Benchmarking Large Language Model Workflows for Efficient HAZOP Study Generation

Jingkang Liang*, Gürkan Sin*

** Process and Systems Engineering Center (PROSYS), Department of Chemical and Biochemical Engineering, Technical University of Denmark (DTU), 2800 Kgs. Lyngby, Denmark *(jinlia@kt.dtu.dk)*

Abstract: Hazard and Operability (HAZOP) studies are widely used in the process industries for systematic identification of process hazards and operational deviations (Crawley, 2015). However, HAZOP studies remain highly labor-intensive, requiring extensive expert discussion and manual documentation. Recent advances in Large Language Models (LLMs) have created new opportunities for supporting engineering workflows in process systems engineering and process safety analysis. Recent studies have also investigated the feasibility of applying LLMs to automated HAZOP generation workflows (Lee et al., 2026).

This work investigates the use of LLM-based workflows for automated generation of structured HAZOP worksheets with the goal of improving study efficiency while maintaining human expert supervision. A multi-step LLM workflow is developed to generate HAZOP table entries including deviations, causes, consequences, safeguards, and recommendations. The study focuses on benchmarking different LLM models and workflow structures to investigate their influence on the consistency, completeness, and practical usefulness of generated HAZOP studies. Rather than positioning the LLM as a replacement for safety engineers, the proposed framework aims to support experts by accelerating the preparation of initial HAZOP drafts and shifting engineering effort toward critical review and validation.

The current workflow is capable of generating approximately 200 HAZOP worksheet rows within 20 minutes for a process case study. Due to the limited availability of industrial ground-truth HAZOP datasets, the present work emphasizes comparative evaluation of workflow strategies and qualitative expert assessment instead of direct industrial validation. The work further discusses challenges including hallucination, reproducibility, and reliability of LLM-generated safety analysis outputs.

The proposed framework represents an initial step toward AI-assisted safety reasoning systems for process industries and highlights the potential of combining LLM workflows with future process simulation, abnormal situation analysis, and operator decision-support systems, building upon recent developments in LLM agents for chemical process simulations (Liang et al., 2026).

Keywords: Process Safety, Large Language Model, Hazard and Operability Study

REFERENCES

Crawley, Frank, and Brian Tyler. *HAZOP: Guide to best practice*. Elsevier, 2015.

Lee, Junseo, et al. "Can large language models automate the HAZOP process without human intervention?." *Safety Science* 194 (2026): 107039.

Liang, Jingkang, Niklas Groll, and Gürkan Sin. “Large Language Model Agent for User-Friendly Chemical Process Simulations.” *Digital Chemical Engineering*, vol. 19, 2026, p. 100312, https://doi.org/10.1016/j.dche.2026.100312