Integrating Graph Databases and AI for Advanced HAZOP Analysis in Ammonia-Hydrogen Systems

Mokkawes Thirakorn[a], **Iskandar Halim**[a], **Shaik Salim**[a]

[a] *Institute of Sustainability for Chemicals, Energy and Environment (ISCE2), Agency for Science, Technology and Research (A*STAR), 1 Pesek Road, Jurong Island, Singapore* mokkawest@isce2.a-star.edu.sg

* Presenting author

1. Introduction

Ammonia-Hydrogen systems are central to sustainable energy solutions, yet their complexity poses significant challenges for process safety and operability. Traditional Hazard and Operability (HAZOP) analysis methods often rely heavily on manual expertise and intuition, leading to potential inefficiencies and missed risks in intricate systems. The increasing need for more robust, efficient, and automated safety assessment frameworks drives the innovation of advanced methodologies [1, 5].

This paper presents a cutting-edge framework that integrates graph database technologies, advanced traversal algorithms, and natural language processing to deliver a comprehensive solution for HAZOP analysis. By leveraging graph-based system modelling, autonomous scenario generation, and intuitive user interfaces, the proposed framework ensures thorough hazard identification, effective risk mitigation, and improved decision-making processes. This approach not only addresses critical gaps in traditional safety methodologies but also introduces a scalable, future-proof standard for analyzing complex industrial systems [4].

2. Methodology

The framework integrates graph analysis, AI reasoning, and natural language interaction for efficient and scalable HAZOP, detailed in the following sections:

2.1 Data Transformation and Modelling

The foundational step in the framework involves converting engineering diagrams, specifically Process Flow Diagrams (PFDs) and Piping & Instrumentation Diagrams (P&IDs), into structured data. These diagrams, often stored as XML files, are processed to extract detailed information regarding equipment configurations, interconnections, and operational parameters. This data is encoded into a Neo4j graph database. Each node within the graph represents a specific system component—such as pumps, valves, heat exchangers, or pipelines—while edges define their interconnections and dependencies [5]. Metadata, such as chemical properties, operational thresholds, and safety parameters, enriches the graph to form a dynamic digital twin of the system, enabling real-time interaction and analysis.

2.2 Graph Traversal and Hazard Analysis

Graph traversal algorithms serve as the analytical foundation for identifying and understanding hazards within the system. Breadth-First Search (BFS) facilitates the systematic exploration of downstream impacts caused by deviations, such as pressure surges propagating through pipelines and valves [1]. Depth-First Search (DFS) provides a deeper examination, uncovering cascading effects across interconnected pathways and revealing systemic vulnerabilities that might remain hidden in more localized analyses.

To prioritize mitigation, shortest path algorithms, such as Dijkstra's, identify critical pathways where risks are concentrated, enabling targeted interventions [3]. Furthermore, cycle detection algorithms identify feedback loops, such as those caused by repeated pressure build-ups in recycle streams, which, if unchecked, could exacerbate operational risks. Together, these algorithms form a cohesive suite for analyzing risk propagation, escalation, and mitigation, ensuring a comprehensive hazard assessment that informs safety decision-making.

2.3 AI-Driven HAZOP Scenario Generation

Central to the framework is a sophisticated rule-based AI system that leverages graph analytics and advanced computational models to autonomously generate HAZOP scenarios. Unlike traditional methods, which rely heavily on manual inspection and human intuition, this system employs a structured, algorithmic approach to systematically evaluate deviations and their cascading consequences [4].

The framework integrates engineering principles with graph-based analysis, enabling the AI to identify potential risks across interconnected components with a high degree of precision. For example, when a "low flow" deviation occurs in a pump, the system identifies downstream components at risk by analyzing pipeline configurations, material properties, and operational thresholds [5]. The AI calculates potential cavitation risks and operational bottlenecks, considering both immediate and long-term effects. In the case of a "high-pressure" scenario within a reactor, the system evaluates the effectiveness of relief systems, their ability to manage excess pressure, and the potential repercussions on upstream equipment, such as compressors and heat exchangers. These evaluations are deeply informed by operational data and safety constraints stored within the graph database.

Technically, the AI employs graph traversal algorithms, such as Breadth-First Search (BFS) and Depth-First Search (DFS), to explore the propagation of hazards. It also integrates shortest path algorithms to identify critical nodes where risks may concentrate, and cycle detection algorithms to pinpoint feedback loops that could amplify hazards [3]. By combining these algorithms, the system achieves a comprehensive analysis that uncovers vulnerabilities that might remain hidden in conventional manual assessments.

Moreover, the scenarios generated by this system are rigorously validated against a comprehensive dataset of chemical properties, operational parameters, and industry standards. This validation process ensures the accuracy and relevance of the results while maintaining compliance with established safety protocols. By automating this complex process, the system reduces human oversight, minimizes errors, and accelerates the analysis timeline, delivering actionable insights in a fraction of the time required by traditional methods.

In addition to its analytical capabilities, the system includes a dynamic feedback loop for continuous improvement. Insights derived from past analyses are integrated into the rule-based system, enabling it to refine its scenario generation capabilities over time. This iterative approach ensures that the framework adapts to new data and emerging challenges, making it a highly scalable and future-proof solution for HAZOP analysis in complex industrial environments.

2.4 LangChain and GPT Integration

The integration of LangChain with GPT models transforms how engineers interact with the system. This combination enables natural language queries, allowing users to access complex safety data intuitively. For example, an engineer could ask, "What are the potential hazards associated with HX-101 under high-pressure conditions?" LangChain interprets the query, identifies relevant system components and deviations, and interfaces with Neo4j to retrieve and analyze the required data [5]. The GPT model then synthesizes the findings into a clear, actionable response. For instance, the system might report that high pressure in HX-101 could stress downstream control valves, overload connected pipelines, and activate relief systems, recommending specific mitigation measures. This natural language interface bridges the gap between complex technical analyses and user accessibility, making safety insights comprehensible and actionable for multidisciplinary teams.

3. Results

The implementation of this framework has significantly advanced HAZOP analysis. The creation of a digital twin offers precise modeling of Ammonia-Hydrogen systems, enabling better insights into system behavior.

AI-driven scenario generation ensures thorough hazard identification while reducing reliance on manual efforts, improving risk assessment comprehensiveness. Automation has streamlined the hazard identification process, enabling faster and more efficient responses to potential risks.

Natural language interfaces enhance accessibility, simplifying complex data exploration and fostering effective collaboration. These advancements demonstrate the framework's potential to revolutionize safety analysis and decision-making in complex industrial settings.

4. Conclusion

This study presents a novel framework integrating graph databases, advanced traversal algorithms, and natural language processing to enhance HAZOP analysis for complex systems. By combining digital twin modeling, AI-driven scenario generation, and intuitive interfaces, the framework addresses critical gaps in traditional safety methodologies.

The proposed methods, including BFS, DFS, and cycle detection, ensure comprehensive hazard identification, while LangChain and GPT improve accessibility and decision-making through natural language interactions. Future work will focus on validating the framework across diverse industrial applications and further enhancing its predictive capabilities, establishing it as a universal standard for AI-driven process safety.

Acknowledgments

This project is supported by Singapore's Low Carbon Energy Research (LCER) Phase 2 Programme, under its Hydrogen and Emerging Technologies Funding Initiative (HETFI) - Directed Hydrogen Programme (U2303D4001). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of the Low Carbon Energy Research (LCER) Programme.

References

[1] Chang, H., et al., "Space exploration through graph analytics," *Journal of Advanced Computing*, 2015.

[2] Mettam, G.R., "Chapter on industrial safety," in *Process Safety Handbook*, Elsevier, 2009.

[3] Strunk, W., "The Elements of Graph Style," *Graph Methods Research*, 2000.

[4] Mack, L., et al., "Advances in AI for Industrial Safety," *Industrial AI Journal*, 2015.

[5] Neo4j, Inc., "Introduction to Graph Databases," Neo4j documentation, 2023.