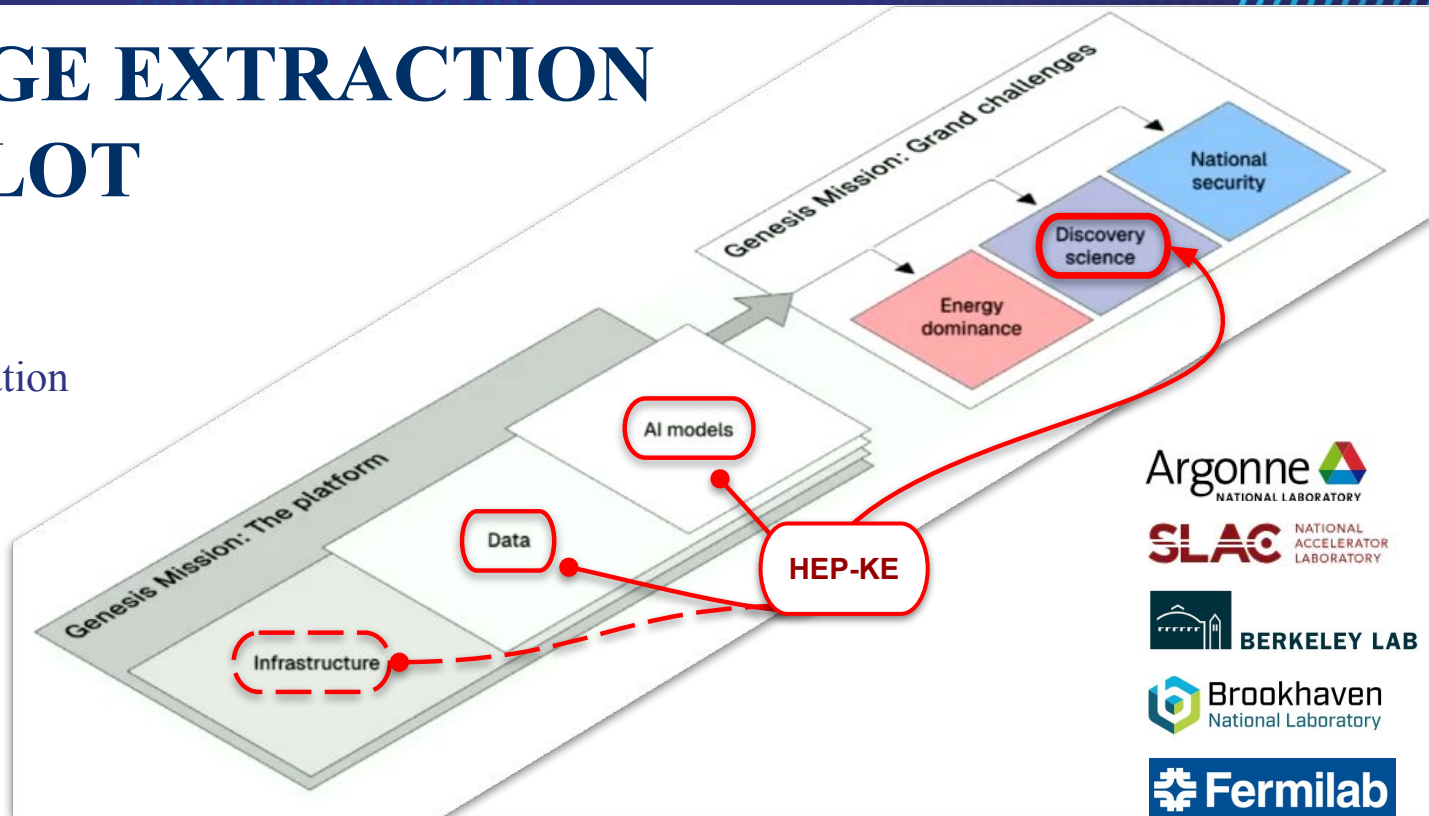


KNOWLEDGE EXTRACTION IDA HEP PILOT

CLARIPHY - AI Collaboration
Meeting, September 2026

Nesar Ramachandra
nramachandra@anl.gov



Argonne
NATIONAL LABORATORY

SLAC
NATIONAL ACCELERATOR
LABORATORY

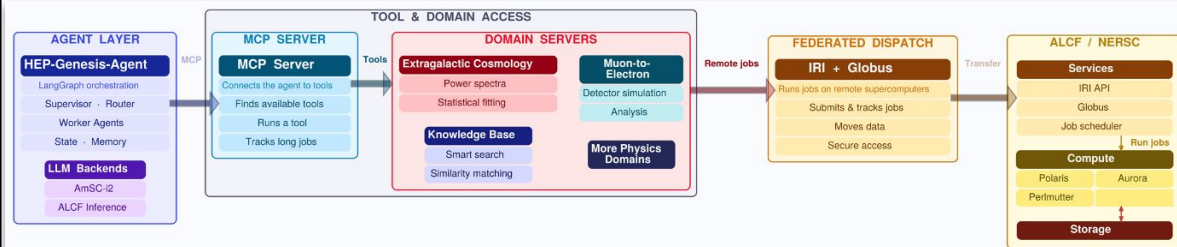
BERKELEY LAB

Brookhaven
National Laboratory

Fermilab

Summary and overview of the project

Overall goal: Develop an HEP-focused framework for **agentic AI** that emphasizes specialization and modularity, where reasoning models **orchestrate** HEP foundation models, simulators, data sources, and numeric and symbolic analysis tools, with **applications such as** data resurrection and investigations of dark matter physics serving to ground and demonstrate the approach.



Flowchart of Knowledge extraction agentic-frameworks. Products from American Science Cloud and Genesis mission are closely integrated for AI-assisted HEP research.

Technical goals:

- Customized agentic workflows for HEP and Cosmology
- Featured with exa-scale data, computationally expensive simulations and data processing, and sophisticated statistical inference algorithms.

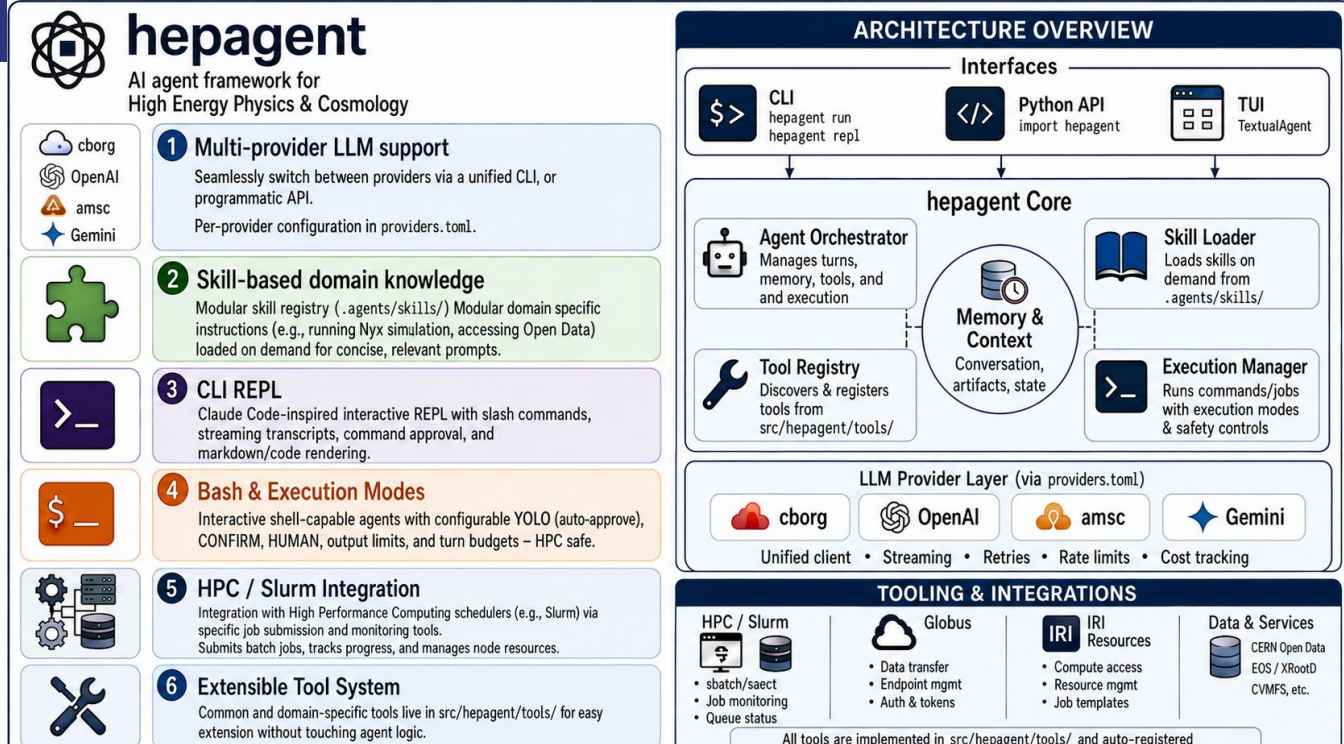
Science goals:

- Resurrection of experimental HEP data from the legacy experiment, re-analyse with modern tools
- Find subtle dark matter signal in simulated Lyman-alpha forest data
- Explore cross-domain scientific problems

Products via HEP-Knowledge extraction: infrastructure

Agentic components:

- Flexible LLM backends
- Multiple sub-agent configurations
- MCP servers per science sub-domain
- Tools for HPC submissions, Globus for transfers, data servers
- Skills, memory files
- Knowledge bases and graphs
- Close integration with AmSC and Genesis platform

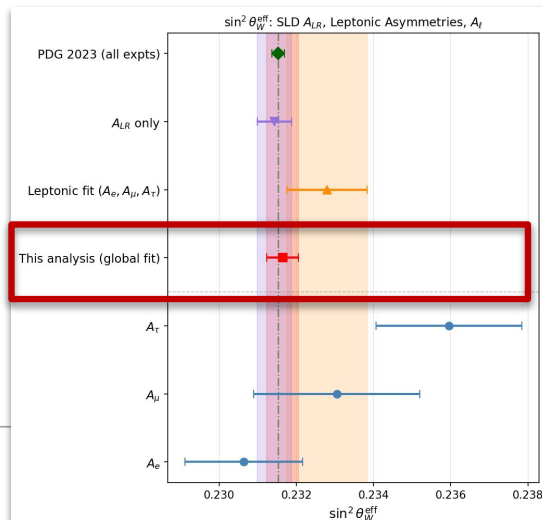


Frameworks being tested for **Lyman-alpha forest data analysis**, collider physics analysis using **CERN open data**, **cosmic emulators**, **HEP archival data analysis**.

Results via HEP-Knowledge extraction

AI-Enabled Data Resurrection and Analysis

- **Task:** resurrect archived SLD data (polarized e^+e^- at SLAC, ~1990s);
- **How:** agents comb old Jazelle registers; convert to modern formats; re-analyze precision electroweak observables, including weak mixing angle extraction
- **Results:** fits land on the PDG 2023 average—a validated reanalysis of legacy DOE precision data



Cheng, Hobbs et al.,
Accepted MLST, 2026:
<https://arxiv.org/abs/2606.00224>

A Tale of Two Teams: Scientists Versus the Multi-Agent System

- **Task:** Explore cosmology simulations to find interesting galaxy clusters
- **How:** experts with or without agent explore independently
- **Results:** Both teams came up with unique results, with team-AI performing better in terms of artifacts

Wells,
Ramachandra et
al., Accepted
Supercomputing-
TPC workshop,
2026

