

Towards Unlocking Insights from Logbooks Using AI

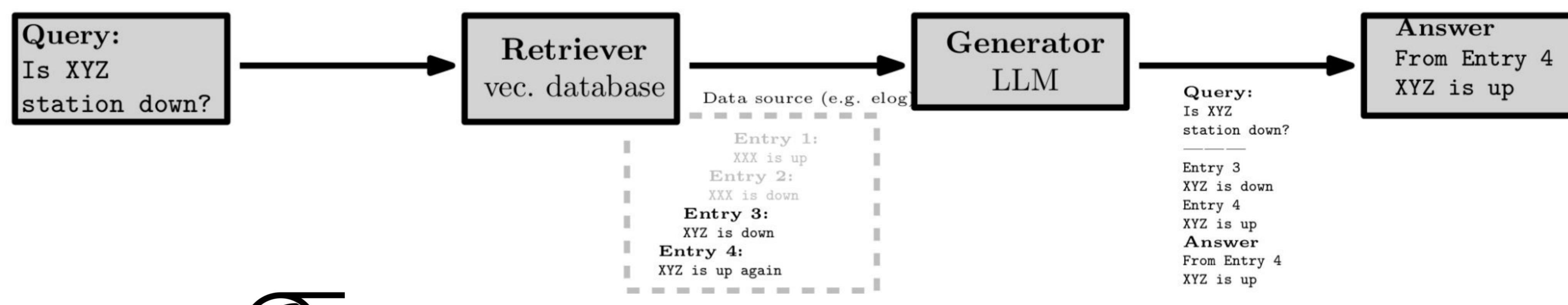
A. Sulc (HZB), Gregor Hartmann (HZN), J. Maldonado (BNL), V. Kain (CERN), F. Rehm (CERN), A. Eichler (DESY), J. Kaiser (DESY), T. Wilksen (DESY), F. Mayet (DESY), R. Kammering (DESY), H. Tuennemann (DESY), J. St. John (Fermi), H. Hoschouer (Fermi), K. J. Hazelwood (Fermi), T. Hellert (LBNL), D. Ratner (SLAC)

Electronic logbooks contain valuable information about activities and events concerning their associated particle accelerator facilities. However, the highly technical nature of logbook entries can hinder their usability and automation. As natural language processing (NLP) continues advancing, it offers opportunities to address various challenges that logbooks present. This work explores jointly testing a tailored Retrieval Augmented Generation (RAG) model for enhancing the usability of particle accelerator logbooks at institutes like DESY, BESSY, Fermilab, BNL, SLAC, LBNL and CERN. The RAG model uses a corpus built on logbook contributions and aims to unlock insights from these logbooks by leveraging retrieval over facility datasets, including discussion about potential multimodal sources. Our goals are to increase the FAIR-ness (findability, accessibility, interoperability, and reusability) of logbooks by exploiting their information content to streamline everyday use, to enable macro-analysis for root cause analysis, and to facilitate problem-solving automation.

FERMILAB-POSTER-24-0055-AD

MOTIVATION

- Using data from highly technical particle accelerator electronic logbooks (**eLogs**) using natural language processing (NLP) and retrieval augmented generation (RAG) models.
- Enhancing the usability, accessibility, interoperability, and reusability (**FAIR-ness**) of eLogs by exploiting their information content.
- Jointly developing a tailored RAG model across multiple particle accelerator facilities (DESY, BESSY, Fermilab, BNL, SLAC, LBNL, and CERN) and **summarizing their progress** to address challenges posed by the technical nature of eLog entries.
- Discussion about use of eLog and operator human-made data and **multimodal sources** (control system data, meeting transcripts, chat logs, etc.) to augment eLog entries



ALS

At the ALS, we have enhanced semantic logbook searches to handle acronyms and jargon better than our initial attempts with word2vec.

We are refining an embedding model and integrating the all-MiniLM-L6-v2 model into a RAG pipeline with Mistral-7B-Instruct-v0.2.

We have started recording our weekly *Operations Critic Meetings* to include in the RAG pipeline, but face challenges with transcribing due to poor audio quality and the prevalent use of acronyms/colloquial language in these meetings.

BNL

- We used Gensim and Doc2Vec to enhance eLog search personalization for C-AD, matching user interests.
- Transformed entries to word vectors, found similarities in accelerator terms. Explored topic modeling for tagging.
- Developed web demo letting users test with word/phrase queries [1].
- Future plans: reinforcement learning from user feedback, eLog system integration.

References

- [1] J. Maldonado, S. Clark, W. Fu, and S. Nemeshure, "Enhancing Electronic Logbooks Using Machine Learning," in Proc. 19th Int. Conf. Accel. Large Exp. Phys. Control Syst.
- [2] F. Rehm, J. M. Guijarro, N. Soufflet, and V. Kain, AccGPT: A Vision for AI Assistance at CERN's Accelerator Control and Beyond, 2024.
- [3] F. Mayet, Building an intelligent accelerator operations assistant using advanced prompt engineering techniques and a high level control system toolkit, 2024.
- [4] F. Mayet, "Gaia: A general ai assistant for intelligent accelerator operations," arXiv preprint arXiv:2405.01359, 2024.
- [5] J. Kaiser, A. Eichler, and A. Lauscher, "Large language models for human-machine collaborative particle accelerator tuning through natural language," unpublished, 2024.
- [6] K. J. Hazelwood, D. Finstrom, M. McCusker-Whiting, and L. G. Mills, "The fermilab accelerator division electronic logbook (ADEL) at 10 years," 2023. 10.2172/2246727
- [7] T. Njeku, "(noise) neural optical image categorizer for the e-log," 2020. <https://www.osti.gov/biblio/1706139>

Acknowledgement

We acknowledge DESY (Hamburg, Germany) and HZB (Berlin, Germany), a member of the Helmholtz Association HGF, for their support in providing resources and infrastructure. This work has in part been funded by the IVF project InternLabs-0011 (HIR3X).

This manuscript has been authored in part by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.

CERN

- Building a pilot called AccGPT [2] to host LLMs with RAG on CERN internal knowledge and fine-tuning them in the future.
- Not incorporating logbooks initially due to data quality concerns.
- Plan for LLMs to fill next-gen logbooks and provide accelerator statistics.
- AccGPT will undergo accuracy tests with experts in mid-2024.

DESY & BESSY (HZB)

- Developing new eLog system simultaneously
- Successful RAG implementation at SINBAD-ARES [3,4], combining multiple data sources
- Using FAISS vectorstore and Mistral-7B for embedding/generation at XFEL/FLASH
- Storing control system metadata with screenshots
- Exploring topic modeling, root cause analysis, autonomous tuning [5].

| Charge | SA1 | SA2 | SA3 | Value |
|-------------------|----------|----------|----------|---|
| | 0.25 nC | 0.25 nC | 0.25 nC | .../TORA.25.II/CHARGE.SA3 0.25 nC |
| Bunch Rate | 2257 kHz | 2257 kHz | 2257 kHz | .../TORA.3098.74D/NUMBEROFBUNCHES.SA3 182 |
| Requested Bunches | 92 | 20 | 182 | .../TORA.25.II/CHARGE.SA2 0.25 nC |
| | | | | .../TORA.3181.TSD/NUMBEROFBUNCHES.SA2 20 |
| | | | | .../TORA.25.II/CHARGE.SA1 0.25 nC |
| | | | | .../TORA.3098.74D/NUMBEROFBUNCHES.SA1 92 |

Fermilab

- ADEL [6] is the primary accelerator operations logbook
- Previous work on OCR for screenshots and image categorization [7].
- Current semantic search prototype providing relevant results
- Using all-mpnet-base-v2 for embedding, Qdrant vector store
- Implementing re-ranking and recommendation features.
- Plan to fully integrate search function into ADEL soon



SLAC

- Investigating text generation (reports, keywords) and natural language queries
- Using operator manuals, technical docs as context
- Initial work on RAG pipeline augmenting logs with internal wiki
- "Cleaning" process to update wiki and generate Q&A pairs
- Working on data processing workflow and IRB review

NATIONAL
ACCELERATOR
LABORATORY

CONCLUSION

- Our work showed initial progress with RAG models for unlocking eLog insights across institutes.
- Technical language in eLogs challenged document retrieval for our RAG models.
- Our findings highlighted AI's potential to enhance eLog usability and accessibility.
- Further research is crucial for us to fully leverage NLP and metadata for augmenting eLogs.
- Continued efforts needed to address technical language barrier and improve retrieval performance.

Can you make a table from all quenches that took place in September 2023?

Based on the provided context, the following quenches occurred in September 2023: A5 quench, A5 quench.

Here's a table summarizing the information:

| Date | Quench |
|------------|--------|
| 2023-09-07 | A5 |