Contribution ID: **93**   Type: **Poster**

Retrieval Augmented Generation for Particle Physics: A Case Study with the Snowmass White Papers and Reports

Wednesday 13 March 2024 16:15 (30 minutes)

Particle physics faces many challenges and opportunities in the coming decades, as reflected by the Snowmass Community Planning Process, which produced about 650 reports on various topics. These reports are a valuable source of information, but they are also difficult to access and query. In this work, we explore the use of Large Language Models (LLMs) and Retrieval Augmented Generation (RAG) to answer questions based on the Snowmass corpus. RAG is a technique that combines LLMs with document retrieval, allowing the model to select relevant passages from the corpus and generate answers. We describe how we indexed the Snowmass reports for RAG, how we compared different LLMs for this task, and how we evaluated the quality and usefulness of the answers. We discuss the potential applications and limitations of this approach for particle physics and beyond.

Significance

LLM's are new - and we are figuring out how to apply them in our field in ways that leverage their power. Search and reasoning over local document collections is one such approach.

This is new work, and hasn't been presented before.

References

Experiment context, if any

None, though both of us are doing this with IRIS-HEP in mind, which isn't actually an experiment...

Primary authors: GALEWSKY, Benjamin (Univ. Illinois at Urbana Champaign (US)); WATTS, Gordon (University of Washington (US))

Presenter: WATTS, Gordon (University of Washington (US))

Session Classification: Poster session with coffee break

Track Classification: Track 1: Computing Technology for Physics Research