The U.S. Department of Energy publishes Lessons Learned documents (LLs) that record safety incidents from national labs. Argonne and FermiLab seek to learn from these safety incidents by matching the LLs to Work Control Documents (WCDs) that describe their projects, but the large and ever-updating volume of LLs makes this process challenging. Argonne has built a machine-learning pipeline for this process using a large language model (LLM) to generate summaries for LLs and an embedded similarity analysis algorithm to match LL summaries and WCDs. The spring clinic group improved the efficiency of this pipeline and adapted the algorithm for use at FermiLab.

The group achieved the following: firstly, a new WCD preprocessor was implemented, resulting in a 37% increase in efficiency and improved processed text quality. Secondly, better LLMs and prompts were identified to generate more meaningful summaries for LLs. Finally, the existing code was tailored to Fermilab data, given its different format. The group tested using LLMs to augment the sparse Fermilab WCD texts, but the number of matches consistently remained low. The next steps would be to refactor the existing code and further test LLM prompts to ensure the pipeline works with WCDs from diverse sources.

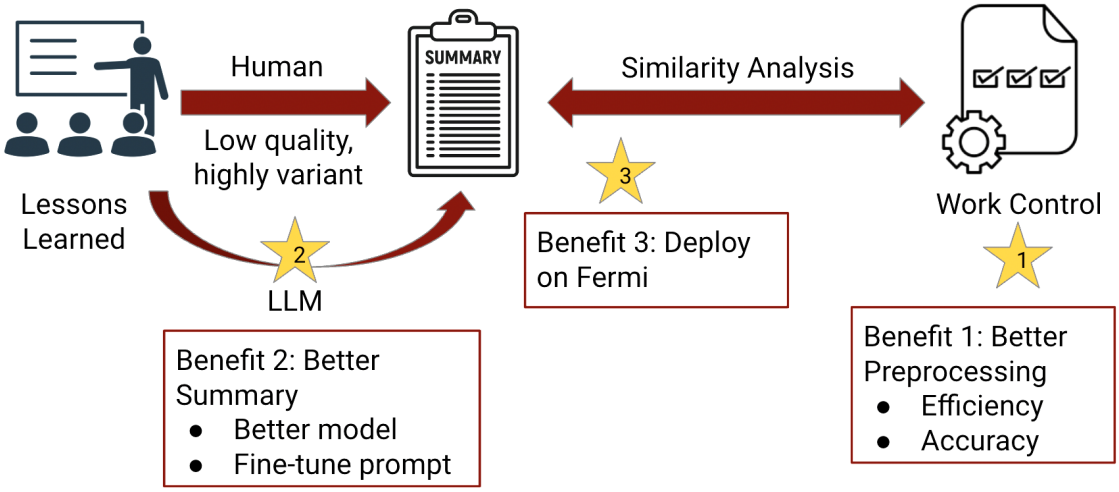

Diagram of Matching Methodology and New Benefits Provided