



U.S. DEPARTMENT
of **ENERGY**

TREASURE-like Efforts

Gabriele D'Amen
Haider Abidi

October 14, 2026



@BrookhavenLab

NSF call

NSF AI Datasets Call (# 26-512)

Unlocking Dataset Value for AI-Enabled Scientific Discovery

- 1) apply **AI-based** capabilities to **feature extraction** and **metadata generation**, and the **integration** of multiple **datasets**.
- 2) Develop robust **data pipelines** necessary for **automated analysis** of **existing datasets** and similar use-cases by AI tools and systems.
- 3) **Augment** and/or **harmonize existing datasets** to better enable use by AI data pipelines and automated analysis.

Being led by Verena, Tulika and Peter Elmer AFAIK

Very similar in scope to TREASURE, we have offered help and tools if they feel it will strengthen their case

CERN AI project



FLARE (CERN + Mistral)

- *Foundation Models at LHC scale for the Reconstruction of Events: a detector- and task-agnostic foundation model for collider physics*
- *Scaling laws measured on jet models: the problem is data-bound, so $\sim 2/3$ of each added order of magnitude of compute must go to more data*
 - *from 10^{21} to 10^{24} – 10^{25} FLOPs $\rightarrow$ target of $O(10^{10})$ events, $O(10^{11})$ jets*

WP1 Dataset

two detectors, from reco objects
down to raw hits, full truth, open

WP2 Tokenisation

detector-agnostic representation;
tested on an unseen detector

WP3 Jet scaling

extend scaling laws in data size,
from reco inputs towards raw hits

WP4 Event model

event-level model, ~ 1000 particles
per collision

- **Dataset: public ATLAS and CMS software releases, Run-3 geometry, HL-LHC conditions (140 pile-up), full GEANT4**
 - $\sim 10^8$ CPU core-hours on **independent resources**, not the experiments' grid allocations

- This was presented @ ATLAS S&C week [Link](#) - (Also by spokes in ATLAS weekly - next slide)
- Have contacted Maurizio Pierini, who is the lead PI
- He is aware of the overlap, and is interested in talking at Genesis level on how to collaborate. We will bring this up Jeremy to discuss IP protections and future scope

CERN AI project

AI/ML strategy

- **CERN AI strategy implementation** (see [talk by M. Pierini](#) on 9 Oct 2026 at S&C Week)

- 8 working groups created within RCS
- Cross-experiment effort led by CERN staff
- CERN resources (3 years): 16 fellows + 16 doctoral students + GPU cluster

WG	Topic	Scope
1	Pattern recognition & object reconstruction	tagging, PID, calibration, tracking
2	Global event interpretation, particle flow & foundation models	end-to-end event reconstruction
3	Generation & fast simulation	calorimeter surrogates, flash sim
4	AI for online processing	on-sensor, FPGA, real-time inference
5	AI for system control & operations	DQ, monitoring, root-cause analysis
6	LLM assistants for scientific production	coding, analysis, plotting, writing
7	Software & hardware infrastructure	GPU platform, storage, model registry
8	AI in theory & phenomenology (TH AI node)	AI for TH, TH for AI

- CERN established a new collaboration with Mistral AI *proposing* to generate sample of $\sim 10^{10}$ events ($\sim 10^{11}$ jets) with Mistral resources for ATLAS & CMS and develop a detector- and task-agnostic foundation model for collider physics: a Large Physics Model ('FLARE')

- **Ongoing discussions with ATLAS**

- Also, discussion on outside collaborations & open data for AI/ML (discussion session on Thu at 2 pm)

- Our understanding is that this will be open to ATLAS/CMS collaborators to join?
 - Training detector agnostic FM with $\sim 10\times$ more data
- We will need to figure out areas of overlap and how TREASURE is uniquely pushing on new fronts in this landscape