

# RHIC Data and Analysis Preservation Round Table

11/20/25

Introduction

# Comments from the PAC

Recommendations of the Nuclear and Particle Physics Program Advisory Committee

Brookhaven National Laboratory

October 16 – 17, 2025

## 5. Data Preservation

RHIC's scientific mission would be incomplete without comprehensive data and analysis preservation (DAP). The PAC commends BNL Management for developing and piloting a comprehensive DAP plan. The plan integrates with DOE data policy, implements FAIR principles, and uses AI technology for accessing the data and analysis.

The PAC is pleased to hear that the development of the DAP plan is coordinated with the RHIC experiments via regular round-table discussions and that a prototype for the AI search is being worked on for STAR's documentation. We encourage BNL to apply for funding to support the components of the DAP plan, especially AI integration.

The analysis of RHIC data will continue over the next decade. The PAC recommends that BNL management plan to allocate adequate resources to support this work and to ensure that RHIC data and associated analysis are preserved for future research.

# Recommendations from the PAC

Recommendations of the Nuclear and Particle Physics Program Advisory Committee

Brookhaven National Laboratory

October 16 – 17, 2025

## **BNL Lab Management:**

- The PAC recommends that BNL management develop a plan to allocate adequate resources to support comprehensive data and analysis preservation (DAP) and to ensure that RHIC data and associated analyses are preserved for future research.
  - We encourage BNL to apply for funding to support the components of the DAP plan, especially AI integration.
- 
- Develop a plan to allocate resources to DAP.
  - Seek dedicated funding, especially for AI components.

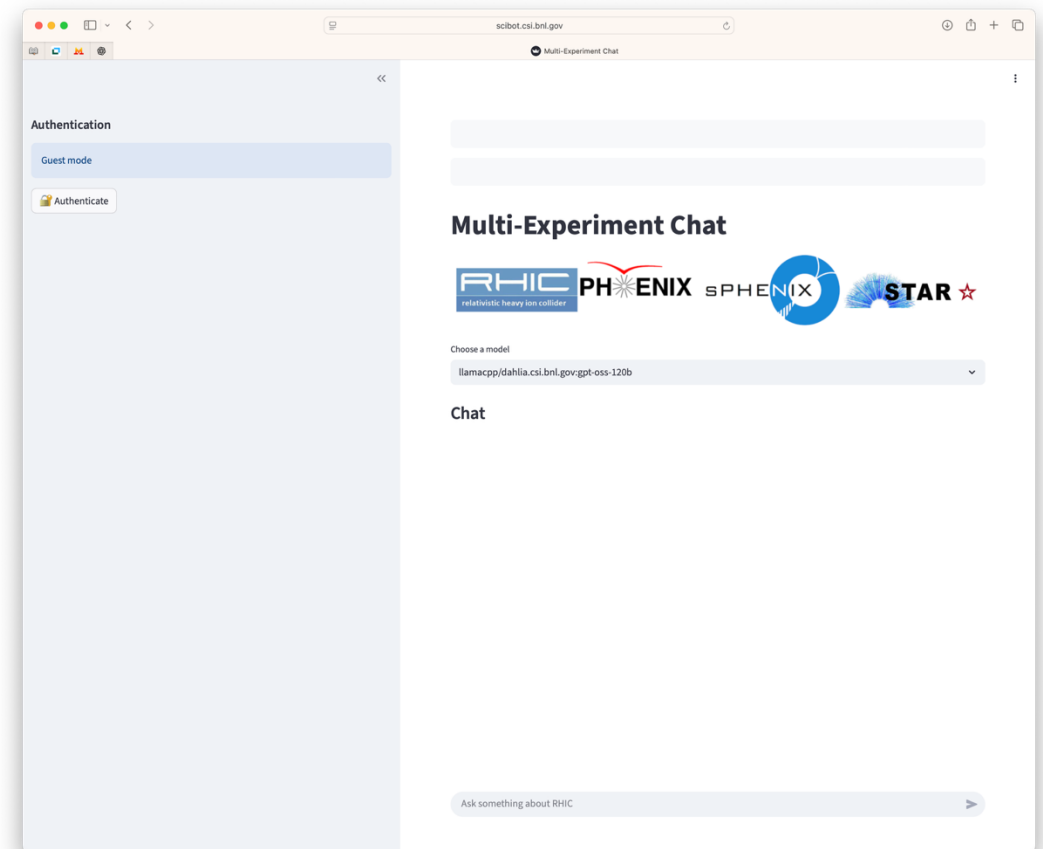
# RHIC university groups and DAP

- University collaborators are essential for the success of DAP
- Many DAP tasks require deep knowledge of RHIC detector systems, software, and past analyses. This expertise primarily exists within university groups.
- Dedicated funding is needed for universities to contribute effectively to DAP activities
- Abhay is engaged and supportive

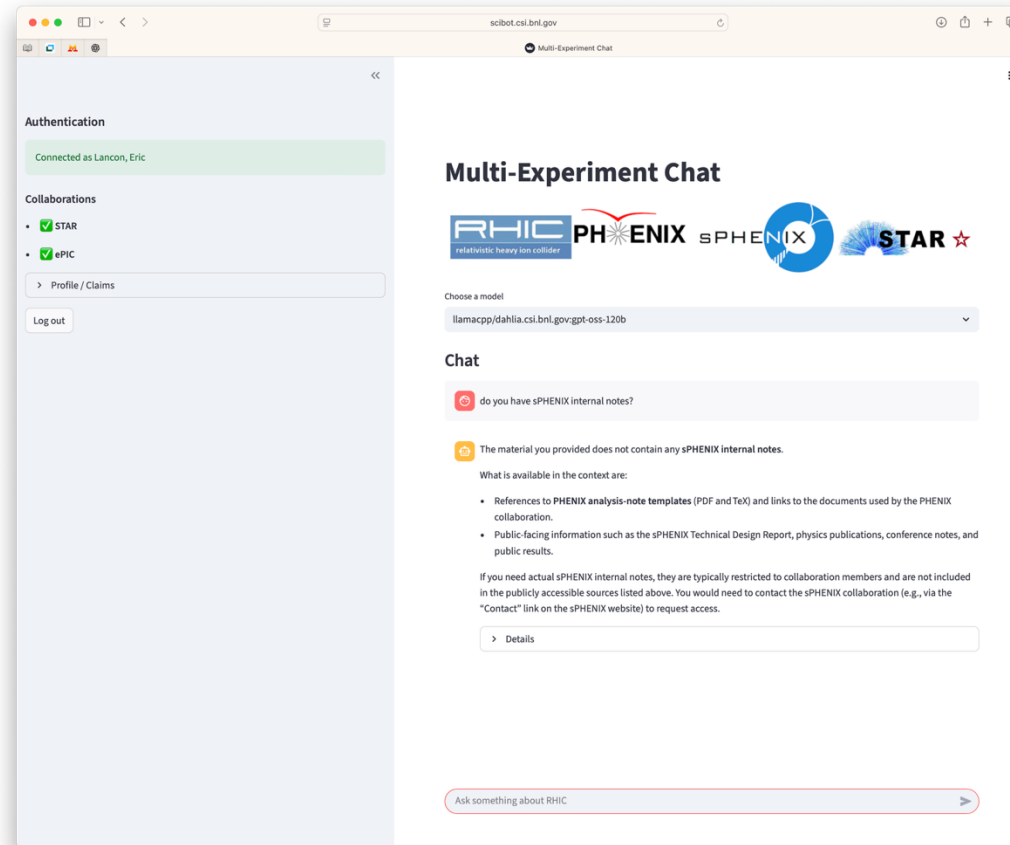
# SciBot is visible from outside BNL network

- Opened to general internet
- Without authentication: access to public information only (from BNL, PHENIX, sPHENIX, and STAR)
- Several thousand of records have been indexed
- With authentication, access only to STAR internal information
- Lacking GPU resources: answers can be slow

<https://scibot.csi.bnl.gov>



# Do you know sPHENIX internal notes?



Answer is no!

# Why not just use ChatGPT?

## The reason is data sovereignty

Collaborations, rely extensively on internal information that is not intended for public release.

Access to internal documentation like operational details, and collaboration-specific data must remain controlled and protected.

All internal data are kept on site and securely accessible via authenticated local AI systems.

# Different LLMs produce different levels of detail and accuracy

GPT-4o-mini

How do I analyze jets in 200 GeV Au+Au collisions from 2014?

Short how-to (based only on the provided STAR/RHIC context):

- Data & event selection
  - Use the 2014 STAR Au+Au  $\sqrt{s_{NN}} = 200$  GeV data set; select central events.
  - Apply event-quality cuts cited by the collaboration (not detailed).
- Track/tower inputs
  - Use charged tracks from the TPC with  $p_T \geq 0.2$  GeV/c (and  $|\eta_{track}| < 1.0$  in one description) and optionally include BEMC neutral towers [1,2].
  - Exclude events containing a track with  $p_T > 30$  GeV/c if following the quoted analysis [1].
- Jet finding
  - Run **FastJet** with the anti- $k_T$  algorithm. Typical resolution parameters used in the context:  $R = 0.2, 0.3, 0.4, 0.5$  (choose  $R$  per physics goal) [2,3].
  - Impose fiducial acceptance  $|\eta_{jet}| < 1 - R$  (or specific acceptance used in each study) [2].
- Background subtraction (heavy-ion specific)
  - Correct reconstructed jet  $p_T$  with the median-area method:  $p_{T,corr} = p_{T,rec} - \rho A$ , where  $\rho = \text{med}[p_{T,0}^2(\text{rec})/A_i]$  and  $A$  is the jet area (KT area method) [2].
  - Estimate residuals/systematics from mixed-event or event-plane dependent techniques as used in the cited analyses [2,3].
- Optional selections & studies
  - Semi-inclusive recoil jets: trigger on BEMC tower with 9–30 GeV ET and select jets in recoil region  $\Delta\phi \in (3\pi/4, 5\pi/4)$  [1].
  - Compare multiple  $R$  to study broadening and energy redistribution; check constituent  $p_T$  cuts (e.g., include/exclude soft constituents) [3].
- Notes/uncertainties
  - The context specifies reconstruction and background subtraction but does not give full unfolding, efficiency or systematic procedures—use STAR collaboration standards or referenced publications for those steps [1,2].

References: [1] STAR 2014 analysis description; [2] Jet reconstruction +  $p_{T,corr}$  (median  $\rho$ ) prescription; [3] Semi-inclusive/BEMC-triggered jet selection (context excerpts).

> Details

Q: How do I analyze jets in 200 GeV Au+Au collisions from 2014?

llama3

How do I analyze jets in 200 GeV Au+Au collisions from 2014?

To analyze jets in 200 GeV Au+Au collisions from 2014, you can follow the steps outlined below:

1. **Event selection:** Select events that have a high-energy trigger tower in the Barrel Electromagnetic Calorimeter (BEMC) and at least one charged track with  $p_T > 0.2$  GeV/c in the Time Projection Chamber (TPC). The trigger towers should be within the pseudorapidity range of  $|\eta| < 1.0$ .
2. **Jet reconstruction:** Use the anti- $k_T$  jet finding algorithm with a resolution parameter  $R = 0.4$  to reconstruct jets from the charged tracks in the TPC. The jets should have  $p_T > 2$  GeV/c and be within the pseudorapidity range of  $|\eta| < 1 - R$ .
3. **Background subtraction:** Subtract the underlying event background using a method based on active jet areas [1, 2]. This is important to ensure that the jets are not contaminated by non-jetlike particles.
4. **Jet shape analysis:** Study the differential jet shapes as a function of  $r$ , where  $r$  is the radial distance from the jet axis. You can use the mixed-event technique to estimate background contributions and subtract them from each  $p_T^{\text{assoc}}$  range [3].

By following these steps, you should be able to analyze jets in 200 GeV Au+Au collisions and compare them with those in p+p collisions. This will help you understand whether there are any significant modifications to the jets in Au+Au collisions.

References:

- [1] STAR Collaboration, Phys. Rev. C 92 (2015) 054906
- [2] STAR Collaboration, Phys. Rev. C 94 (2016) 024908
- [3] STAR Collaboration, Phys. Rev. C 95 (2017) 034903

> Details

Gemini-2

How do I analyze jets in 200 GeV Au+Au collisions from 2014?

To analyze jets in 200 GeV Au+Au collisions from 2014 using data collected by the STAR experiment, consider the following:

- **Data Source:** The data was collected by the STAR experiment at RHIC in 2014 [1].
- **Track Reconstruction:** Use charged tracks with  $0.2 < p_{T,track} < 30.0$  GeV/c and  $|\eta_{track}| < 1.0$  reconstructed by the Time Projection Chamber (TPC) [1]. Exclude events containing a track with  $p_{T,track}$  higher than 30.0 GeV/c [1].
- **Jet Reconstruction:**
  - Use the anti- $k_T$  jet clustering algorithm from the FastJet package [1].
  - Use a resolution parameter of  $R = 0.4$  [1].
- **Jet Selection:** Employ a semi-inclusive approach [1]. Consider jets in the recoil region ( $3\pi/4 < |\varphi_{jet} - \varphi_{trig}| < 5\pi/4$ ) of a high energy Barrel Electromagnetic Calorimeter (BEMC) trigger tower ( $9.0 < E_T < 30.0$  GeV) [1].

> Details

We must quantify and verify reliability.



# Validating SciBot

- Wrong information is worse than having no information.
- Once ‘validated’ information is fed into SciBot,
- There is still a need to develop a multi-layer validation process:
  - **Expert review:** domain experts vet answers and curate corrections.
  - **Automated checks:** semantic similarity, provenance scoring, and regression tests.
  - **Ensemble & benchmarking:** multi-model cross-checks + [domain Q&A benchmarks](#) (community-sourced).

[Need help to collect validated Q&As](#)

# Contribute with Q&A to validate SciBot

## Help Build SciBot — Preserve RHIC's Legacy

Every Q&A you submit keeps 20+ years of RHIC knowledge alive for future scientists


### What We Need From You:

- ⚡ Help benchmark and improve SciBot
- 🧠 Pose a challenging RHIC physics question that ChatGPT cannot answer
- 💬 Share your answer (brief or detailed—both help!)
- 🕒 Time: ~5-10 minutes per Q&A
- 📊 Goal: Build a comprehensive RHIC knowledge base

### Recognition & Impact:


- 📄 Contributors acknowledged in validation paper
- 🌟 Help shape AI-powered future of nuclear physics
- 📖 Make RHIC discoveries accessible to next generation
- 👥 Join colleagues across all RHIC experiments

### Get Started:



Or visit:

<https://shorturl.at/irxuk>

 **Submit Today!**

*The RHIC Data and Analysis Preservation Plan*

Short  
URL HERE  
is **correct**

# Call for Sample Analyses

- Need a small set of representative RHIC analyses to help implementing AI-driven (re-)analysis
  - Just ONE to start with
- Sample analyses to validate and test whether AI can:
  - interpret documentation,
  - reconstruct analysis logic,
  - identify missing information,
  - and help automate and verify reanalysis steps
- Needed to get a realistic estimate of required human intervention