

Towards Control Room Automation with a Chatbot for DAQ and Shifter Assistance

Luigi Podda,

with G. Avolio, S. Chapeland, J. Hoya, A. Marzin.

1 July 2025

AI in DAQ systems (EP R&D WP9.1)

The EP R&D is a cross-experiment collaboration that aims to explore cutting-edge technologies for use at CERN.

In particular, WP9.1 works on exploring new and AI applications to automate the DAQ systems for all experiments.

Short Term

Generative AI

A Chatbot to assist shifters and operators.

Create reports and summaries of interesting events at the end of a run.

Medium Term

Anomaly detection

Detect and anticipate imminent failures in complex scenarios.

Further automate and improve the monitoring of a HEP experiment.

Long Term

Experiment Autopilot

Break down plans of the day into executable actions.

Initiate the run once all systems are properly configured and ready.

Generative AI (Chatbot)

Data & Running Infrastructure

Documentation

- Sourced from **ALICE DAQ** and **ATLAS Trigger** systems
- **~400 Markdown documents**
- Rich in technical procedures, configurations, and internal workflows

Infrastructure

GPU Resources

- Initially: **NVIDIA RTX 5000 Ada Generation (32GB VRAM)**
- Recently upgraded to: **NVIDIA RTX 6000 Pro Blackwell (96GB VRAM)**

Storage & Processing

- **Framework:** LangChain (Orchestrates document loading, chunking, embedding, retrieval, and generation)
- **Vector Store:** ChromaDB
 - Stores embedded chunks with metadata (e.g., doc name, section headers)
 - Enables fast semantic search and context retrieval

The LLM-as-Judge Paradigm

What is LLM-as-Judge?

- A Paradigm where a **large language model evaluates the output** of an LLM system.
- The judge model assesses **accuracy, relevance, coherence, and contextual quality**.
- Acts like a human evaluator.

Why use it?

- **Automated, scalable evaluation** without relying on manual annotation.
- Captures **semantic understanding** and not just words matches.
- Can explain **why** something is right or wrong.

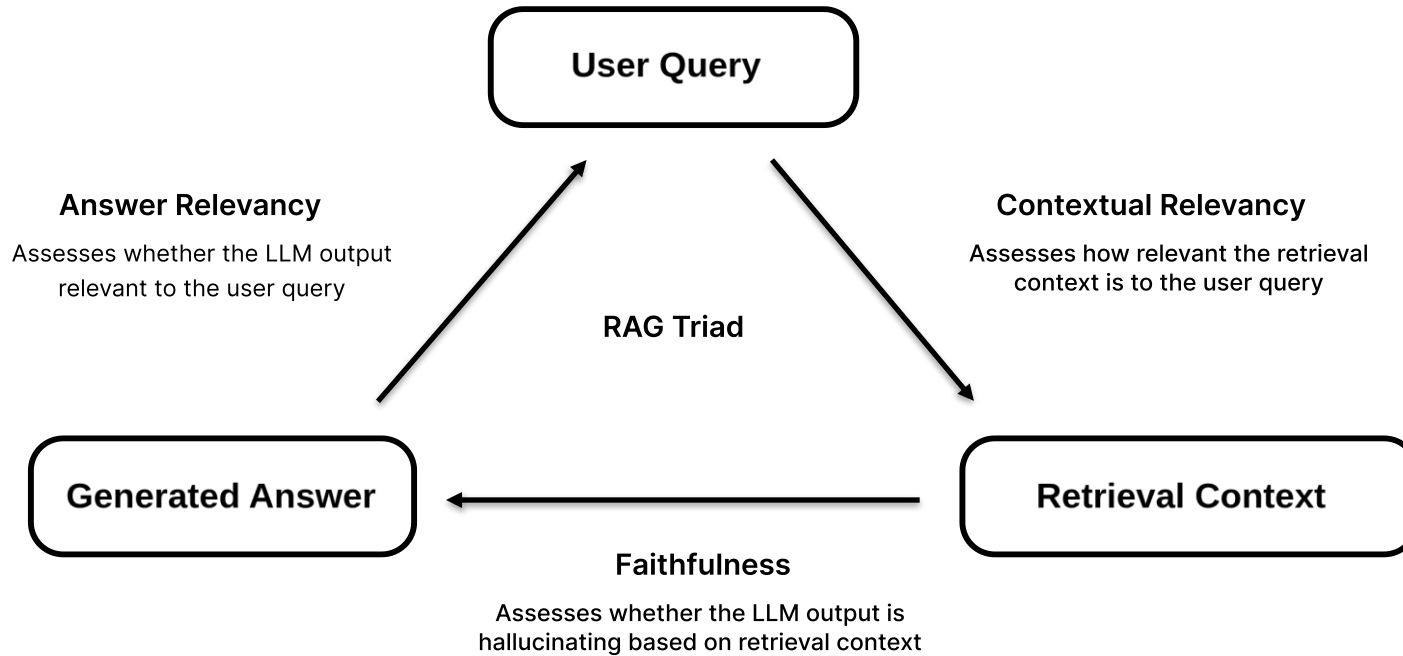
Computes metrics like:

- Answer Relevancy
- Faithfulness
- Contextual Relevancy

Returns a **score + explanation** for transparency

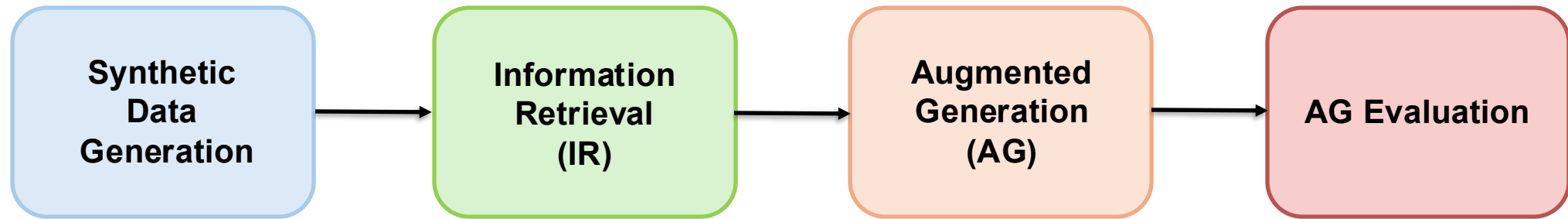


Evaluation Metrics



- **Answer Relevancy** can be improved by enhancing examples and instructions in prompts.
- **Faithfulness** can be improved by changing or fine-tuning the LLM if it can't use retrieved context effectively.
- **Contextual Relevancy** can be improved by using better embeddings and chunking strategies

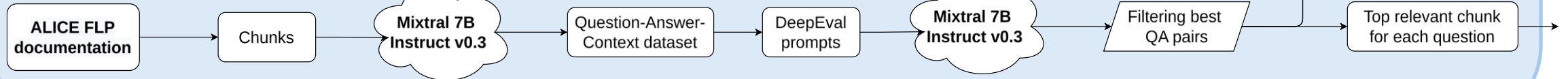
RAG pipeline



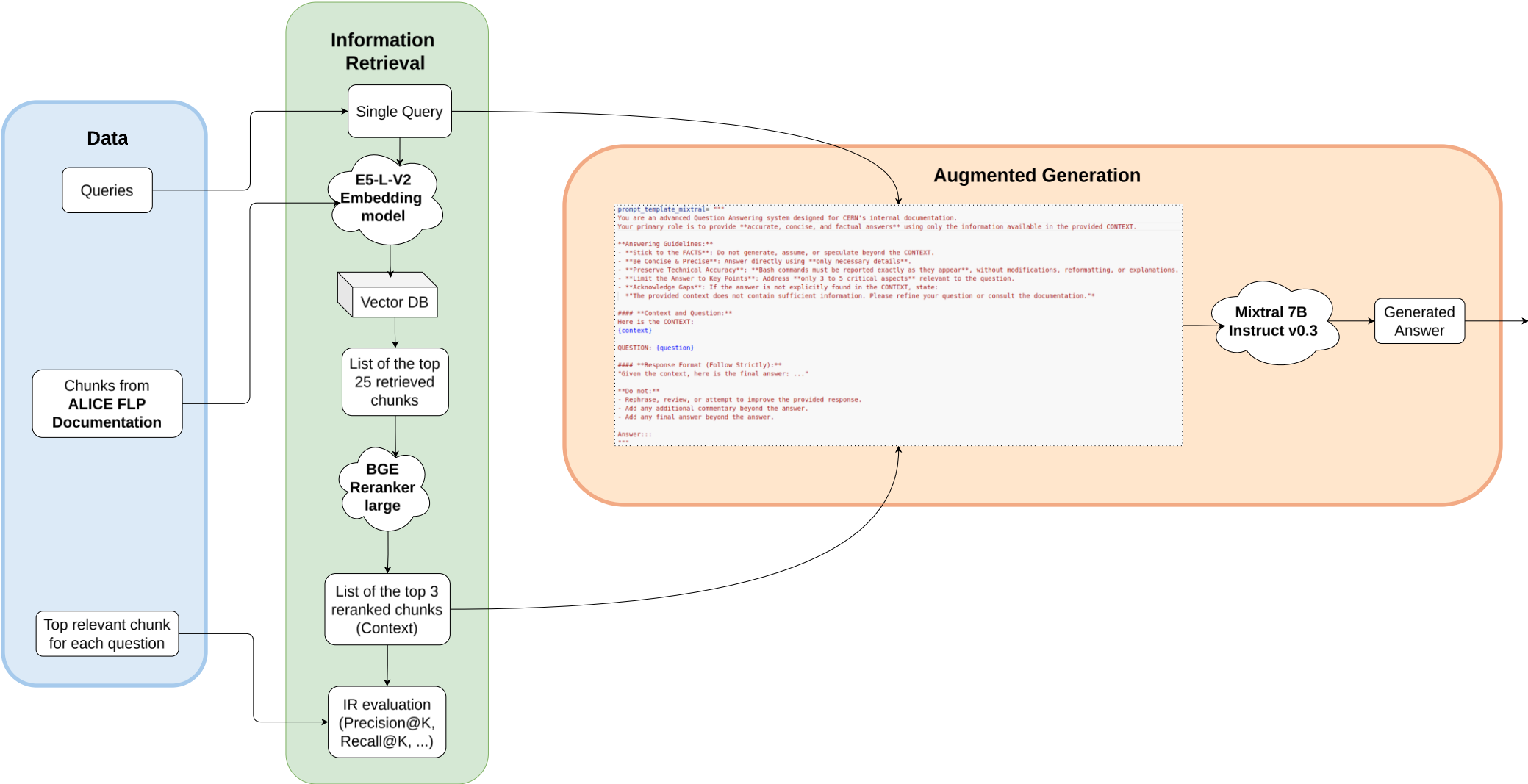
RAG pipeline (Data Generation)

Synthetic data generation

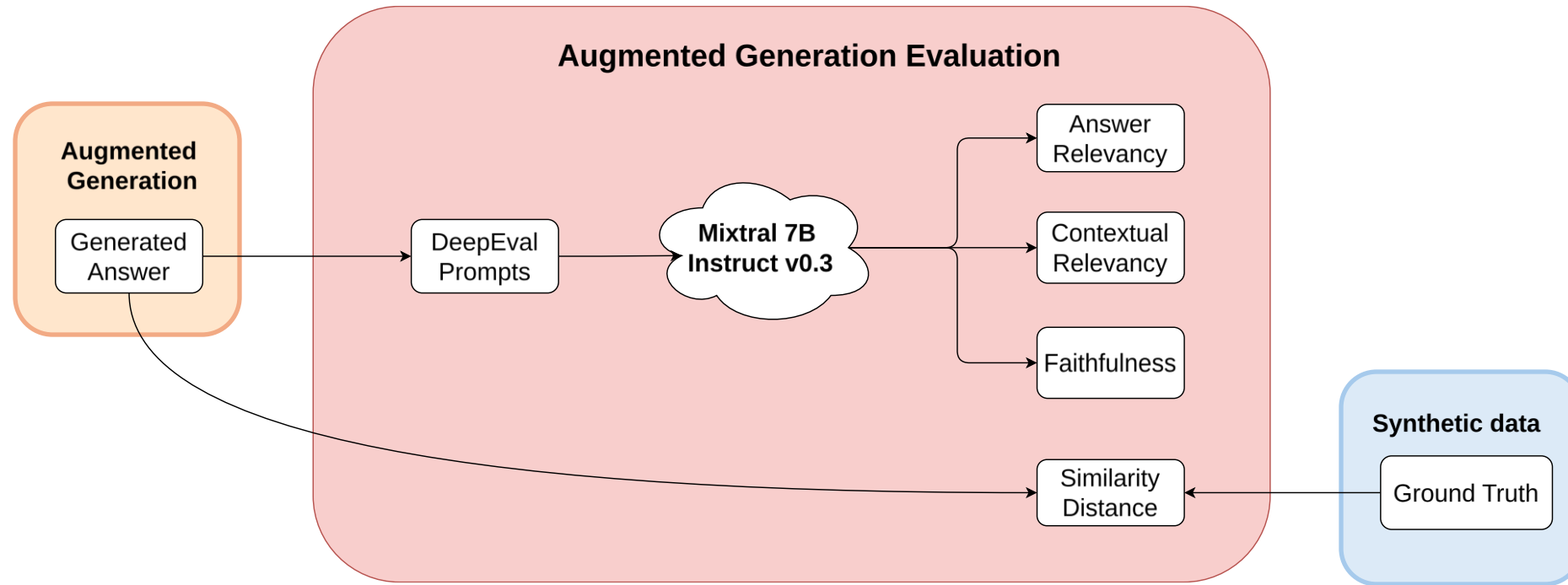
```
QA_generation_prompt = """  
Your task is to write a factoid question and an answer given a context.  
Your factoid question should be answerable with a specific, concise piece of factual information from the context.  
Your factoid question should be formulated in the same style as questions users could ask in a search engine.  
This means that your factoid question MUST NOT mention something like "according to the passage" or "context".  
  
Provide your answer as follows:  
  
Output::  
Factoid question: (your factoid question)  
Answer: (your answer to the factoid question)  
  
Now here is the context.  
  
Context: {context}\n  
Output::"""
```



RAG pipeline (Data Generation)



RAG pipeline (Data Generation)



Example

Query:

Which packages need to be uninstalled prior to upgrading FLP Suite from v0.10.0 to v0.10.1?

Ground Truth:

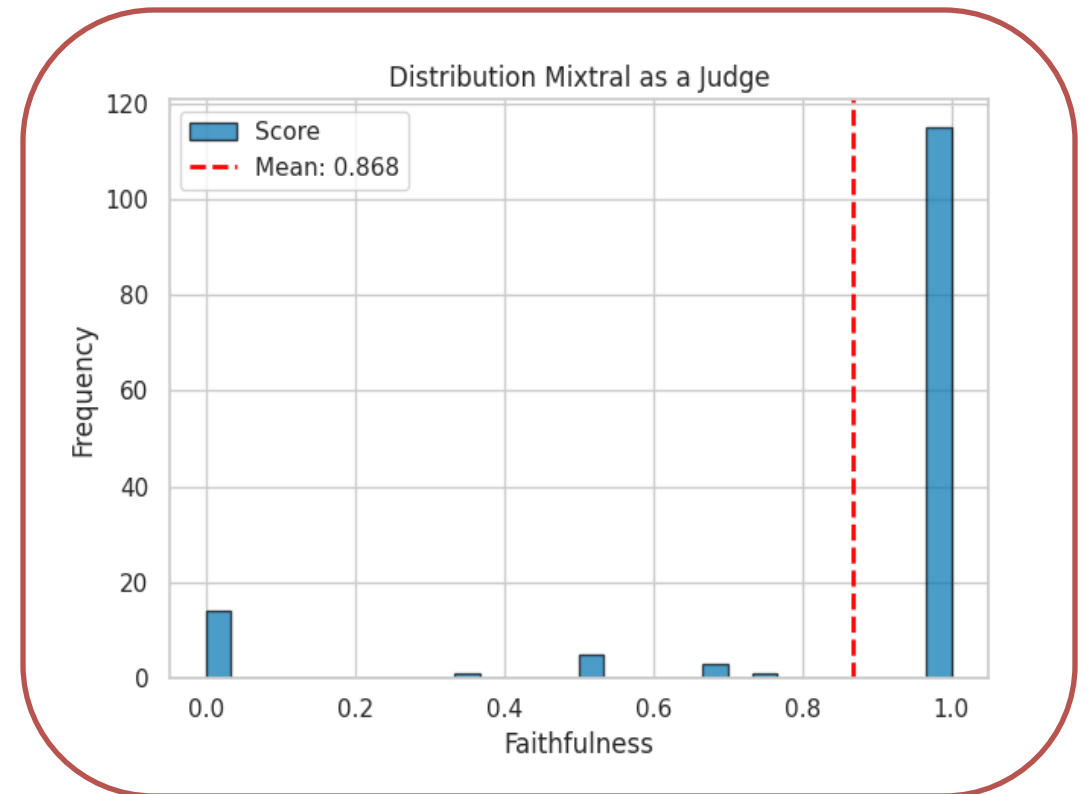
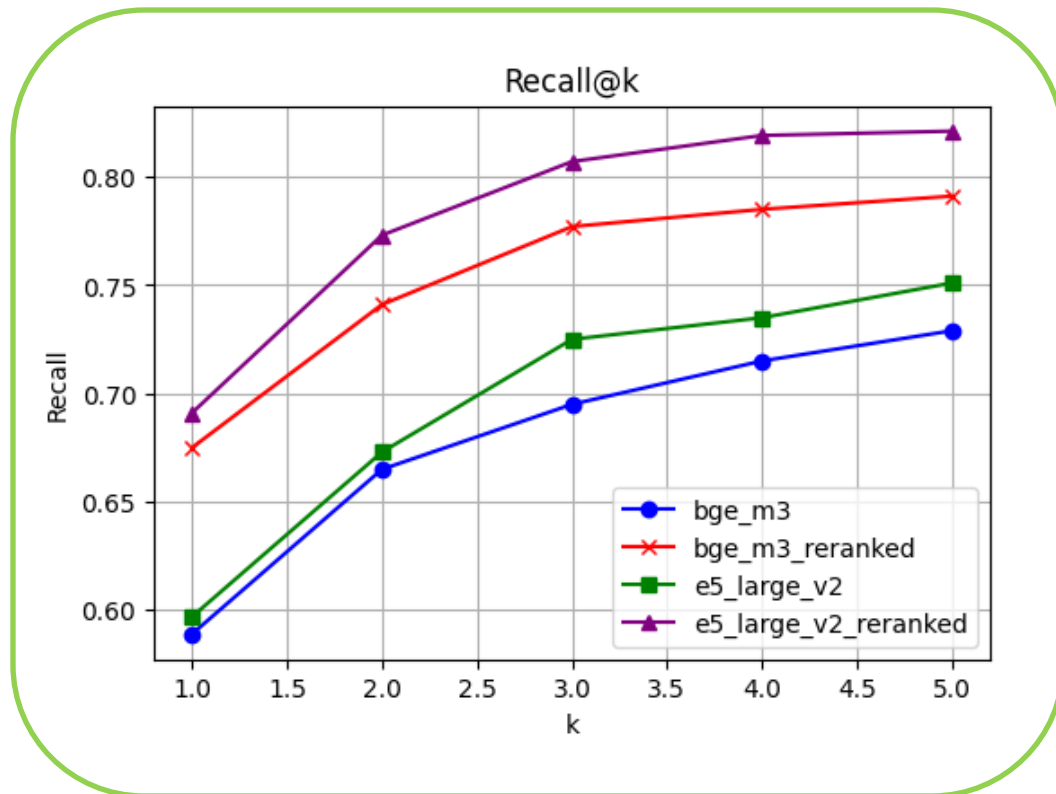
The packages `alisw-coconut`, `alisw-Configuration`, `alisw-Control` and related, `alisw-Monitoring`, `alisw-PDA` and `alisw-ReadoutCard`, and `alisw-QualityControl` must be removed before installing the latest version of FLP Suite due to compatibility issues.

RAG Answer:

The following packages need to be uninstalled prior to upgrading FLP Suite from v0.10.0 to v0.10.1:

- alisw-coconut
- alisw-Configuration
- alisw-Control and related
- alisw-Monitoring
- alisw-PDA and alisw-ReadoutCard
- alisw-QualityControl
- alisw-Readout.

RAG pipeline (Evaluation)



Open-WebUI

Integration of Open-WebUI of testing purposes.



Conclusions

Experiments ask for increasing automation in the **Control Room** and **DAQ systems** to reduce manual effort and improve efficiency.

Developing a system that assists **shifters** with quick access to **DAQ documentation**.

A next step could be the integration of **summaries of interesting events or issues** during experiment runs (per run, day, or week)

How it can be done?

LLM agents can break down complex tasks into smaller, manageable steps

Final Goal -> Experiment Autopilot that:

- Translates daily plans into executable steps
- Controls detector components dynamically
- Coordinates run start-up once all systems are ready



Thank you for your attention!

