demands would probably be even higher, since this challenge would not pre-compress any of the raw detector data.
- **Metrics:** Evaluation would be some combination of data compression efficiency and the supernova pointing resolution that is possible with the resulting compressed data. However, it is not obvious how to balance these two in the evaluation.
- **Other Notes:** Since we need to know the pointing resolution that is possible with the compressed data, this challenge would probably also require participants to submit a pointing algorithm intended for use with their compression approach. This pointing algorithm would not be subject to particular hardware or runtime constraints, but the organizers must have a way to verify that its output comes from the compressed data, and not directly from the raw test dataset.

Long-term Proposal Example

Proposals where we don't have a good idea of how to actually turn them into open DCs

Some of these just don't seem like good options for open DCs, but we tried to be inclusive

Detector Design Optimization When developing new techniques, the detector design is often treated as immutable, simply because it has already been built or is already under construction. When looking toward future experiments, there is the potential to use ML-based techniques to explore the space of possible detector designs and better optimize the detector's sensitivity to particular processes.

- **Dataset:** The primary content of the dataset would be a set of truth-level interactions in some near-infinite medium of interest. Solutions would propose a detector design to construct around this volume. The dataset may need to include sets of allowed detector components and sensors, along with defined responses for them, so that designs cannot simply propose unrealistically high-performing components.
- **Metrics:** Many possible metrics exist, but it is unclear how to choose and combine them in an appropriate way. Detector performance metrics, such as energy resolution, spatial resolution, and particle identification capabilities, would be relevant, but many of these depend on downstream analysis tools as well. Metrics could also be aimed at specific physics processes instead of low-level detector metrics, but this has similar difficulties.
- **Other Notes:** The amount of information needed to constrain detector designs to be realistic and the number of metrics that need to be optimized make this difficult to define as an open data challenge.

Contributing More Data Challenge Proposals

We welcome authors to write additional data challenge ideas to put in this white paper, but it's not meant to be exhaustive

- New ones will have to be sorted into the near/mid/long-term categories too

Also welcome people to expand on the descriptions of the data challenges we already included

- Aimed for them to not have overly detailed explanations of the physics cases for everything
- Currently also not including tons of references to existing work and state-of-the-art, for fear of getting into questions about what should be cited
- But these decisions can be discussed

Timeline

Initial draft prepared by very limited group, using notes from the satellite workshop

First draft sent to NPML 2025 satellite list 9/7, signatures and comments by 9/14

Second draft sent to full NPML 2025/2026 lists + various collaboration mailing lists on 9/15, with deadline for more signatures and comments on 9/25

Now: have some new feedback of substance. Will plan to expand writing group to include people who volunteered to help or had substantial comments.

Plan for a new draft by end of October, release paper in November

- Definitely aim for release before first DORAEMON challenges

If you haven't seen the white paper and are interested, let me know!