

RHIC Data and Analysis Preservation Round Table

SciBot updates

- Reminder: high value of the SciBot for researchers, provides context grounded answers based in internal and public information
 - Controlled access to non-public documentation and internal materials
 - Public information can be accessed via standard public tools (e.g., ChatGPT, Claude, Gemini, Mistral, Copilot, Perplexity, etc.) without the need for SciBot.
- There are now experiment-tailored SciBot instances (+ RHIC generic and dev)
 - RAG's DataBase content agreed with relevant experiment
- Inference on publicly deployed LLMs:
 - Local hardware at CDS,
 - Azure,
 - Under test: connection to ANL inference service (access to large GPU clusters with open LLMs)
- Atif developing a framework to validate LLMs from calibrated answers

Re-Analysis

- **Reanalysis from an external perspective (STAR)**
 - Ankush Kanuganti (STAR) is investigating re-running a STAR analysis from a new/external-user perspective (benchmarking DAP).
 - Evaluating implicit assumptions, documentation completeness, and reproducibility.
 - More details at an upcoming meeting.
- **AI-assisted datasheet documentation**
 - Goal: automated generation of standardized datasheets for low-level datasets (e.g., ROOT files) used in analyses.
 - Centralizes essential information, including implicit knowledge, required to understand and reuse datasets.
 - Helps identify missing elements required for reanalysis and verification.
 - Enhances reproducibility, AI-readiness and long-term usability of RHIC data

Today

Apologies from Christine & Frank

1. PHENIX data preservation updates – Christine / Gabor
2. Updates on SciBot – Atif / Vincent

- Next meeting: TBD