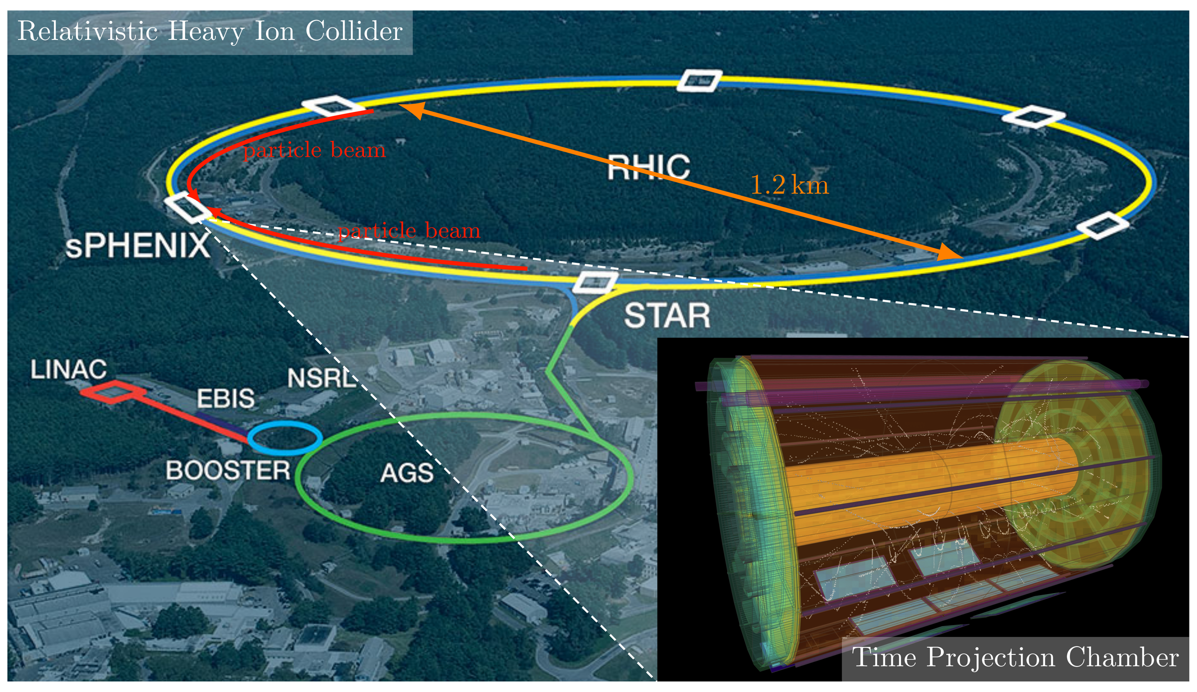
FM4NPP: A Scaling Foundation Model for Nuclear and Particle Physics

code & data

David Park[1], Shuhang Li[2], Yi Huang[1], Xihaier Luo[1], Haiwang Yu[2], Yeonju Go[2], Christopher Pinkenburg[2], Yuewei Lin[1], Shinjae Yoo[1], Joseph Osborn[2], Jin Huang[2], Yihui "Ray" Ren[1]

[1]AI Department and [2]Nuclear and Particle Physics Department, Brookhaven National Laboratory, Upton, NY



Abstract

Large language models have revolutionized artificial intelligence by enabling large, generalizable models trained through self-supervision. This paradigm has inspired the development of scientific foundation models (FMs). However, applying this capability to experimental particle physics is challenging due to the sparse, spatially distributed nature of detector data, which differs dramatically from natural language. This work addresses if an FM for particle physics can scale and generalize across diverse tasks. We introduce a new dataset with more than 10 million particle collision events and a suite of downstream tasks and labeled data for evaluation. We propose a novel self-supervised training method for detector data and demonstrate its neural scalability with models that feature up to 188 million parameters. With frozen weights and task-specific adapters, this FM consistently outperforms baseline models across all downstream tasks. The performance also exhibits robust data-efficient adaptation. Further analysis reveals the representations extracted by the FM are task-agnostic but can be specialized via a single linear mapping for different downstream tasks.

Key Questions

- Can a scaling FM be trained for sparse detector data using self-supervised learning without labeled data?
- Will the FM pre-training scale?
- Will the learned representation from the FM useful for downstream tasks?
- Will a larger FM also lead to better downstream task performance?
- Will adapting from an FM be more data efficient?

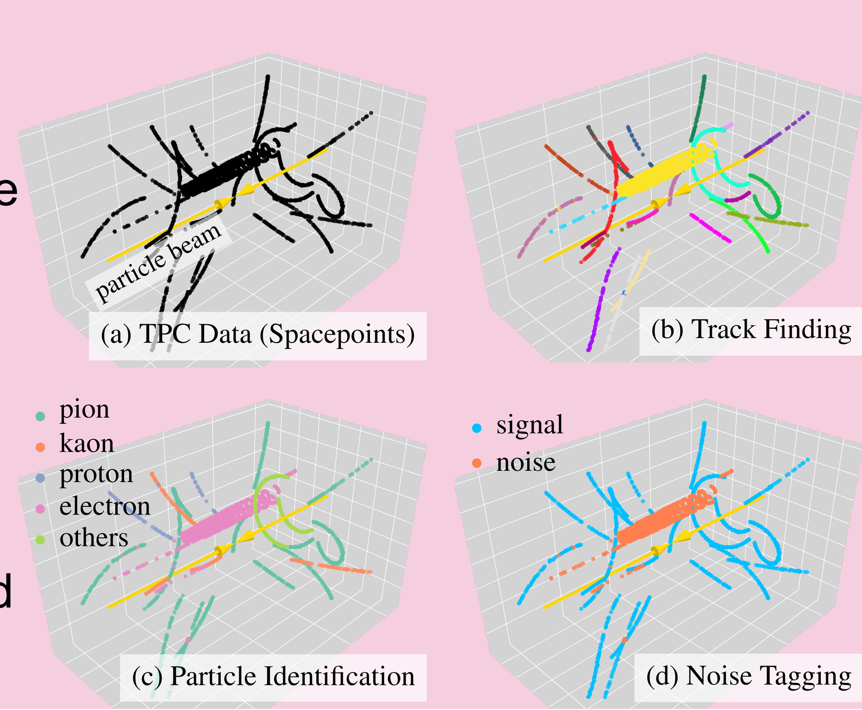
Dataset and Downstream Tasks

sPHENIX Time Projection Chamber (TPC) spacepoints:

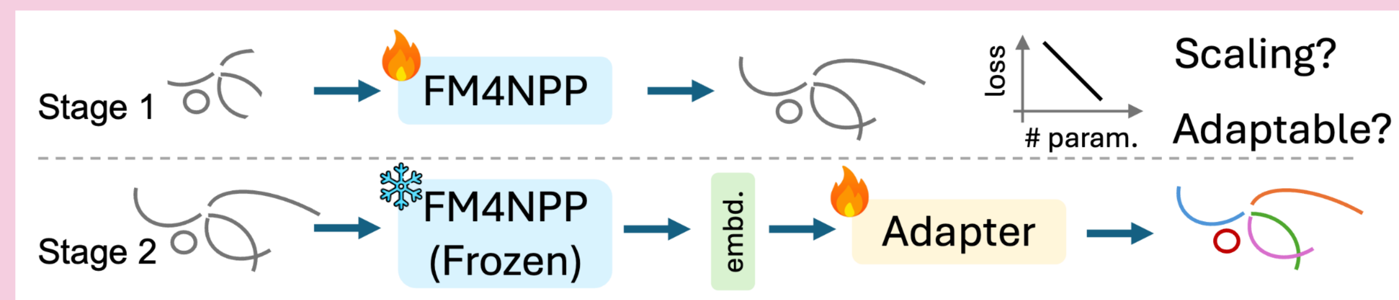
- 10 million events for self-supervised pre-training
- 75,000 events for downstream tasks supervised training.
- p+p collisions at $\sqrt{s}$ = 200 GeV
- Realistic simulation includes continuous energy loss, multiple scattering, secondary particle production, and decay processes when interacting with the detector material.

Three Downstream Tasks:

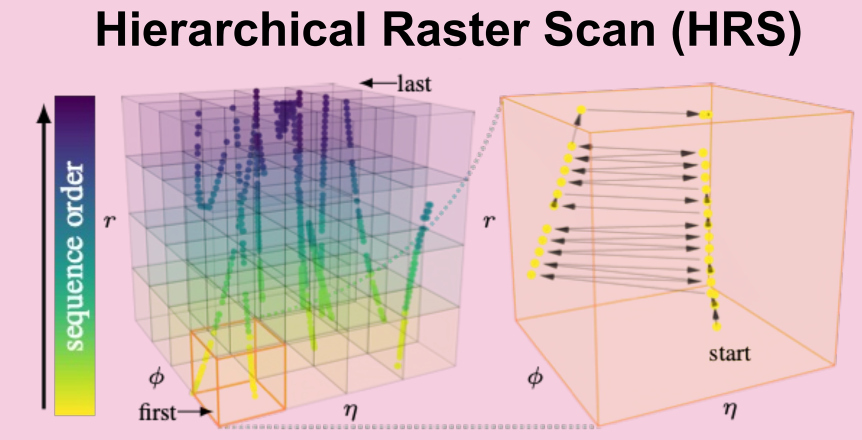
- **Track Finding**: group spacepoints from the same particle together
- **Noise Tagging**: classify spacepoints from low momentum secondary particles ("noise")
- **Particle Identification**: classify spacepoints based on their particle species



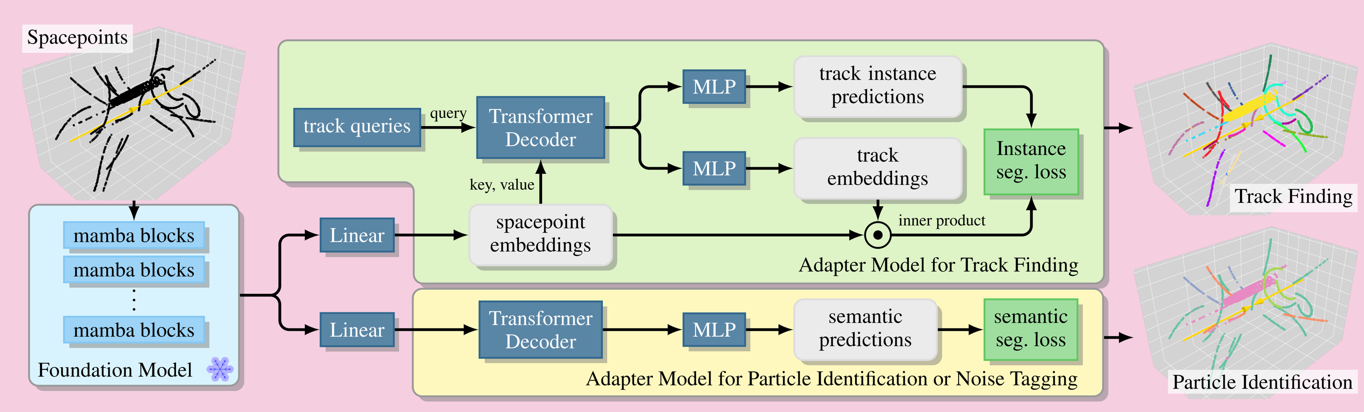
Method: A Scalable and Generalizable FM



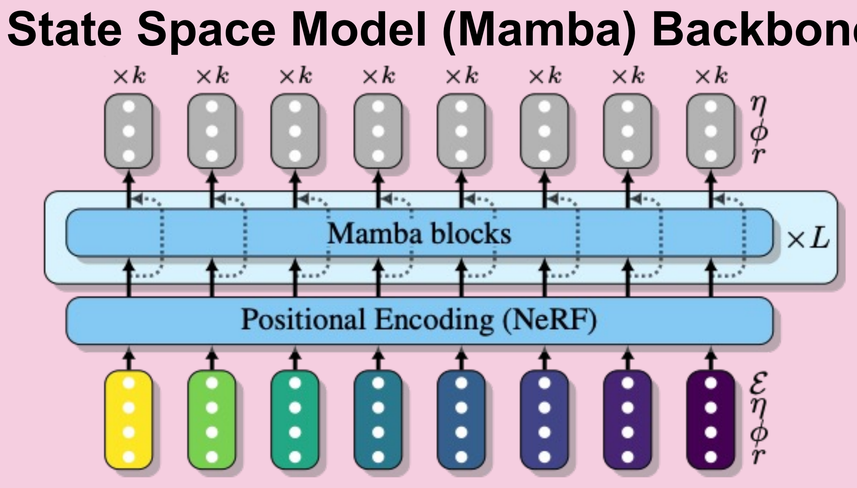
Hierarchical Raster Scan (HRS)



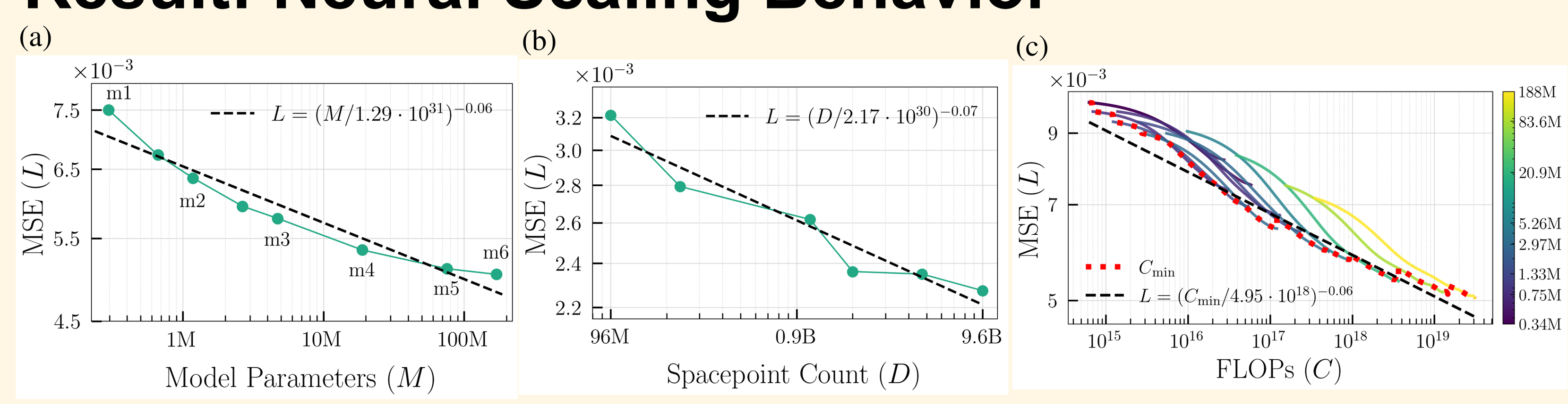
Lightweight Adaptor Transformer Models



State Space Model (Mamba) Backbone



Result: Neural Scaling Behavior



	Model Sizes					
	m1	m2	m3	m4	m5	m6
Model Width	64	128	256	512	1024	1536
Model Params	0.34M	1.3M	5.3M	21M	84M	188M
	Compute Resources					
NVIDIA GPU	H100 80GB		A100 80GB			
Num GPUs	1	1	4	8	24	64
Train Hrs	10	12	20	32	50	72

Clear power-law scaling with model size, data size, and compute.

Result: Cutting-edge Performance on Downstream Tasks

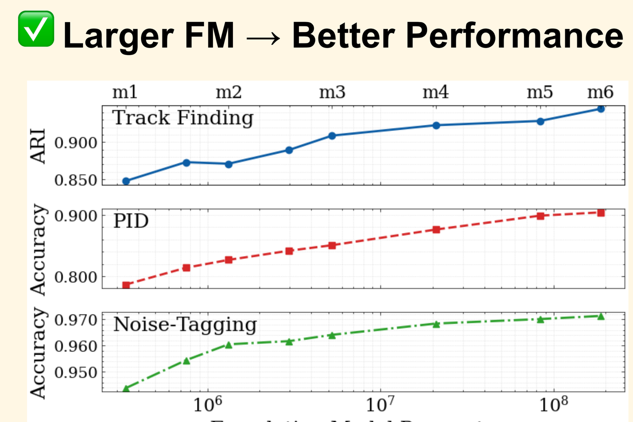
		Track Finding		
model	#trnbl para.	ARI↑	efficiency↑	purity↑
EggNet	0.16M	0.7256	74.19%	<u>75.14%</u>
Exa.TrkX	3.86M	<u>0.8765</u>	<u>91.79%</u>	66.42%
AdapterOnly	2.39M	0.7243	78.01%	64.54%
FM4NPP (m6)	2.39M	**0.9448**	**96.08%**	**93.08%**

		Particle Identification			Noise Tagging		
model	#trnbl para.	acc.↑	recall↑	pre.↑	acc.↑	recall↑	pre.↑
SAGEConv	0.91M	0.7262	0.4563	<u>0.6502</u>	0.9174	0.7227	0.8165
OneFormer3D	44.95M	<u>0.7701</u>	<u>0.4897</u>	0.5767	<u>0.9646</u>	**0.9404**	<u>0.8948</u>
AdapterOnly	0.74M	0.6631	0.3387	0.6111	0.9111	0.6215	0.8359
FM4NPP (m6)	0.74M	**0.9039**	**0.7652**	**0.8782**	**0.9713**	<u>0.9367</u>	**0.9190**

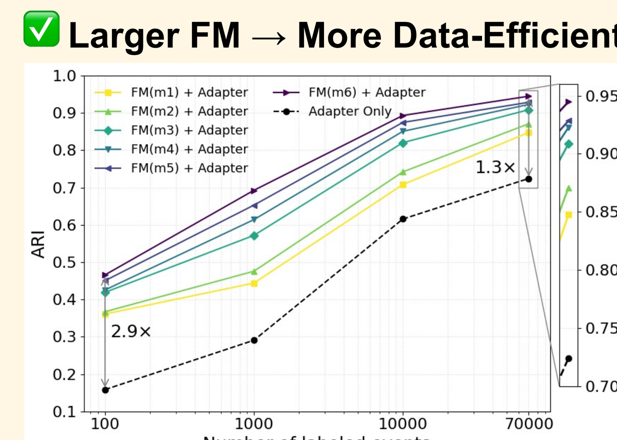
By comparing with the "AdapterOnly" model, we verify the performance gain is from FM4NPP base model.

Park, D., Li, S., Huang, Y., Luo, X., Yu, H., Go, Y., ... & Ren, Y. FM4NPP: A Scaling Foundation Model for Nuclear and Particle Physics. **ICLR 2026** arXiv:2508.14087.

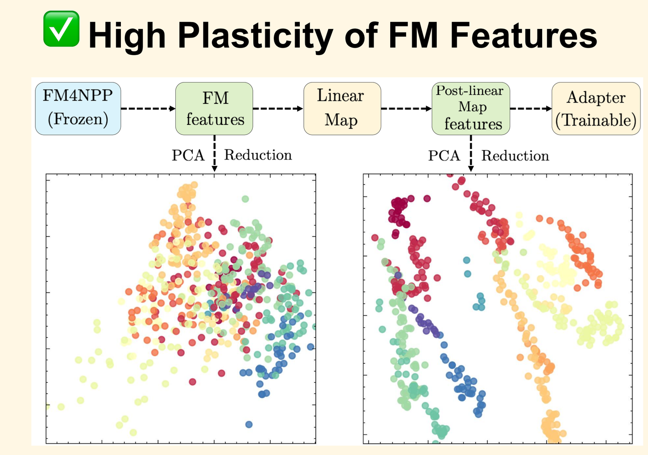
✅ Larger FM → Better Performance



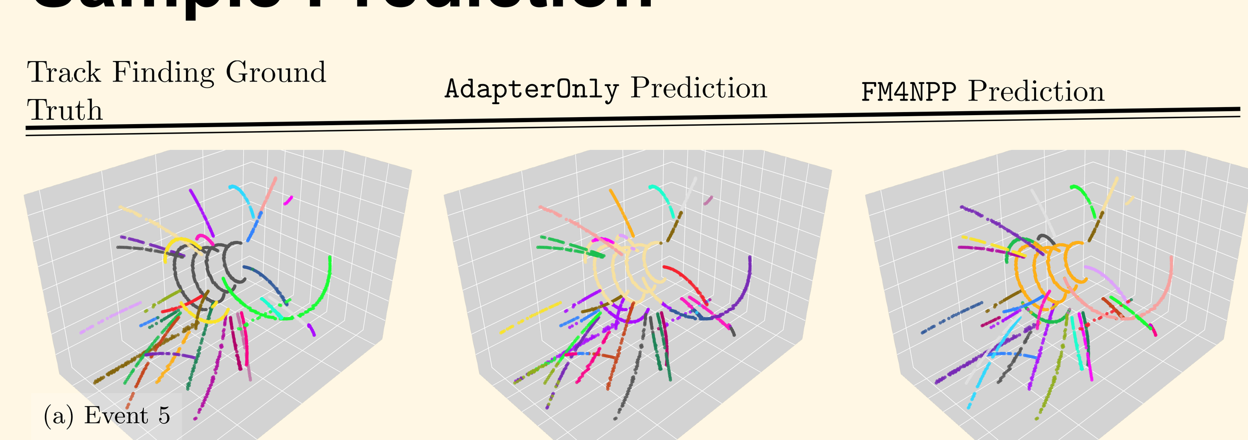
✅ Larger FM → More Data-Efficient

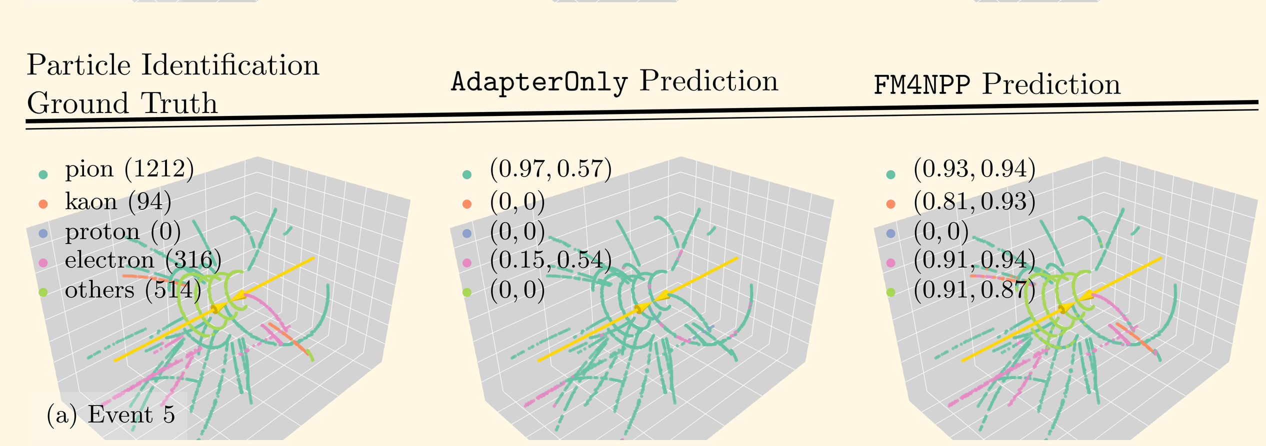


✅ High Plasticity of FM Features



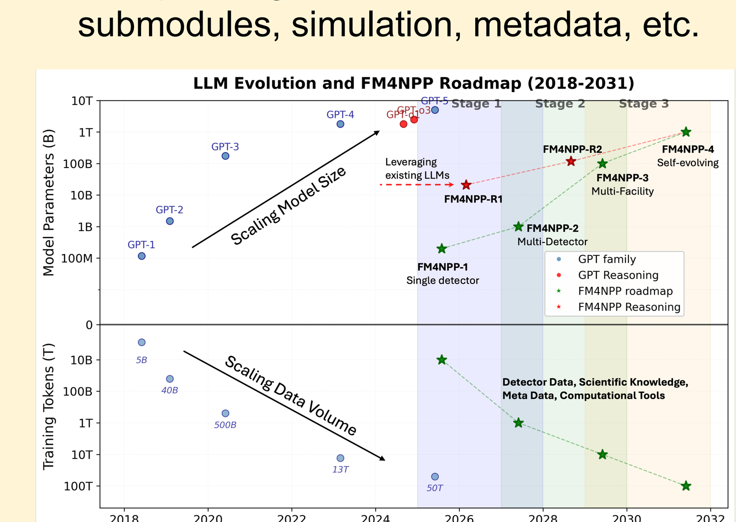
Sample Prediction





Future Work

1. Scaling to bigger models, incorporating more data, and validating on other tasks.
2. Exploring other architectures and self-supervised learning tasks.
3. Expanding to other experiments in NP and HEP (ATLAS, CMS, Belle II, etc.)
4. Incorporating other modalities: detector submodules, simulation, metadata, etc.



Acknowledgments

- This work was supported by the Laboratory Directed Research and Development (LDRD) Program at Brookhaven National Laboratory, LDRD 25-045, which is operated and managed for the U.S. Department of Energy (DOE) Office of Science (SC) by Brookhaven Science Associates under contract No. DE-SC0012704.
- Yihui Ren, Xihaier Luo, and Shinjae Yoo were partially supported by DOE-SC through the Office of Advanced Scientific Computing Research and the Scientific Discovery through Advanced Computing (SciDAC) program.
- Shuhang Li was partially supported by DOE-SC through the Office of Nuclear Physics under Award No.~DE-FG02-86ER40281.
- This research also used resources of the National Energy Research Scientific Computing Center (NERSC) under the "GenAI@NERSC" program. NERSC is a DOE-SC User Facility with Award No. DDR-ERCAP0034059. The authors are grateful to the NERSC staff for their support, particularly Shashank Subramanian and Wahid Bhimji.

U.S. DEPARTMENT of ENERGY

www.bnl.gov

Brookhaven National Laboratory