Brookhaven National Laboratory

U.S. DEPARTMENT OF ENERGY

FY27 NPP LDRD Type B Pre-Proposal

Global HENP event reconstruction with Foundation Models

PIs: Joe Osborn (PO) and David Park (CDS)

January 9, 2026

@BrookhavenLab

FY27 NPP LDRD Type B Pre-Proposal

Proposal title: Global HENP event reconstruction with Foundation Models

Primary Investigator: Joe Osborn (PO), David Park (CDS) (both ECA eligible)

Other Investigators: PO: Jin Huang, Yeonju Go
CDS: Yihui (Ray) Ren, Yi Huang
Postdoc

Indicate if this is a cross-directorate proposal: Yes _X_ No___

If yes, identify other directorates/organizations: NPPS, sPHENIX in PO, AI in CDS

Proposal Term: From: 10/01/2026 To: 09/31/2028

Brookhaven National Laboratory

FY27 NPP LDRD Type B Pre-Proposal

Proposal title and brief abstract: Global HENP event reconstruction with Foundation Models

In high energy nuclear and particle physics, data reconstruction is a complex and CPU intensive task that involves using tens or hundreds of algorithms working together to produce high level physics objects, such as tracks or jets, which are then used to make scientific discoveries. These algorithms are typically developed by scientists to perform a single task for single or multiple detectors, reducing the data step by step. Foundation models (FMs) are a recent breakthrough in Artificial Intelligence (AI) research that utilize self supervised learning on unlabeled data. FMs are rapidly growing in use due to their ability to identify generalized representations that can be adapted to a variety of downstream tasks. These characteristics make FMs an excellent candidate to perform global event reconstruction on HENP datasets rather than relying on complex individual algorithms that must interface with one another. A prototype FM for track finding has already been demonstrated utilizing data from a time projection chamber. In this LDRD, we will utilize recent breakthroughs in FM research to demonstrate event reconstruction with low level tracking and calorimeter hit data to produce particle flow like objects that could be used in high level physics analysis. This work supports the overall Genesis mission by leveraging AI tools to enhance scientific discovery potential.

Program: NP

Return on Investment: Would leverage new AI technology for HENP, fitting into the overall Genesis mission

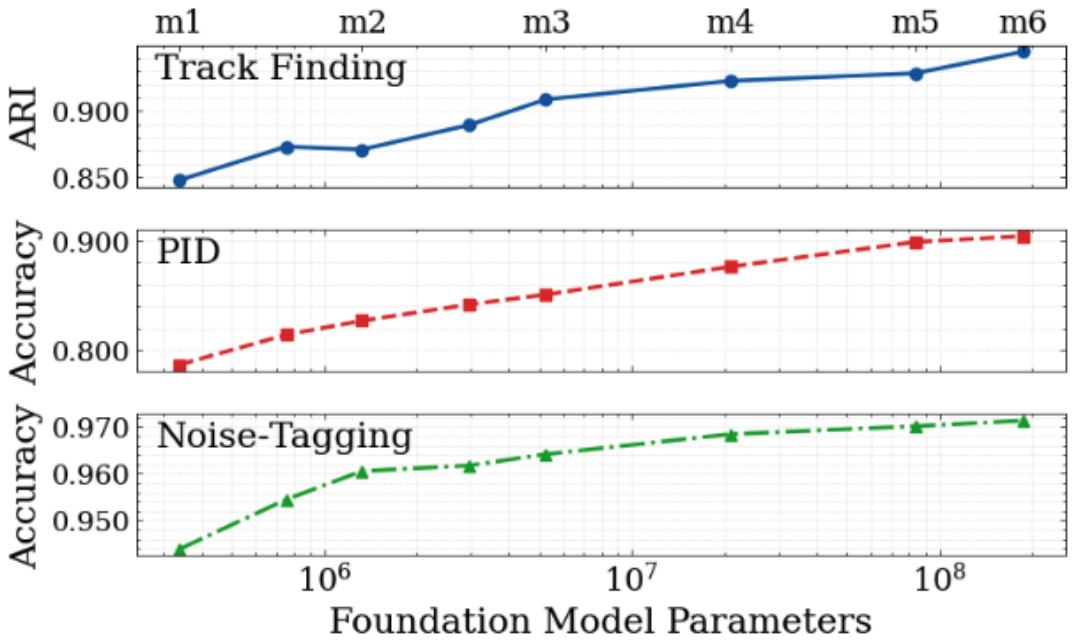
Broader impact on the activities at the laboratory: Expand AI and cross directorate collaboration at BNL

Total planned funding per year in FY27 and FY28: 250k in FY27 and FY28

FM4NPP – A prototype

- Initial prototype demonstrated on sPHENIX TPC data
- Research focused on validating expectations of Foundation Models
- FMs are expected to show:
  - Improved performance with model size ✅
  - Adaptation to multiple tasks ✅

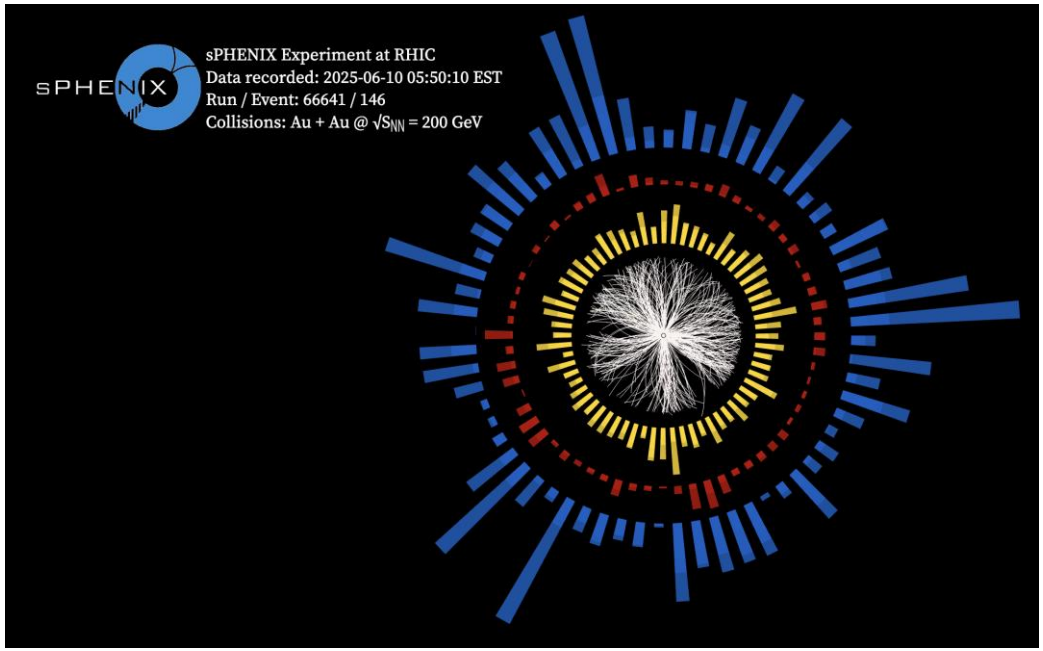
arXiv:2508.14087



Brookhaven National Laboratory

Towards Physics Analysis

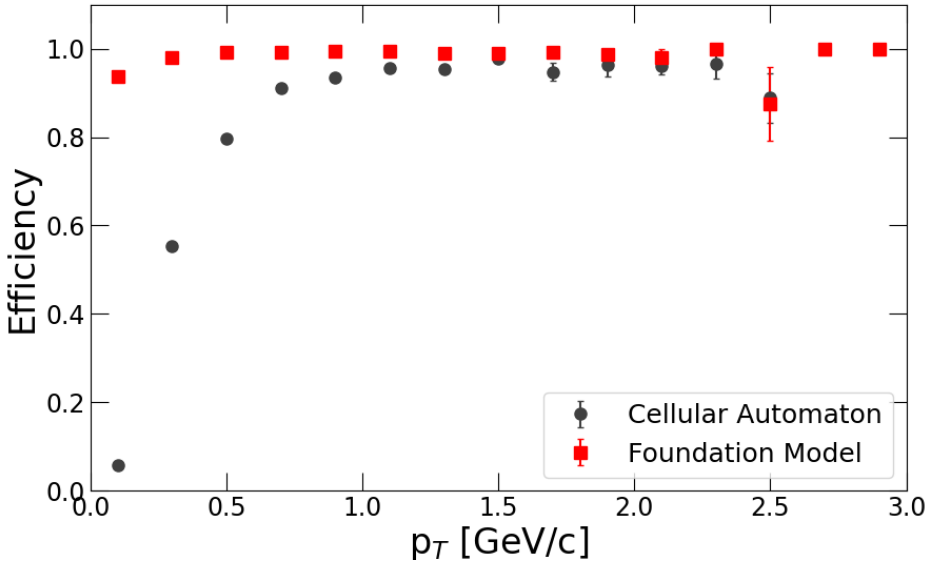
- Motivation for this LDRD – expanding to create a product that produces physics ready analysis objects
- Include silicon and calorimeter detectors to perform particle flow reconstruction from low level data (hits/clusters)
- Demonstrate physics ready object measurement (e.g. resonance or jet reconstruction)

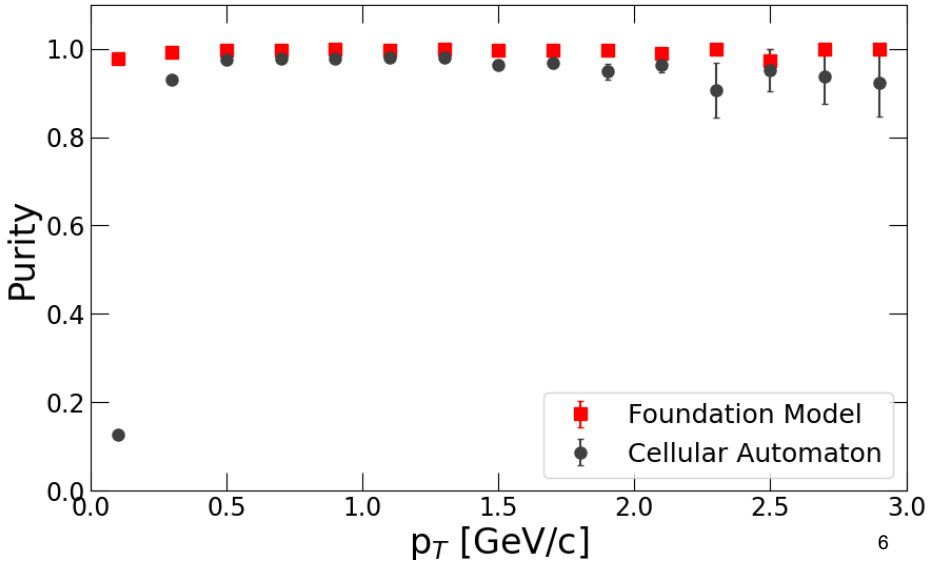


Brookhaven National Laboratory

Initial Demonstration

- Initial exploration showing potential physics improvement in TPC only track reconstruction, comparing FM4NPP and “traditional” Cellular Automaton based method
- Significantly improved efficiency and purity for track seeding





Brookhaven National Laboratory

Deliverables and Impact

- Deliverable: given input low level reconstruction data (hits,clusters), demonstrate jet and/or resonance reconstruction with a FM using all sPHENIX subsystems
- Broader Impacts
  - Would be impactful for collider based reconstruction in both NP and HEP
  - Cross division collaboration at BNL, supporting applied AI research
- Tasks/milestones
  - Tokenize silicon and calorimeter sPHENIX subsystems (Y1)
  - Demonstrate global reconstruction capabilities (Y1)
  - Extend to higher complexity systems (e.g. Au+Au) (Y2)
  - Integration into a production workflow (Y2)
- Potential future research directions
  - Domain adaptation/generalization between experiments/colliders (e.g. EIC/LHC)
  - Prepare laboratory for future funding opportunities emphasizing advanced data analysis methods within the Genesis mission

Conclusions

- Foundation Models have been shown to be very good at generalizing information for many downstream tasks
- In this project, we aim to develop a FM that can be applied to a wide variety of detector data to perform global event reconstruction
- Future research directions aplenty
  - Domain adaptation between experiments/colliders
  - Additional reconstruction and/or physics analysis related tasks
  - Comparison to existing algorithms
- Aiming to identify further funding from AI related calls within NP/HEP e.g. within Genesis mission