Foundation Models for Nuclear and Particle Physics (FM4NPP)

Yihui (Ray) Ren yren@bnl.gov

Sr. Computational Scientist (AI/ML) & AI-CoDesign Group Lead

AI Department, Brookhaven National Laboratory

@BrookhavenLab

CLARIPHY AI Collaboration Meeting | Sept. 14, 2026

Foundation Models (FMs) are envisioned as a generalized counterpart to text-based LLM.

- Built on large-scale, primarily **unlabeled** data.
- Trained via **self-supervised learning** on surrogate tasks.
- Pre-trained once and **adaptable to diverse downstream applications**
- Exhibit strong neural scaling behavior in pre-training.
- Achieve state-of-the-art performance across application tasks.

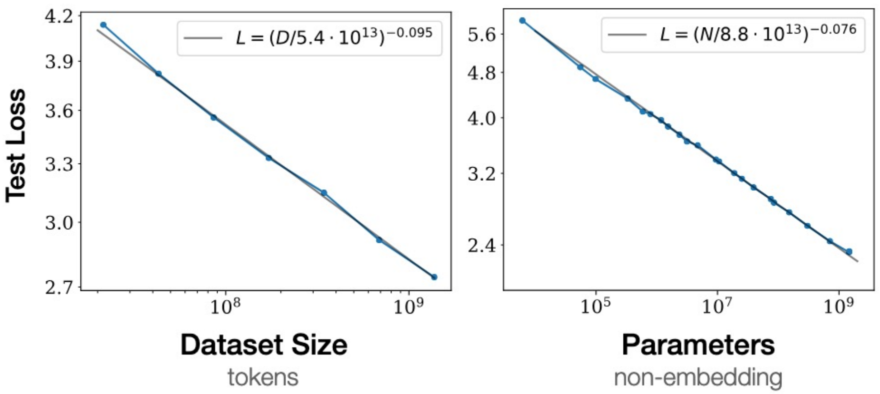


Fig. 1 Neural Scaling Behavior [1]

High-energy Nuclear and Particle Physics (NPP) experiments have a lot of data and exhibiting complex patterns. However, detector data are sparse spatial points, unlike text nor dense images.

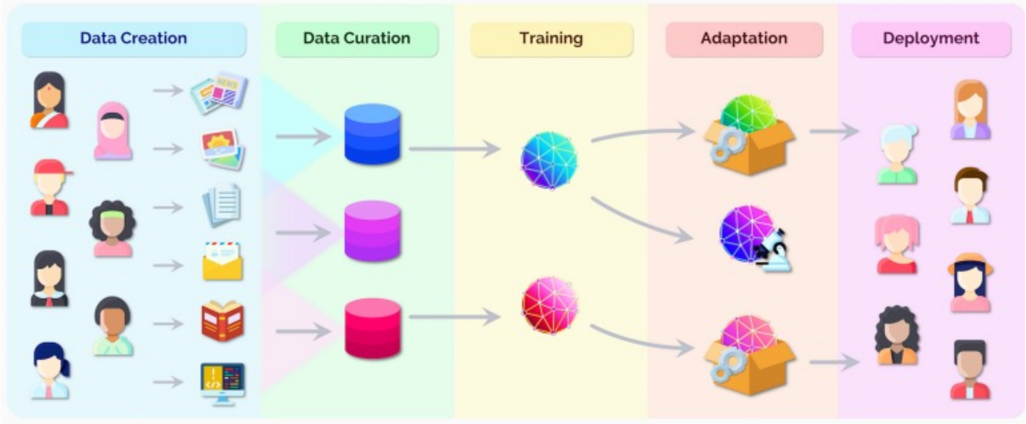


Fig. 2 Foundation Model Paradigm [2]

Key Questions:

1. What is the right self-supervised learning task?
2. Will the FM pre-training scale?
3. Will the representation learned from FM useful?
4. Will larger FM also lead to better downstream tasks?
5. Will adapting from FM be more data efficient?

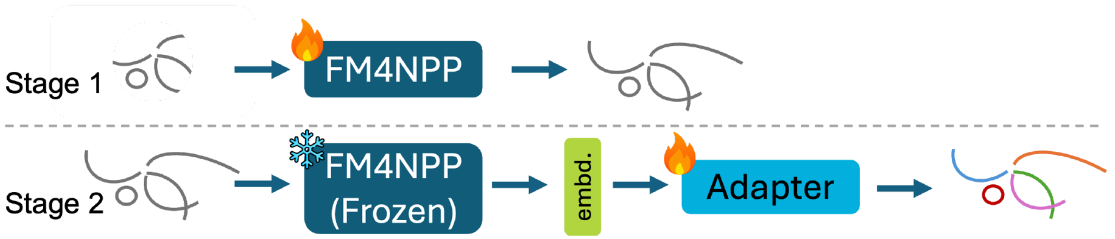
[1] Kaplan, Jared, et al. “Scaling laws for neural language models." *arXiv:2001.08361* (2020).
[2] Bommasani, R. et al. On the Opportunities and Risks of Foundation Models. *arXiv*:2108.07258 (2022).

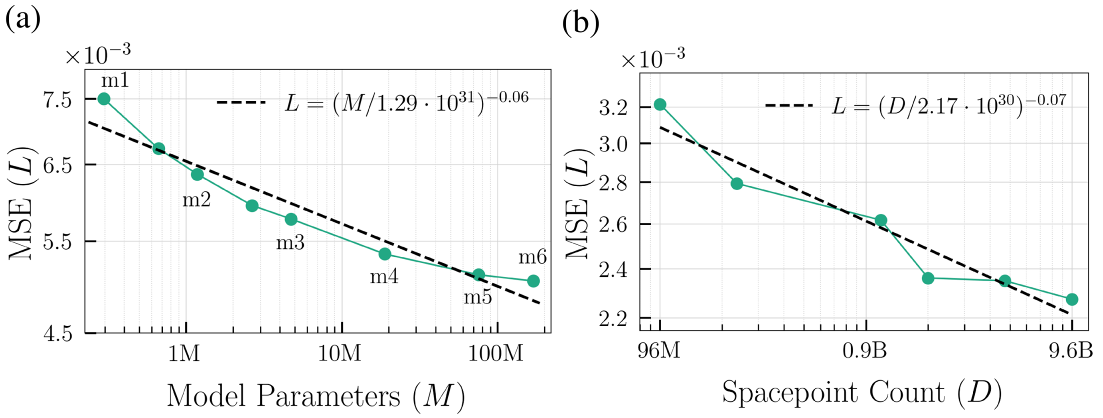
On simulated sPHENIX TPC data, we employed two-stage training strategy:

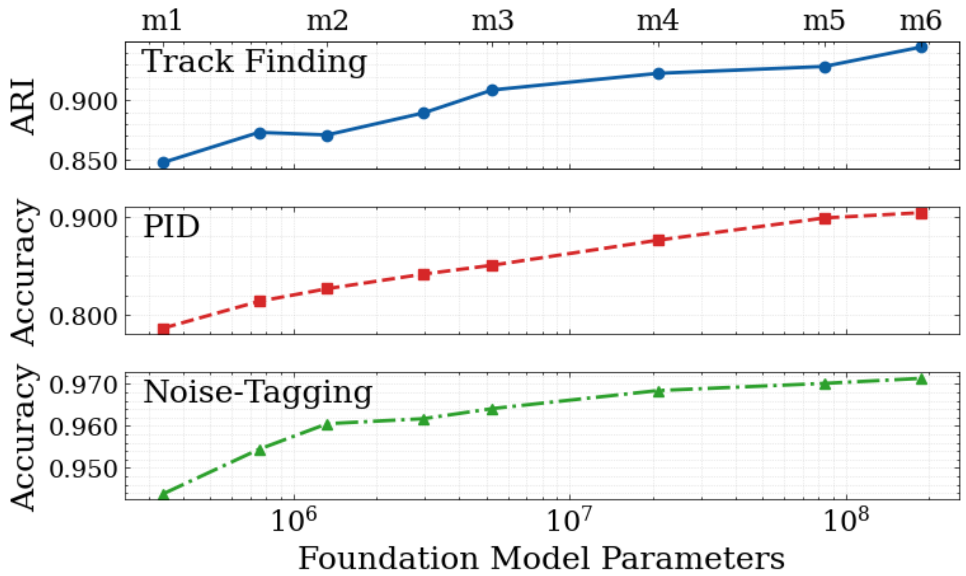
- Stage 1: Self-supervised next-k-nearest-neighbor pre-training using Mamba backbone.
- Stage 2: Frozen backbone with lightweight adapters for different downstream tasks.

Key findings:

- Neural scaling behavior up-to 200M parameters and 10B tokens.
- All three downstream tasks reach the state-of-the-art comparing to baseline models. The scaling behaviors also exhibit in the downstream tasks.





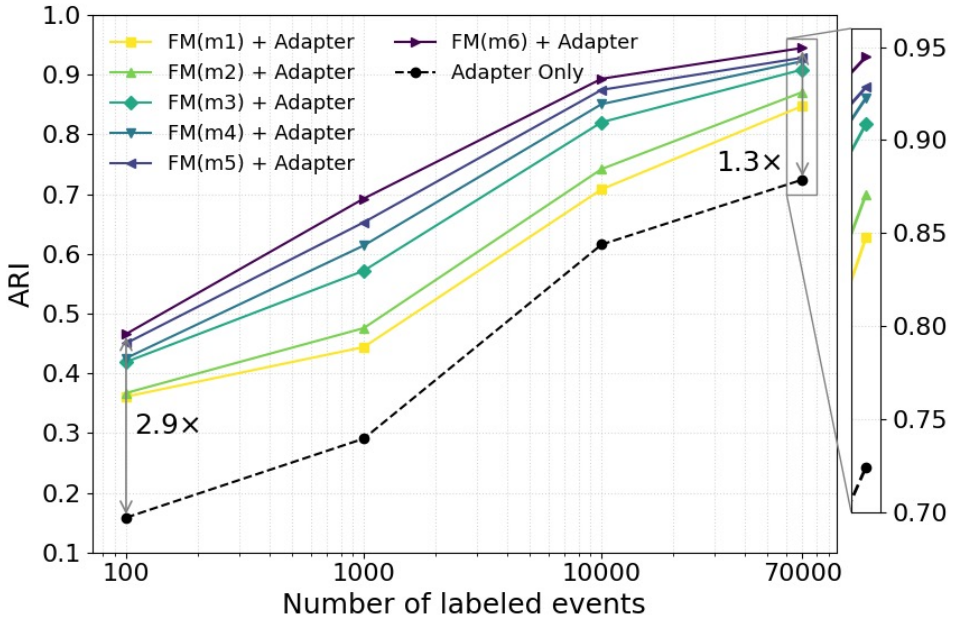


On simulated sPHENIX TPC data, we employed two-stage training strategy:

- Stage 1: Self-supervised next-k-nearest-neighbor pre-training using Mamba backbone.
- Stage 2: Frozen backbone with lightweight adapters for different downstream tasks.

Key findings:

- Neural scaling behavior up-to 200M parameters and 10B tokens.
- All three downstream tasks reach the state-of-the-art comparing to baseline models. The scaling behaviors also exhibit in the downstream tasks.
- The larger model is more data efficient, requiring less annotated data to reach the same level of accuracy for downstream tasks.



Key Answers:

We found one way to scale pre-training.

The representation learned from FM pre-training are useful, even with a frozen FM backbone, we can surpass other models.

Larger FM leads to better pre-training accuracy, downstream task performance, and data efficiency.

FM4NPP: A Scaling Foundation Model for Nuclear and Particle Physics

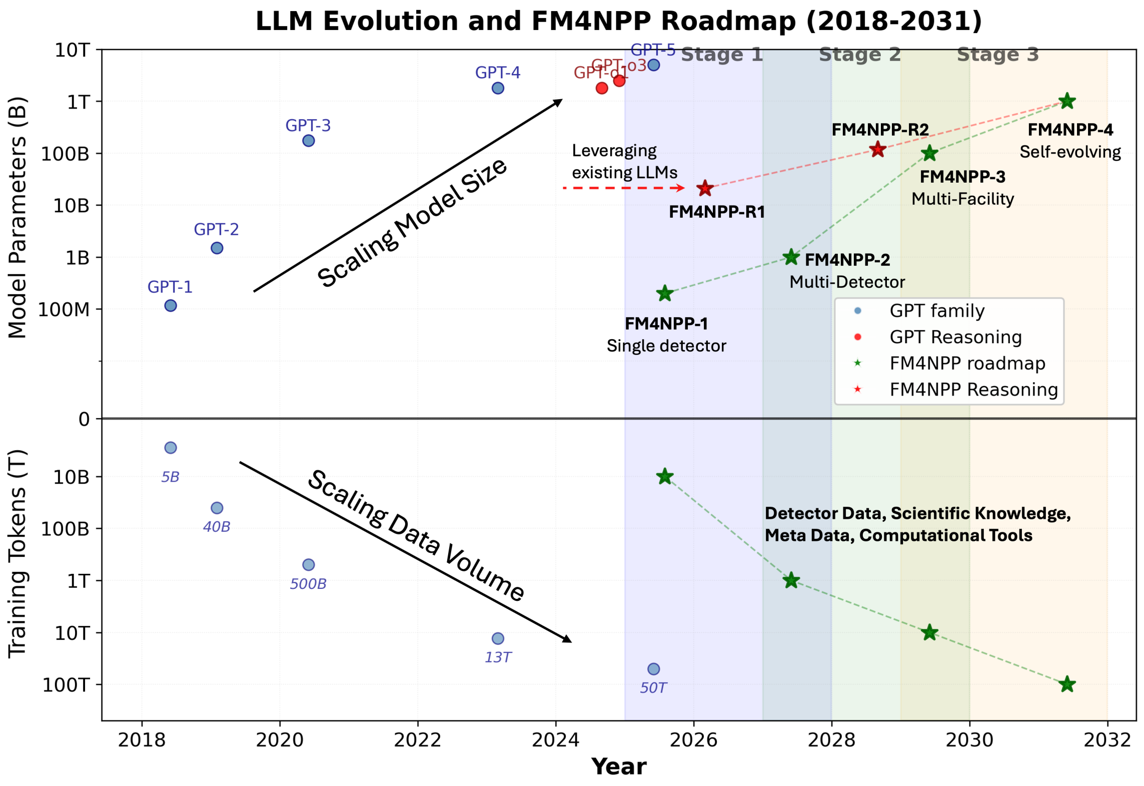
David Park*[1], Shuhang Li*[2], Yi Huang*[1], Xihaier Luo[1], Haiwang Yu[2], Yeonju Go[2], Christopher Pinkenburg[2], Yuewei Lin[1], Shinjae Yoo[1], Joseph Osborn[2], Jin Huang[2], Yihui Ren[1]

arXiv: 2508.14087; Published in ICLR 2026

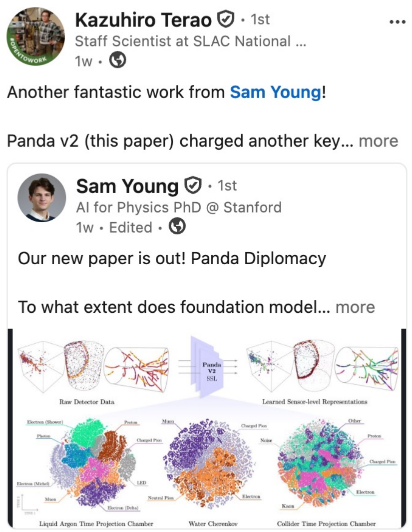
This is exciting, but leads to more questions:

- Will the scaling behavior of FM4NPP continue?
- How to integrate other sub-detectors?
- Will FM4NPP paradigm work beyond sPHENIX?
- How will FM interact with LLM? Etc.

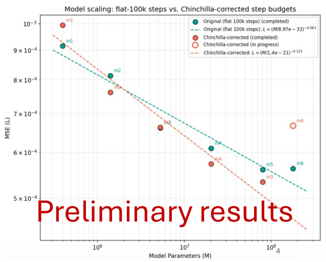
Ren, Y., Osborn, J. D., Brost, E., Yu, H., Abidi, S. H., Go, Y., ... & Ma, H. (2026). A roadmap toward scaling, reasoning and self-evolving foundation models for nuclear and particle physics. *International Journal of Modern Physics A*, *41*(15n16), 2650086.



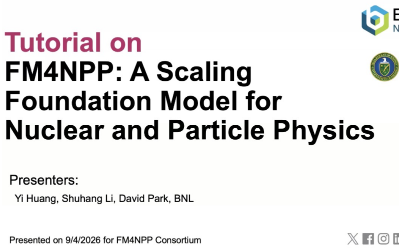
Open sourced code, dataset, pre-trained model, ICLR2026 reviewers' comments and rebuttals are open.



arXiv:2609.00611



Prof. Kyungseon Joo's group: scaling on CLAS12



Leading authors of FM4NPP: David, Shuhang & Yi gave tutorials, Yeonju & Jin org.
~ 60 attendees

Recording avail: indico.bnl.gov/event/34234/

FM4NPP: Foundation Models for Nuclear and Particle Physics

Pretrained state space model (Mamba / Mamba2) checkpoints for particle physics, from the paper Foundation Models for Particle Physics (accepted to ICLR 2026). Developed at Brookhaven National Laboratory.

- Paper: https://arxiv.org/abs/2508.14087 · OpenReview
- Code: https://github.com/FM4NPP/PP_collision
- Dataset (TPCpp-10M): https://doi.org/10.5281/zenodo.16970029

Reproduced by Google's team

FM4NPP Open Questions <=> Genesis Mission

- *Will the scaling continue with more data size and larger model size?*
  - DOE AmSC Intelligent Data Activities: HEP (TREASURE, Gabriele D'Amen) & NP (DaPP, James Dunlop)
- *Will FM4NPP generalizable to other detectors?*
  - "Foundation Models for Transferable Particle Tracking in Nuclear Physics" (Jan Bernauer, SBU)
  - "Developing a Foundation Model to Unify Neutrinoless Double Beta Decay Search Technologies" (Zepeng Li, UH)
- *Alternative backbone architecture?*
  - "Mixture-of-Experts Foundation Models for Scalable Holistic Reconstruction, Modeling, and Interpretation of Particle Interactions" (Cristiano Fanelli, W&M)
- *How to handle multi-modality data to unlock additional downstream tasks?*
  - "Multimodal Enhancements to the FM4NPP" (Ron Soltz, LLNL)
  - "Multi-modal and multi-facility application of the FM4NPP foundation model: silicon trackers and electron colliders" (Gunther Roland, MIT)
- *Will FM4NPP work on real experiment data?*
  - TREASURE + DaPP (sPHENIX), STAR data, MIT electron data. Zepeng's and Gunther's GM.
- *How to enable FM4NPP for high-level reasoning?*
- *What are historically collected data that can be readily used?*

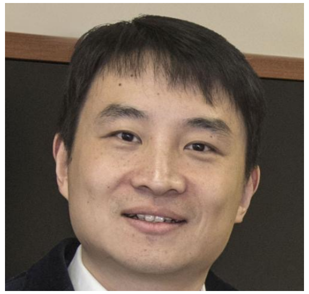
Pending: GM Phase-IIs, Japan's JST ARiSE; On-going NSF (NSF 26-512) AI Dataset call.
Great time to have such discussion 🚶➡️ parallel sessions & poster sessions

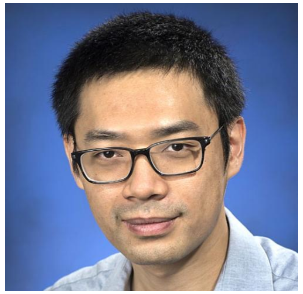
Team members from three BNL-participated FM4NPP Phase-I RFAs with LLNL, MIT, and SBU as lead are attending in person: Joe Osborn (BNL), Yeonju Go (BNL), Cameron Dean (ISU), Pedro Marin (ISU).

Acknowledgement

- We thank sPHENIX collaboration to share the simulated TPC data.
- Sponsor: BNL LDRD, SciDAC RAPIDS. New LDRD in FY27 (Joe & David).
- Computing: NERSC Award No. DDRERCAP-m4722.



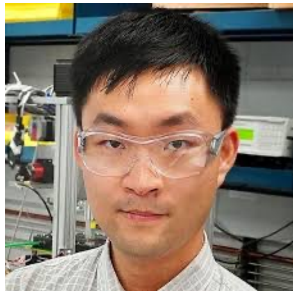





(AI Dept.) David Park, Yi Huang, Xihaier Luo, Yuewei Lin, Shinjae Yoo Yihui “Ray” Ren (yren@bnl.gov)







(Phys Dept.) Shuhang Li, Haiwang Yu, Joe Osborn, Yeonju Go, Jin Huang