



ID de la contribución : 73

Tipo : Talk

Towards Agent-Driven HEP Analysis: Physics Closure and Harness Development for a RAG-Based AI Co-scientist

We present an update on our prototype Retrieval-Augmented Generation (RAG) and agentic LLM tool designed to accelerate and support high-energy physics analyses. As a case study, we applied the system to the published 2016 $\Lambda_b \rightarrow \Lambda\gamma$ Run 2 LHCb analysis. Reproducing legacy workflows is often slow and error-prone due to fragmented code, dispersed documentation, personnel turnover, and software evolution over multiple data-taking periods. Our tool integrates semantic retrieval, code-understanding capabilities, and autonomous agent actions to navigate these challenges.

The system ingests heterogeneous knowledge sources, including archived analysis scripts, notes, published documentation, and the modern LHCb software stack, and uses a vector database to build a unified retrieval layer. Built on general-purpose LLMs, open-source and closed-source IDEs, and agentic frameworks, the assistant can propose analysis strategies, generate or repair scripts, and autonomously perform several reconstruction and analysis tasks. With the current prototype we reach near physics closure with the published analysis. In internal tests, the prototype demonstrated the potential to compress workflows that normally require months or years of manual effort.

We have also used the framework to revive a stale LHCb analysis, whose last analysis note dates from 2023, and to drive it to near closure in a matter of months, pending physics approval.

We are now extending the tool with an agent harness that records the analysis state, so the agent can resume from a checkpoint, compare runs, and roll back after a failed step instead of restarting from scratch. On top of this harness we are implementing an autoresearch-style loop in which the agent proposes, tests, and evaluates candidate selection cuts against a defined figure of merit, starting with the optimisation of the $\Lambda_b \rightarrow \Lambda\gamma$ selection and extending to further physics tasks such as reconstruction.

We discuss the tool's architecture, agentic capabilities, and safeguards, as well as its current limitations. Verification is ongoing to evaluate reliability, detect errors, and benchmark performance against non-RAG LLM baselines. These case studies highlight the promise of RAG-augmented, agent-driven systems in restoring, verifying, and accelerating complex HEP analyses, pointing toward future end-to-end AI-assisted analysis pipelines.

Primary author(s) : PEREZ, Cilicia Uzziel (La Salle Barcelona, LHCb); VILASIS-CARDONA, Xavier (La Salle - Universitat Ramon Llull); CALVO GÓMEZ, Míriam (La Salle, Universitat Ramon Llull); OYANGUREN, Arantza (IFIC- Valencia); MARTORELL I GRANOLLERS, Alexia (La Salle, Ramon Llull); ABIA SARRIA, Carlos

Presenter(s) : PEREZ, Cilicia Uzziel (La Salle Barcelona, LHCb)

Clasificación de la sesión : COMCHA

Clasificación de temáticas : COMCHA