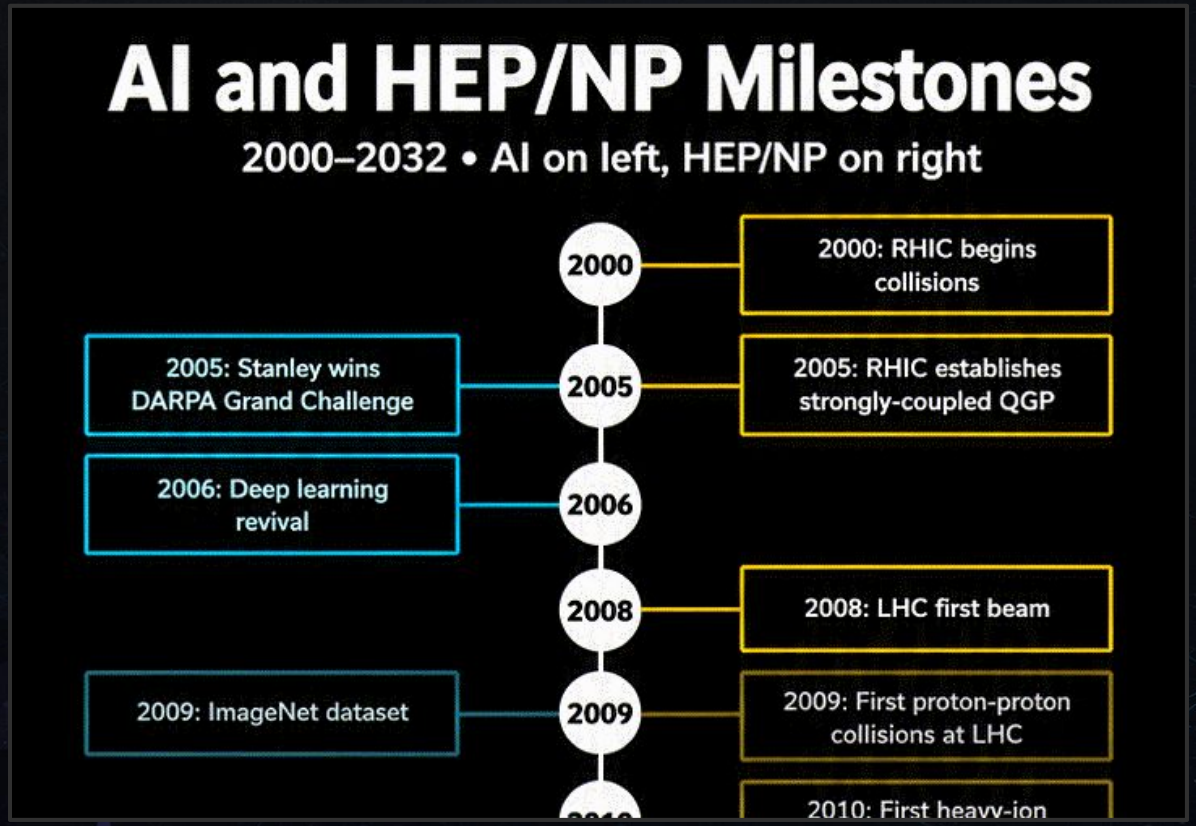
The AI Revolution and Nuclear Physics

David Lawrence, JLab
July 13, 2026

Milestones in AI and HENP

AI **HENP**



HENP has generally been good at adopting new technologies*.

AI is no exception. (See the last 3-4 CHEP conferences).

This technology is outpacing construction of the EIC.

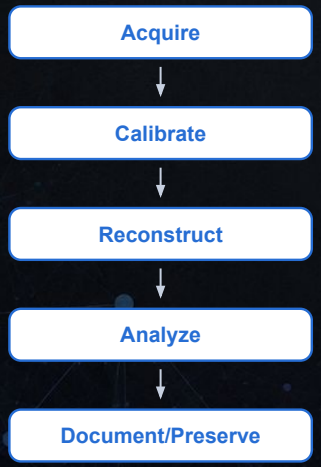
*cernlib and Fortran users excused

The revolution is a workflow revolution

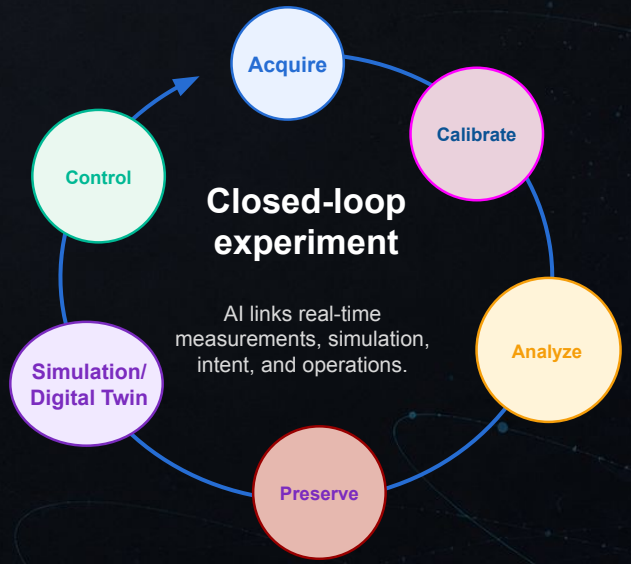
The opportunity is not “use a model”; it is to redesign the scientific loop.

Today

Sequential pipeline



Tomorrow



We design and execute experiments to meet measurement goals.

AI will allow dynamic adjustment of experiments to achieve the physics goals while minimizing resource usage.

Automated, holistic calibrations of interdependent systems will enable experiment pipelines*

**ePIC already incorporates much of this type of workflow in its design.*

Sources: CHEP Track 6: AI/LLM tools · CHEP Track 8: analysis + open data · CHEP Track 1: FAIR data

Coding with AI

AI is changing the way we code and will likely continue ..

CLAUDE CODE

CURSOR

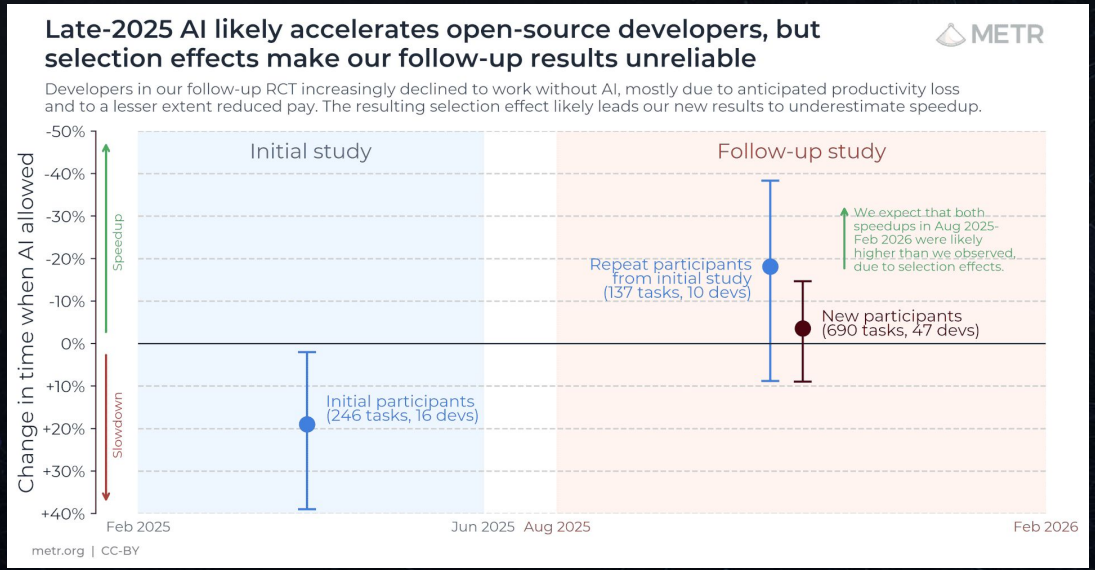
Ask Copilot

Codex

Vibe Coding: *"I describe the vibe of what I want, the AI hands me a complete app, and I pray it works."*

Agentic Coding: *"I give an autonomous AI agent a high-level goal, and it manages its own step-by-step execution, testing, and debugging loop."*

https://metr.org/blog/2026-02-24-uplift-update/



Studies of improved development time are very mixed but tend to say 25%-50% more FTEs when using AI for coding.

Experienced coders can expect to drive AI agents better.

Future coders will train their brains to think differently about code development.

"What language will people be coding in 10 years from now?"

"Probably English."

- overheard at CHEP 2026

Data as a first-class product

F.A.I.R. Data Principles

The first step in (re)using data is to find it. **Metadata** and data should be **easy to find** for both **humans** and **computers**.

Once found, need to know how it can be **accessed**, possibly including **authentication** and **authorization**.

Able to **integrate with other data**. Interoperate with **applications** or **workflows** for **analysis**, **storage**, and **processing**.

Metadata and data should be **well-described** so that they can be **replicated** and/or **combined** in **different settings**.

AI-Ready Data

"AI-Ready" ≠ File format

*"AI Readiness for Data refers to a spectrum of how ready a given set of data is to be used by an AI process **without need for further pre-processing or curation.**"*

- Kirkpatrick & Chuang @ SciDataCon2025
https://doi.org/10.5281/zenodo.17368124

AI readiness is not a one-time binary state, but an ongoing process as AI methods, workflows, and data uses evolve.

We are now in a state of transition …
… the EIC will operate on the other side.

- Literature is how we convey information. Especially to the next generation of scientists.

- Traditionally, only small expectation that event-level data will be useful years from now.
  - All the useful information was already published
  - Effort it would take to access/process/analyze old data by non-expert is too high of a bar

- Agentic tools are lowering this bar substantially, but only if the data is well-documented enough for the agent to figure it out.

➔ *Coding (by humans) is becoming much less valuable than it used to be*

➔ *Making data accessible to auto-generated code is becoming invaluable*

What the EIC can do

Generate DPI for data as it is acquired-

Digital persistent identifier (DPI or digital PID) – A digital identifier that is globally unique, persistent, machine resolvable and processable, and has an associated metadata schema* (like an ORCID or DOI)

** DOE Public Access Plan 2023 https://doi.org/10.11578/2023DOEPublicAccessPlan*

Generate DataCards to store alongside the data as it is stored-

Includes an overview and context, organization & file structure, data details, and methodological info
Human and AI readable
Published and unpublished datasets

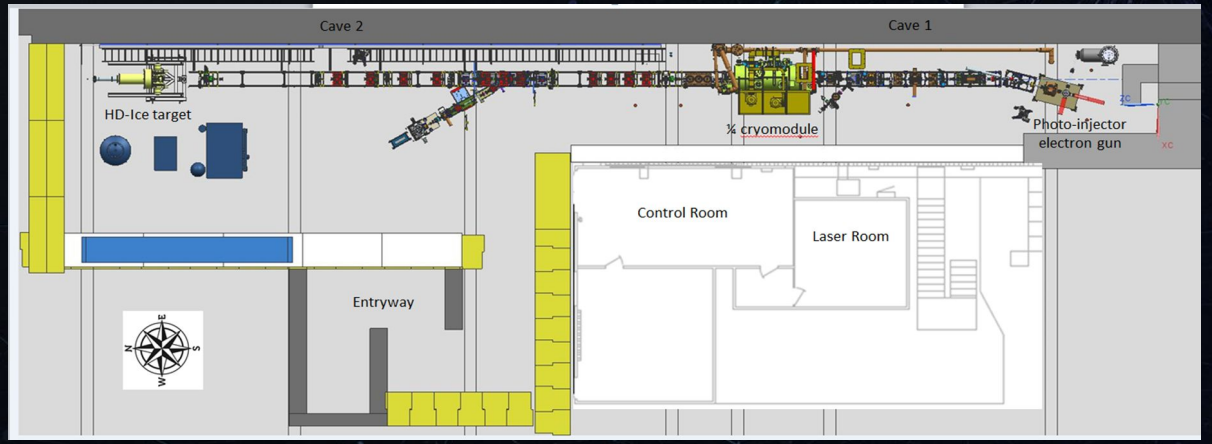
see Laura Biven's talk at the recent JLab User's Organization meeting:
https://indico.jlab.org/event/1000/

Large Language Models (LLMs)

Facility Context: UITF

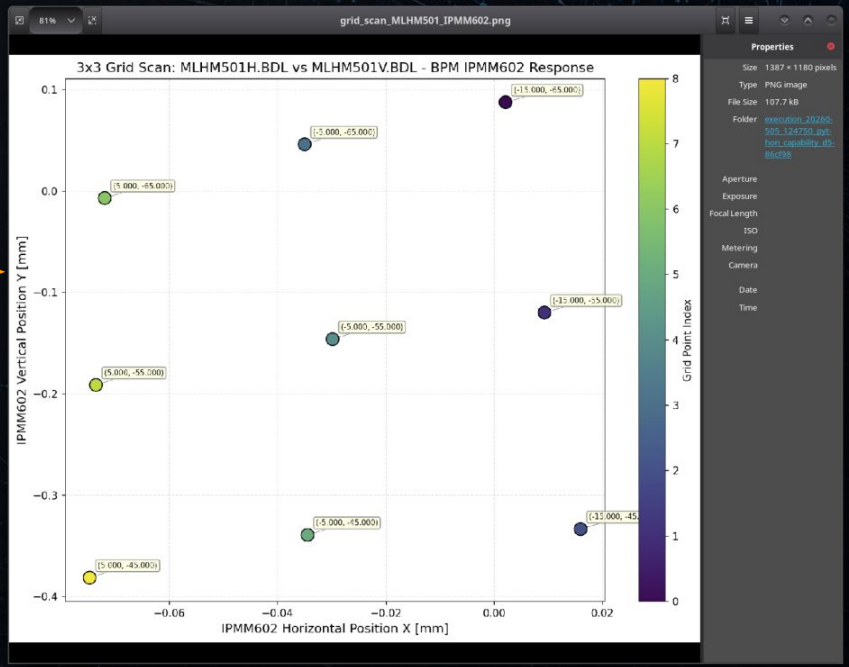
- The Upgrade Injector Test Facility (UITF) is a test bed to evaluate new hardware destined for CEBAF
  - e.g. Wien filters, 200 kV photogun, quarter cryomodule, non-invasive polarimeter
- UITF is also a member of BeamNetUS – an organization of test accelerators that offers beamtime to external Users

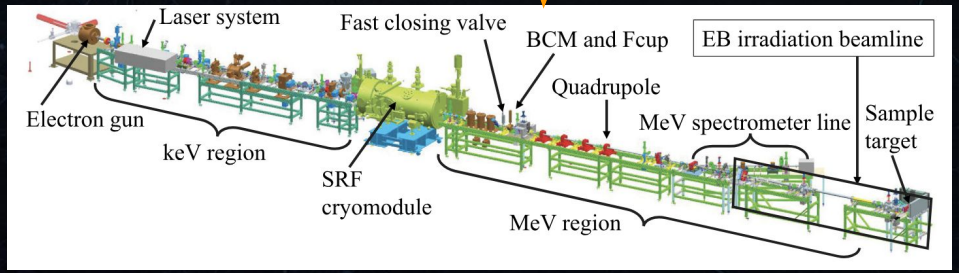
- For the AmSC Osprey Demo:
  Build a channel database for most of the archived PVs
  - *489 PVs (BLMs, BCMs, BPMs, RF cavities, correctors, quad-ru poles, solenoids, dipoles, ion pumps, Wien filter)*



Demo: Control System WRITES

PROMPT: "Vary the integrated field of MLHM501H.BDL and MLHM501V.BDL in a 3 by 3 grid scan from their baseline values in steps of 10, e.g. for each setpoint of MLHM501H.BDL, scan three steps of MLHM501V.BDL. Before each setpoint, ask for verification. Wait 7 seconds for signals to settle and record the measured horizontal and vertical beam positions from BPM IPMM505. Return corrector values to their baseline values. Plot X vs Y for all 9 points."



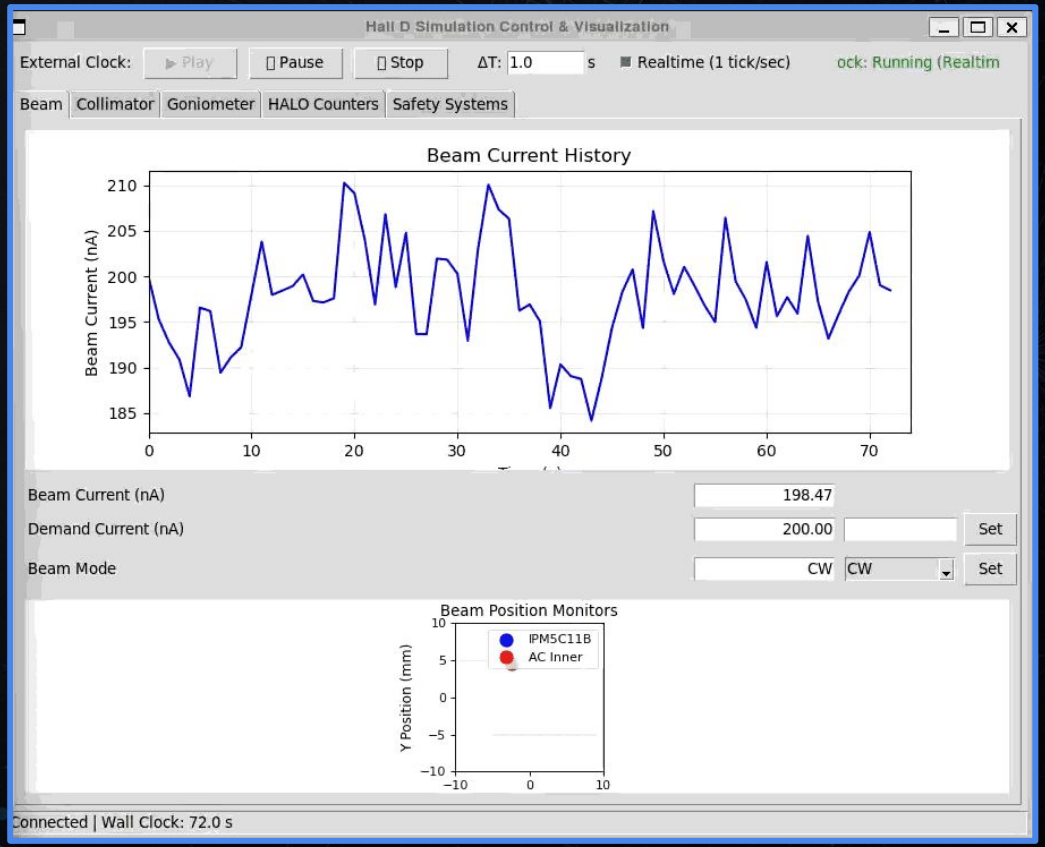


Hall-D Online Digital Twin

Natural Language is being developed as the communication protocol between **multiple AI agents** representing the **accelerator** and **experiment**.

Written instructions developed for human operators can be used for AI operator agents.

Each agent implements its own **independent guardrails** for the systems it represents.



Genesis Mission
and the American Science Cloud

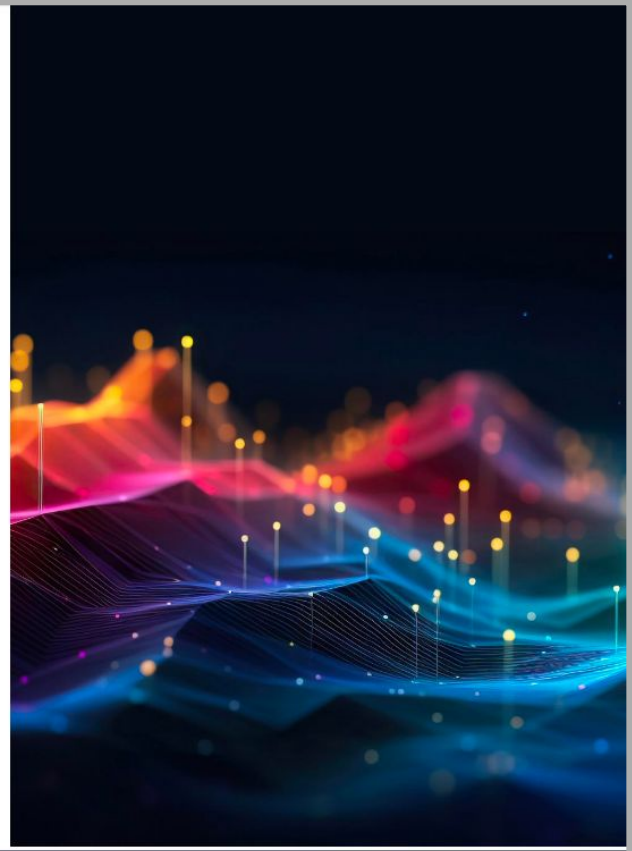
Genesis Mission

About the Genesis Mission

The **Genesis Mission** is a historic national effort to catalyze new industries, create high-skill jobs, and usher a new golden era of American discovery through artificial intelligence innovation.

The Genesis Mission's **American Science and Security Platform** will connect the world's best supercomputers, AI systems, and next-generation quantum computers with the most exquisite scientific instruments in the nation, and its intelligence layer will be trained with the singular scientific datasets and expertise housed in the National Laboratories.

Once complete, it will be the world's most complex and powerful scientific instrument ever built.



U.S. DEPARTMENT of ENERGY

Genesis Mission

Genesis Mission Pillars

Energy Dominance

AI will be used to accelerate sustainable fusion power, optimize advanced nuclear reactor design and operation, and enable a more intelligent and resilient electrical grid.

Discovery Science

AI will be used to illuminate molecular dynamics, unify data to understand the universe from quarks to cosmos, and generate new quantum algorithms.

National Security

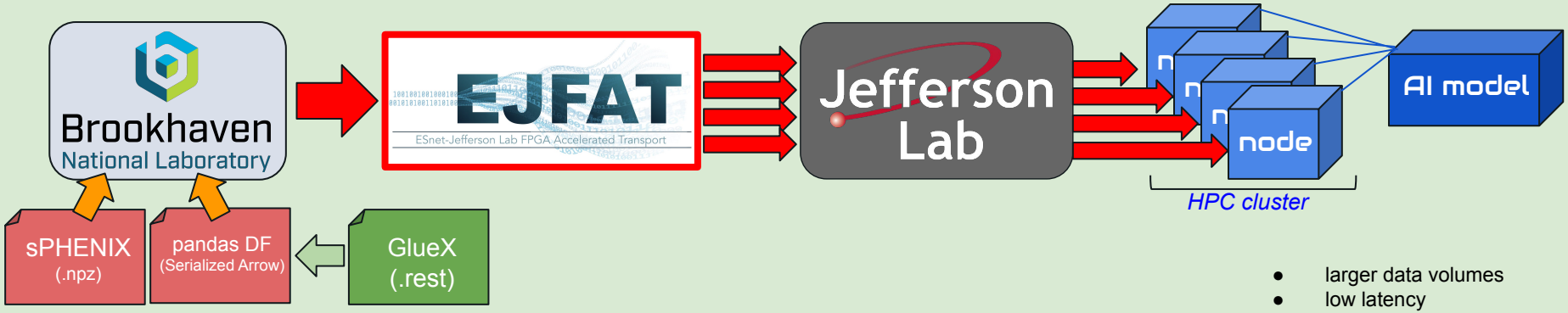
AI will be used to secure critical materials, accelerate advanced manufacturing, and discover mission-ready materials for defense and industry.

U.S. DEPARTMENT of ENERGY

D2.1: NP Workflow Demonstrator

(Thrust 1: User data and science)

What this will demonstrate: Streaming large detector datasets between facilities in AI-ready format, distributing to multiple nodes in an HPC (or HTC) cluster for processing



- larger data volumes
- low latency
- real time streaming

What pieces can be supplied by AmSC:

Virtual Organization:	Authentication, Access Control, Allocations
Data Services:	Preprocessing, Curated S/W stacks, *(Inference)*
Model Services:	Model Catalog
At-Scale Services:	Training, Inference
IRI-API:	WAN Data Fabric (ESnet) L3VPN

Partnerships:

This leverages data and models from the FM4NPP effort that is the basis of the NPDaPP-QCD: Nuclear Physics Data Providers Program - QCD AmSC project

BNL:
Meifeng Lin
Mohammad Atif

JLab:
David Lawrence
Nathan Brei
Armen Kasparian
Raiqa Rasool

Genesis Mission

Genesis Mission Funding

American Science Cloud: $150M OBBB Act end of FY25

$30M The Transformational AI Models Consortium (ModCon) (DE-SCL0000139)

…

$50M Data Providers Program (BER, BES, FES, HEP, NP)

$10M for NP

$3.67M **BNL** *Preparing QCD Data for Foundation Models* (DE-SCL0000128)

$4.00M **JLab** *Developing AI-Ready Data Framework for DOE NP Particle Accelerators* (DE-SCL0000128)

$1.97M **LBNL** *AI/ML-Ready Data Labeling for Low Energy Nuclear Physics* (DE-SCL0000128)

Genesis Mission RFA: $293M DE-FOA-0003612 (see Executive Order 14363)

- ~5k applications
- Public announcement of awards expected July 22, 2026

Disclaimer: While I work at a DOE funded lab, I am not a DOE employee and the information here is not an official statement from the DOE

Predictions

These are my own opinions …

- English will be primary programming language
- Focus shifted to Data
  - Much better meta-data to allow non-collaborators to access data in meaningful way
- Calibrations will be done more holistically and, consequently, faster
- Younger scientists will become less expert in ROOT, LaTex, bash, …
- Large dependency on agents
  - e-mail/messaging, scheduling, etc.
  - “Reading” papers will become more conversational
- Compute resources will become more in demand/scarce … or, the opposite
  - easier for one or two people to create huge campaigns increasing demand
  - more formally competitive process for using compute
  - manufacturers may respond to current commercial demand causing excess supply
- Cross-disciplinary research will become easier

If you take only two things from this talk …

1. Start thinking of data as similar to publications as a product of the experiment.

2. Embrace agentic coding so you can wield it professionally.

If you want to see a really cool talk on this subject check out Daniel Murnane's site here:
https://www.danielmurnane.com/ETH-Agentic-AI-for-Collider-Physics/

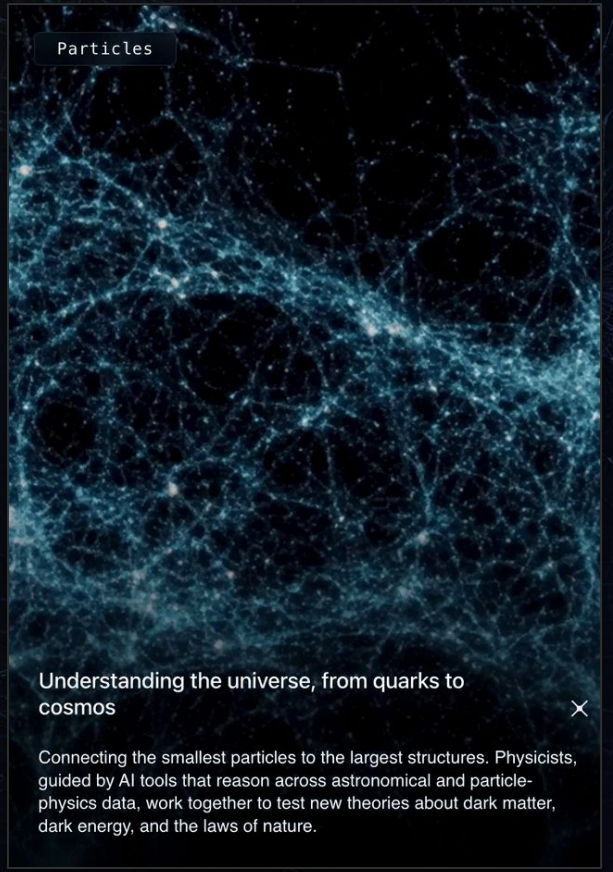
Backups

Genesis Mission

A National Mission to Accelerate Science Through Artificial Intelligence

Goal:

Genesis Mission will develop an integrated platform that connects the world's best supercomputers, experimental facilities, AI systems, and unique datasets across every major scientific domain to double the productivity and impact of American research and innovation within a decade.



Genesis Mission

The Labs provide leadership across five teams,
as well as leadership for the American Science and Security Platform

National S&T Challenges

Focusing on how AI, high performance computing, and data are being applied to address the U.S. government's cross-cutting priorities.

Data

Building a continuously evolving data ecosystem that curates, connects, and prepares DOE and NNSA data.

Infrastructure

Delivering the software and hardware infrastructure for the American Science and Security Platform.

Models

Delivering a comprehensive portfolio of AI models, agents, and tools.

Public-Private Partnerships

Supporting connections and activity between DOE, NNSA, and external orgs, other government agencies, and international allies.

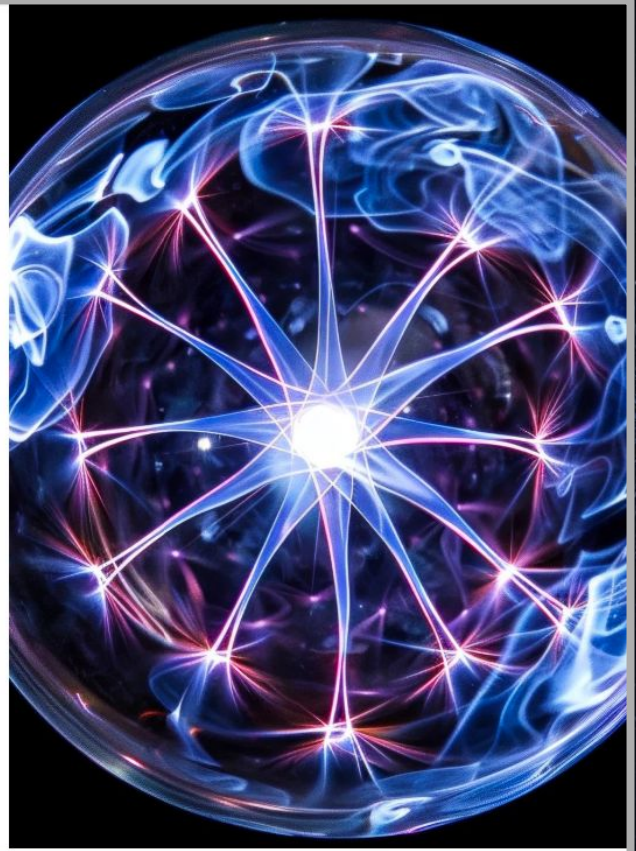
Genesis Mission

The American Science and Security Platform

The core technology engine of the Genesis Mission

The platform integrates America's most advanced high-performance computing systems, scientific experimental facilities, data resources, and production capabilities into a single, coordinated AI-driven discovery system.

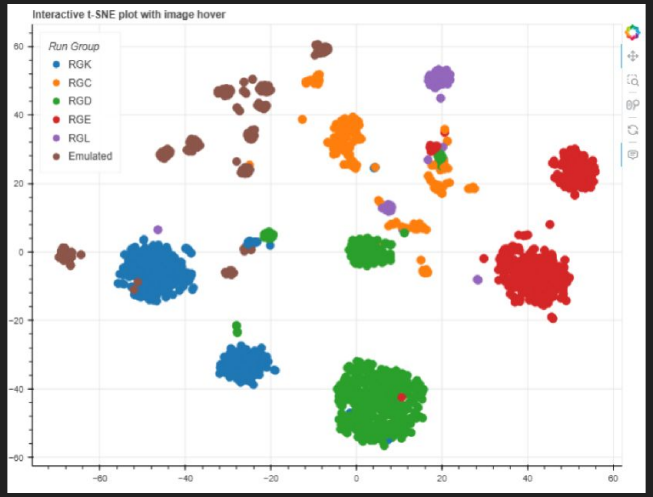
The result is one of the most exquisite scientific instruments humans have ever built—an instrument that rapidly turns ideas into trustworthy scientific output, strengthening U.S. competitiveness and amplifying national security.



U.S. DEPARTMENT of ENERGY

New Feature Developments in Hydra

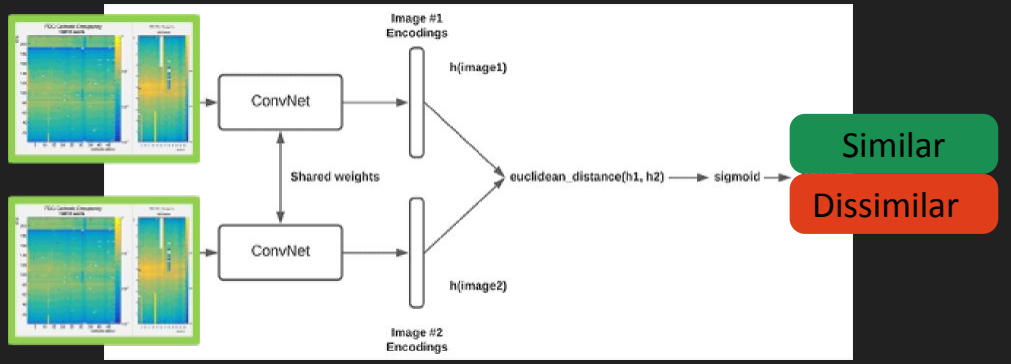
Semi-Automated Labeling



t-distributed stochastic neighborhood embedding: similar objects are modeled by nearby points and dissimilar are modeled by distant points. This link, https://distill.pub/2016/misread-tsne/ contains fun visualizations and explanations.

Automating Comparison to Reference Plots

Siamese models learn to compare two inputs (in our case images) and determine how *similar* they are. Think of this as Hydra's equivalent to the shift crew comparing images to their reference.



Work being done by grad. students Darren Upton(ODU) and Jordan O'Kronley (UT)

Streaming Data Acquisition

https://indico.cern.ch/event/1471803/contributions/6910053/attachments/3282186/5865711/CHEP2026_DrSai_KeLi_V3.1.pdf

https://indico.cern.ch/event/1471803/contributions/6967027/attachments/3280639/5862516/AIE4ML_CHEP26_Dimitrios_Danopoulos.pdf

https://indico.cern.ch/event/1471803/contributions/6967268/attachments/3280862/5863122/main.pdf

https://www.danielmurnane.com/ETH-Agentic-AI-for-Collider-Physics/ Daniel Murnane “Agentic AI for Collider Physics”

What we are looking for is a comprehensive overview of how the AI revolution may change how we do nuclear physics in the time between now and when the EIC begins taking data. It can be broad in it’s approach, and not just focus on the current Genesis mission (although this is certainly a part of it). In addition, it would be helpful to consider the entire process from data taking, through calibration, analysis and data preservation. Of course, we look forward to you bringing your own approach to the topic as well.