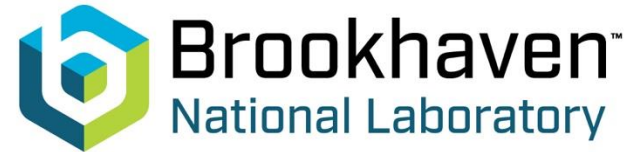


Improving LArTPC Clustering through AI-Based Scientific Reasoning and Fine-grained Domain-specific Tools toward Enhanced Discovery Capability in DUNE and SBND



Principal Investigator(s):

Haiwang Yu (PI, PO, Presenter) , Yi Huang (co-PI, CDS)

List of the proposal participants and their organizations if other than NPP:

Yousen Zhang (PO), Xin Qian (PO), Brett Viren (PO), Yihui Ren (CDS)

ECA eligibility: **No**

Proposal term from: **Oct. 2026** to: **Sep. 2028**



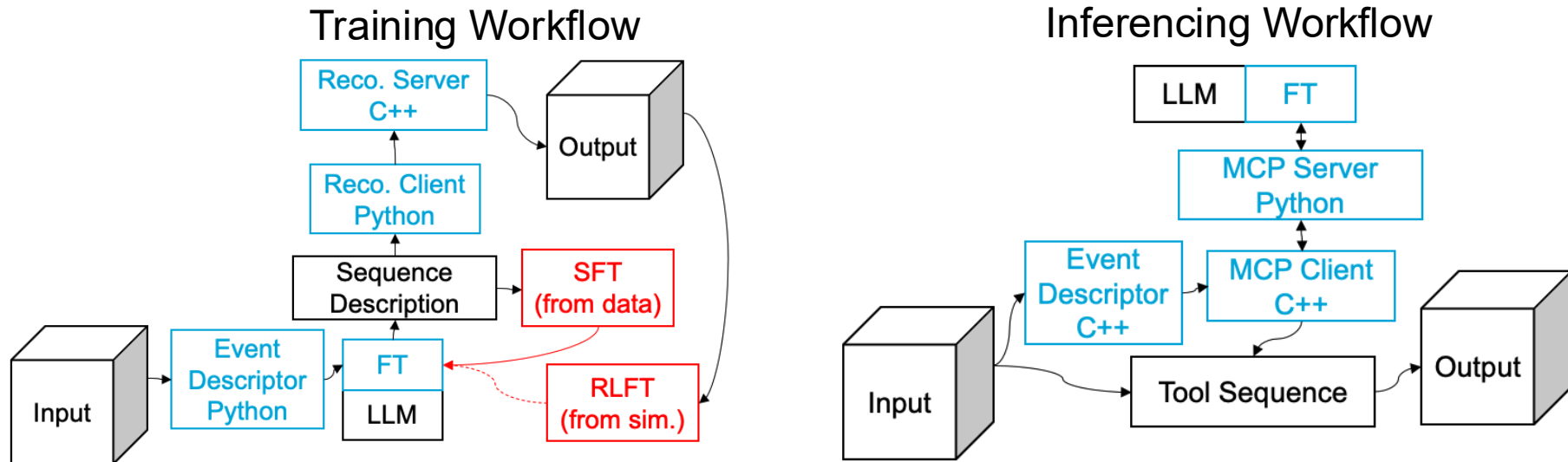
Annual funding: FY27: **\$250k** FY28: **\$250k**

Description of the LDRD Proposal

Goals & Deliverables

- A validated workflow to enhance LArTPC reconstruction, with photon-related topologies as a representative use case. The targeted performance improvement is **from the current 56% to the 75–85% range**.
- An extensible MCP-based framework integrated into Wire-Cell that is capable of algorithm-level tool orchestration.
- A reusable platform for testing and advancing AI methodologies in HEP reconstruction.
- (Extension) Photon-related analyses in SBND using the proposed workflow conducted through collaboration with universities.

Method:



Description of the LDRD Proposal

Timeline with Milestones

Period	Task	Milestone
FY27, H1	Package core clustering algorithms as MCP tools; LLM selection; Define and develop the event descriptor schema.	A library of modular C++ reconstruction tools, a selected LLM and a finalized event descriptor format.
FY27, H2	Develop the C++ Reconstruction server prototype; Build the initial Python fine-tuning framework.	A functional Reconstruction server capable of executing requested algorithms and a Python framework that can interface with the server to begin model training.
FY28, H1	Conduct supervised fine-tuning with human labeling; Perform initial validation and performance studies on photon-rich events.	An initial validated workflow based on a supervised-tuned model that shows performance improvements over traditional fixed workflows. (relative 10 – 20% improvements)
FY28, H2	Apply Reinforcement Learning Fine-Tuning (RLFT) to further optimize the workflow; Finalize and integrate the MCP-based framework into Wire-Cell; Complete documentation for the reusable platform.	A final, validated workflow incorporating RLFT, fully integrated with Wire-Cell and released as a reusable platform for the community. (relative 30 – 50% improvements)

Personel

- Haiwang Yu (PI, PO): Experienced in Wire-Cell and AI/ML
- Yi Huang (co-PI, CDS): Experienced in AI/ML for scientific physics
- Yousen Zhang (PO): Experienced C++ and Python developer
- Xin Qian (PO): Lead algorithm developer of Wire-Cell
- Brett Viren (PO): Lead architect of the Wire-Cell Toolkit
- Yihui Ren (CDS): Experienced in AI/ML for scientific applications

Intellectual Merit Summary

- **A novel framework that simultaneously achieves robustness and high performance**, by tightly coupling
 - **Robustness** from well-tested domain knowledge embedded directly within physics-driven reconstruction tools, and
 - **High performance** enabled by large language models (LLMs) leveraging industry-scale innovation.
- **Pioneering the systematic capture of human scientific reasoning** through supervised fine-tuning, translating expert decision-making into machine-actionable knowledge.
- **Advancing beyond current agentic AI paradigms**, which primarily emphasize high-level workflow orchestration, by enabling reasoning and decision-making at the **algorithmic and physics-tool level**, where scientific fidelity and performance are determined.

Return on Investment or Potential Future Funding

The proposed work is well aligned with the objectives of the Genesis Mission and directly supports DOE investments in AI-enabled scientific discovery. The developed agentic AI workflow is expected to benefit the DUNE experiment and establish a technical foundation that can be leveraged for future funding opportunities in AI for High Energy Physics (HEP). **In particular, this work positions BNL to compete for Genesis-aligned funding released through the Office of Science, including HEP and cross-cutting AI-related solicitations.**

Beyond HEP, the proposed agentic AI and fine-grained domain-tool integration framework is broadly applicable to other data-intensive scientific domains. **As such, the outcomes of this LDRD are expected to be relevant to future funding opportunities from the DOE Office of Advanced Scientific Computing Research (ASCR) and other Office of Science programs supporting AI infrastructure, scientific workflows, and methodology development.** These opportunities are anticipated to emerge over the FY 2026–FY 2028 timeframe.

The Broader Impact on the Laboratory

- Expands BNL's core AI capabilities by advancing AI reasoning at a fine-grained, algorithmic level. This approach is broadly applicable across scientific disciplines and strengthens the Laboratory's science and engineering capabilities.
- Enhances BNL's leadership in reconstruction algorithms and AI research, positioning the Laboratory to effectively pursue emerging AI/ML-focused funding opportunities across DOE programs.

Names of Suggested BNL Reviewers

- Viviana Cavaliere, PO
- Jin Huang, PO
- Yuewei Lin, CDS
- Dmitrii Torbunov, CDS

If ECA Eligible – How this LDRD Benefits Your Application

1 slide

LDRD Funding Table from BOM

Title: Improving LArTPC Clustering through AI-Based Scientific Reasoning and Fine-grained Domain-specific Tools toward Enhanced Discovery Capability in DUNE and SBND
PI: Haiwang Yu/Yi Huang

Resource Category	DESCRIPTION	FY27	FY28
050 Salary - Scientific		14,219	14,787
051 Salary - Research Assoc		43,596	45,340
050 Salary - Professional		78,669	74,482
050 Salary - Technical		0	0
050 Salary - Management & Admin.		0	0
Total FTEs		0.87	0.84
TOTAL SALARY/WAGE & FRINGE		136,483	134,610
various Contracts - Low Value		0	0
280 Foreign Travel		0	0
290 Domestic Travel		0	0
various Purchases		10,000	10,000
TOTAL MSTC		10,000	10,000
TOTAL DIRECT COSTS		146,483	144,610
251 Electric Distributed (Electric Power Burden)		1,365	1,346
700/701/481 Organizational Burden		24,878	26,764
TOTAL ORGANIZATIONAL BURDEN		26,243	28,111
745 Procurement (Material Handling)		700	700
735 G&A Burden		0	0
730 Common Institutional Support		76,574	76,580
722 Safeguards & Security Assess		0	0
TOTAL LABORATORY BURDEN		77,274	77,280
705 LDRD Burden		0	0
TOTAL PROGRAM COSTS		250,000	250,000
740 Full Cost Recovery		0	0
TOTAL PROGRAM COSTS		250,000	250,000

Labor Band	Name	FY27		FY28	
		FTE	Amount	FTE	Amount
PROF2	Haiwang Yu	0.22	41,481	0.19	35,807
PROF2	Yi Huang	0.20	37,187	0.20	38,675
RA1	Yousen Zhang	0.40	43,596	0.40	45,340
SCI2	Yi Hui Ren	0.05	14,219	0.05	14,787
Total		0.87	136,483	0.84	134,610