Contribution ID: **1767** Contribution code: **THPM120**   Type: **Poster Presentation**

Application of large language models for the extraction of information from particle accelerator technical documentation

Thursday 5 June 2025 15:30 (2 hours)

The large set of technical documentation of legacy accelerator systems, coupled with the retirement of experienced personnel, underscores the urgent need for efficient methods to preserve and transfer specialized knowledge. This paper explores the application of large language models (LLMs), to automate and enhance the extraction of information from particle accelerator technical documents. By exploiting LLMs, we aim to address the challenges of knowledge retention, enabling the retrieval of domain expertise embedded in legacy documentation.

We present initial results of adapting LLMs to this specialized domain. Our evaluation demonstrates the effectiveness of LLMs in extracting, summarizing, and organizing knowledge, significantly reducing the risk of losing valuable insights as personnel retire. Furthermore, we discuss the limitations of current LLMs, such as interpretability and handling of rare domain-specific terms, and propose strategies for improvement. This work highlights the potential of LLMs to play a pivotal role in preserving institutional knowledge and ensuring continuity in highly specialized fields.

Footnotes

Paper preparation format

LaTeX

Region represented

Europe

Funding Agency

Author: DAI, qing (Paul Scherrer Institute)

Co-authors: GRYCNER, Adam (Google (Switzerland)); Dr SAPINSKI, Mariusz (Paul Scherrer Institute); ISCHEBECK, Rasmus (Paul Scherrer Institute)

Presenter: Dr SAPINSKI, Mariusz (Paul Scherrer Institute)

Session Classification: Thursday Poster Session