Utilization of Artificial Intelligence in HAZOP Studies and Reports

Ehab Elhosary[1], Osama Moselhi [1,2]

[1]Department of Building, Civil and Environmental Engineering, Concordia University, Montreal, QC, Canada
[2]Centre for Innovation in Construction and Infrastructure Engineering and Management (CICIEM), Concordia University, Montreal, QC, Canada
ehab.elhosary@mail.concordia.ca, moselhi@encs.concordia.ca

Abstract –

The Hazard and Operability (HAZOP) study is the most widely used hazard analysis technique in the process industry, aimed at enhancing safety and preventing accidents. It identifies potential hazards and operability malfunctions by dividing the design into small sections, called nodes, and analyzing each node's process separately. This paper briefly describes HAZOP studies and their role in enhancing operational safety across process industries. It also describes the advancements made to facilitate the performance of these studies and their challenges. To address these challenges, a new framework integrating BERTopic into HAZOP studies is proposed for enhanced efficiency and accuracy. The framework leverages historical data to categorize HAZOP elements into topics and extract node process and equipment descriptions to generate an intelligent pre-populated HAZOP analysis table. This paper focuses on categorizing causes into main risk factors for each HAZOP node and prioritizing them based on the likelihood of occurrence for each factor. The BERTopic model, incorporating embedding generation, dimensionality reduction, clustering, and topic representation, was applied to 1,574 HAZOP records from an oil pump station. The model achieved coherence and diversity scores of 80% and 92.4% respectively, outperforming Latent Dirichlet Allocation (LDA) model at 45.4% and 88.8%. It identified 13 topics, validated against hazard causes in oil pump stations and pipelines from literature. This model can be extended to categorize consequences and countermeasures, prioritizing them by severity and risk levels to generate a prepopulated table. This table can guide participants during sessions, significantly reducing the time required for the final HAZOP report.

Keywords –

HAZOP; Artificial intelligence; BERTopic; LDA; Coherence Score; Topic Diversity Score.

1 Introduction

Hazard analysis encompasses identifying potential hazards, assessing scenarios that could result in adverse outcomes, and developing countermeasures to eliminate or mitigate these hazards [1]. In this regard, hazard and operability (HAZOP) analysis has been the most widely used technique in the process industry. This is due to its simplicity, systematic approach, thoroughness, structured brainstorming process, and applicability to a wide range of systems [2]. The HAZOP concept dates to 1974 following a large explosion in England, which resulted in the deaths of 28 workers and injuries to 36 others. An international standard for HAZOP studies [BS IEC 61882] was published in 2001 and updated in 2016.

A HAZOP study is typically conducted during the design stage to identify and assess potential hazards, their causes, consequences, safeguards, and recommendations. The primary inputs are Piping and Instrumentation Diagrams (P&IDs) and the expertise of participants. P&IDs provide detailed design and engineering information about the plant. Participants contribute specialized knowledge of processes, drawing upon their professional experience [3]. The process begins by dividing the P&IDs into clearly defined sections, referred to as nodes, to ensure that each piece of equipment in the process is analysed thoroughly. The study employs guidewords (e.g., no, more, less of) in conjunction with parameters (e.g., pressure, temperature, flow) to identify deviations from the normal operating conditions. This systematic approach is applied within a specific node. Once deviations are identified, the team investigates their potential causes and consequences and identifies safeguards and recommendations to prevent or mitigate the hazardous situation [4]. The HAZOP study has gained regulatory acceptance in the chemical industry, and has been extended to other industries, including oil and gas, petrochemical plants, nuclear power, environmental engineering, and infrastructure projects [5]. However, it often faces quality issues stemming from complexity,

time consumption, reliance on expert judgment and knowledge loss. Completing the analysis for a typical chemical process takes 1–8 weeks and for large-scale processes, such as plants with over 200 P&IDs, the required time increases significantly [6]. To overcome these issues, intelligent systems with various levels of automation have been developed to reduce time, cost, and human bias [5].

In this respect, this paper introduces the evolution of HAZOP automation within the process industry and highlights the challenges of existing systems. To address these limitations, it proposes a novel framework that leverages artificial intelligence (AI) tools to generate an intelligent pre-populated HAZOP analysis table, designed to assist participants during workshop sessions.

2 Literature Review

HAZOP automation has been a research focus for over 30 years, developing intelligent systems, including knowledge-based systems, model-based approaches, and data-driven models

Knowledge-based systems are computer systems that use knowledge representation to provide expert knowledge or draw conclusions [7]. These systems, include expert systems, computer simulations, integrated based tools, ontologies, and Bayesian networks [5].

Expert systems emulate human logic through knowledge databases that store domain knowledge, inference engines to deduce conclusions based on predefined rules, and user interfaces, enabling diagnosis, fault-finding, and problem-solving. Notable examples include Stateflow [3] and preHAZOP [8]. While these systems automate routine tasks like identifying common causes and consequences, they are often industry-specific and struggle with adapting to new technologies or dynamic process changes or changing environments [4].

Ontologies, represented as directed acyclic graphs (DAGs) with nodes as concepts and edges as relationships. For instance, Yan et al. [9] applied named entity recognition (NER) to extract knowledge from HAZOP reports, organized it into an ontology knowledge base via Protégé and web ontology language (OWL), and employed the HermiT inference engine for automated analysis. This approach improved scenario coverage and safeguard descriptions, though its effectiveness depends on the ontology's scope and detail. Notably, the model focuses on text analysis and lacks automated risk assessment. Additionally, ontologies is time-consuming and biased as they rely on human's experience [4].

Bayesian networks (BNs) are acyclic graphs where nodes represent process variables and arcs depict cause–effect relationships. Each node has probabilistic states used to extrapolate the likelihood of other events. BNs have been applied in HAZOP to quantify risk and event probabilities, with experts assigning prior and conditional probabilities [10]. However, building BNs relies heavily on expert input for structure and prior probabilities, which can be subjective and challenging.

The model-based approaches use detailed mathematical models, combining the HAZOP technique with dynamic simulations to understand system behaviour during failures. This integration enhances hazardous scenario identification and risk assessment, reducing subjectivity in evaluating event severity and likelihood. Dynamic HAZOP simulation uses two main approaches: custom mathematical models tailored to specific units, offering full control over simulation methods or commercial process simulators, such as Aspen Plus, Aspen HYSYS, and k-Spice, which allow flexible adjustments to process flow sheets [11]. While these tools provide robust environments for integrating simulation with process hazard analysis, HAZOP studies rely on teams, and analysing new units requires developing and verifying new equations for each case.

Knowledge-based and model-based approaches have supported HAZOP studies, but acquiring and updating domain-specific knowledge remains challenging, subjective, and time-consuming. Recently, data-driven models have advanced HAZOP studies by leveraging historical data to improve safety outcomes and reduce manual effort [5]. Various classification models have been applied to automate specific aspects of HAZOP reports. For example, the BERT-BiLSTM-Attention model was used to predict severity from consequence descriptions, achieving a precision of 88.6% [12]. Bag of Words (BOW) was combined with ML to predict deviations based on causes, reaching 92% accuracy [13]. Peng et al. [14] proposed an ELMo-DCNN-BiLSTM-CRF model for NER to identify material and equipment terms, achieving higher recall for equipment (93.52% recall) than materials (86.08% recall). Term Frequency Inverse Document Frequency (TF-IDF) was applied with Naïve Bayes to predict likelihood, severity, and risk levels using causes, and consequences, with accuracies exceeding 80% [15]. Ekramipooya et al. [16] utilized Bidirectional Encoder Representations from Transformers (BERT) and multi-layer perceptron (MLP) to predict recommendations using two separate models, one using causes and the other using consequences.

These classification models face some limitations. HAZOP studies involve interconnected elements, such as process parameters, guidewords, causes, consequences, safeguards, and recommendations. Effective automation requires considering all these elements and addressing their interdependencies to avoid inaccuracies. Additionally, they lack prediction of safeguards and recommendations, based on these interconnected HAZOP elements.

Topic modelling methods have been developed to uncover latent topics in HAZOP reports. Wang et al. [15] used Latent Dirichlet Allocation (LDA) to uncover hidden cause and consequence from HAZOP data. TF-IDF, LDA, Part-of-Speech (POS) tagging, and the Apriori algorithm were used to explore correlations between causes and consequences, clustering them into 20 topics [17]. Despite its strengths, LDA does not consider the semantic relationships between words and order of words. BERT technique address this by generating contextual word and sentence vector representations. Ekramipooya et al. [16] leveraged BERT, Uniform Manifold Approximation and Projection (UMAP), and Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) to cluster recommendations as a first step for prediction models. However, they did not use topic representation algorithms, limiting the analysis of the generated topics.

The topic modelling literature has focused on identifying the main topics of causes and consequences for entire plants. However, HAZOP teams typically analyse nodes individually, necessitating the identification of these topics at the node level and their corresponding risk levels for effective risk assessment. Additionally, existing studies fail to categorize safeguards and recommendations, which are essential for guiding preventive measures. Moreover, previous research does not map the identified topics to their associated P&ID nodes, which is crucial since the division of P&IDs into nodes depends on the perspectives of HAZOP participants, based on process flow or main equipment. To address these gaps, this study leverages historical data to generate an intelligent HAZOP table that assists participants during workshops in developing the final report. In 2022, BERTopic, a state-of-the-art topic modelling algorithm built on pretrained sentence transformers, was introduced by Grootendorst [18]. It has shown superior performance in capturing semantic context across various domains. For its ability to provide a comprehensive semantic view and generate coherent topics representing different factors of causes, consequences, and countermeasures, BERTopic was chosen for this study to enhance the efficiency and accuracy of topic modelling and analysis.

3 Proposed Method

Figure 1 illustrates an intelligent analysis method, designed to enhance the efficiency and accuracy of conducting HAZOP studies and aid less experienced engineers by leveraging historical HAZOP data (reports and P&IDs' nodes), closely matching the plant under investigation. Using BERTopic, the method identifies the most frequent topics of root causes, consequences, and countermeasures for each HAZOP node and prioritizes them by likelihood, severity, and risk levels. These topics are also linked to their corresponding deviations. A multimodal model like Gemini [19] extracts detailed process and equipment descriptions from P&ID nodes, providing a deeper understanding of the process and associated risks. Mapping these topics to the P&ID nodes offers participants a clearer understanding of hazard identification and assessment for each node under investigation. By combining these insights, the method generates an intelligent, pre-populated HAZOP analysis table to guide participants during workshops, reducing time and manual effort for similar projects to produce the final HAZOP report. Due to space limitations, this paper will focus on categorizing causes into main risk factors, identifying the most frequent factors for each node and prioritizing them based on the likelihood of associated hazards.

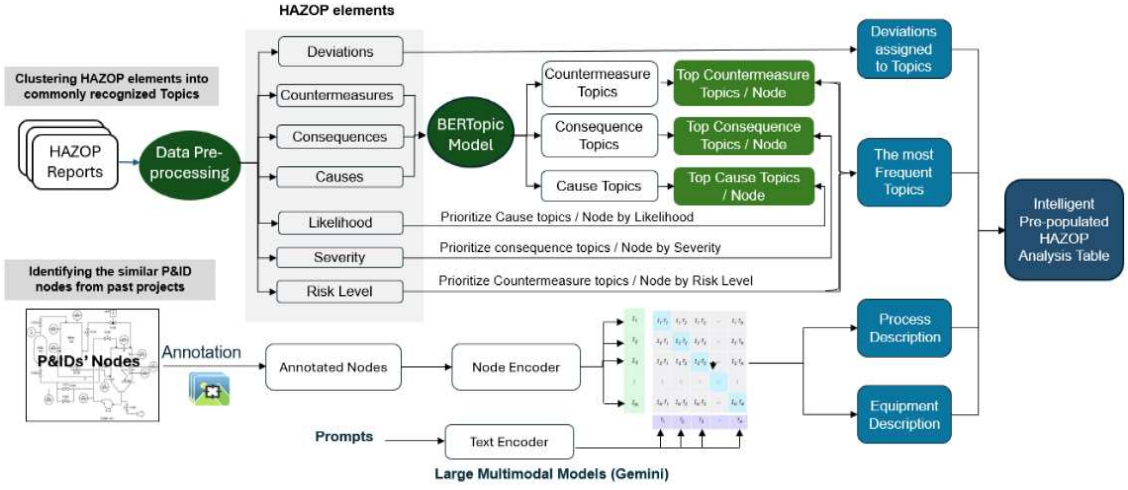

Figure 1. Flowchart of intelligent HAZOP analysis method using Topic Modeling.

The process starts with collecting HAZOP reports and P&IDs' nodes. HAZOP reports include deviations, causes, consequences, safeguards and recommendations, likelihood, severity, and risk. The dataset is shuffled to minimize order bias and enhance model generalization [20]. The preprocessing step involves tokenization, text cleaning, and normalization. Tokenization breaks text into smaller units (tokens) for granular analysis. Text cleaning removes irrelevant elements like stopwords, punctuation, and special characters while normalization uses techniques like lemmatization and lowercase conversion to standardize text [21].

Following preprocessing, the BERTopic model is applied as shown in Figure 2, including vector embedding, dimensionality reduction, clustering, tokenization, and topic representation. It begins with embedding extraction using Sentence-BERT (S-BERT) to generate dense -size embeddings that capture semantic meaning. These embeddings include token embeddings, segment embeddings, and position embeddings. To produce a single-size embedding for each sentence, S-BERT applies mean pooling across token embeddings, resulting in 768-dimensional vectors [22].

To optimize the HDBSCAN process, UMAP is employed for dimensionality reduction, which preserves both local and global structures of high-dimensional embeddings in a lower-dimensional space, facilitating efficient clustering [23]. Key hyperparameters, such as the number of neighbours, the number of components, and the distance metric (e.g., cosine similarity), are fine-tuned to maintain semantic relationships and improve clustering performance [24].

The datapoints are then tokenized and vectorized. Tokenization breaks the text into smaller units like words or phrases. CountVectorizer was used to transform these textual units into numerical vectors [25]. Topic representation is then refined using class-TF-IDF, that identifies distinguishing features between clusters. KeyBERT further enhances topic representation by selecting the most relevant keywords based on their cosine similarity to cluster embeddings [25].

Topic quality is evaluated using coherence and diversity metrics. The coherence score (C_V) assesses the semantic relatedness of topic words, with higher scores (e.g., 0.6–1.0) indicating well-defined and interpretable topics [26]. The Coherence score is calculated as follows:

$$C\_(T) = 2 / (|T|(|T| - 1)) * \Sigma\_\{i=1\}^\{|T|\} \Sigma\_\{j \neq i\} sim(Word\_i, Word\_j) \quad (1)$$

The coherence score (C_V (T)) for a given topic T is calculated based on the number of words in the topic (|T|), the representation of two distinct words (Wordi and Wordj) within that topic, and the similarity measure (sim (Wordi, Wordj)) between these word pairs.

The diversity score measures the uniqueness of topic words across clusters, with values closer to 1 reflecting minimal redundancy. While metrics provide quantitative insights, human evaluation remains essential for validating the interpretability of topics [27].

To enhance topic exploration and reduce outliers, Zero-shot BERTopic uses predefined topics and cosine similarity thresholds to guide clustering. By adjusting the similarity threshold, the model flexibly assigns documents to predefined topics or generates new topics as needed [27]. This iterative process, paired with human judgment, ensures the final topics are actionable.

The model was evaluated against the LDA model, a probabilistic method for uncovering latent topics in text. LDA uses a document-term matrix, treating each document as a mixture of topics and attributing words to these topics based on co-occurrence patterns [28]. TF-IDF was used to enhance feature representation, as it weights terms based on their frequency within documents and importance across the corpus. The resulting TF-IDF matrix serves as input for LDA, enabling it to analyse term significance and distribution to uncover topics.

The models were implemented on Google Colab (CPU runtime), with Key libraries including NLTK and SpaCy for preprocessing, PyTorch, BERTopic, and scikit-learn for topic modelling, and Gensim for evaluation metrics. Core Python packages like NumPy, pandas, Matplotlib, and Seaborn supported data manipulation and visualization.

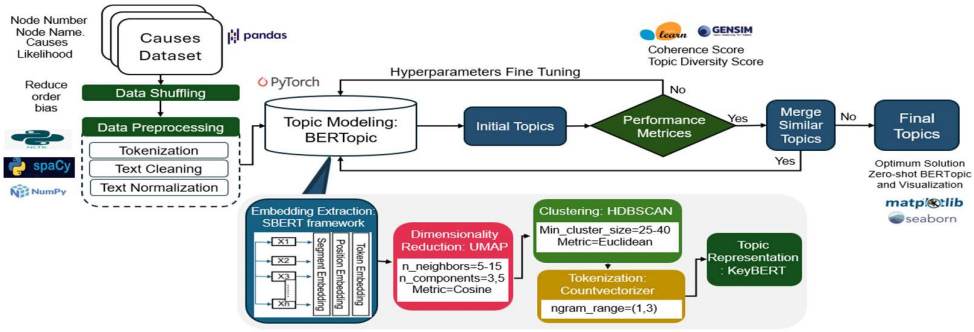

Figure 2. Methodology to apply BERTopic Model on HAZOP Causes.

4 Case Study and Results

The data utilized in this study comprised node name, causes, and likelihood derived from HAZOP reports focused on oil pump stations, which play a vital role in transporting oil through pipelines between various locations. The dataset was sourced from the thesis of Kilincet al. [29] report of an oil pump station located in Sivas, Turkey, designed to ensure safe and efficient transfer of oil via pipelines. At the station, oil enters from the pipeline and first reaches the pig receiving system. It then passes through the main pump filters before reaching the pumps, where it is energized and sent through the pig launching section back into the pipeline toward its next destination. Additionally, the wax handling system, used to remove wax (solid oil), represents one of the most challenging processes in the operation. The dataset encompasses six nodes and 1,574 data points required for safe oil transfer within the plant.

The BERTopic model was fine-tuned by experimenting with various hyperparameters to optimize topic generation. Key parameters included min_topic_size (20–50) to control the minimum documents per topic, affecting topic granularity; top_n_words (15) to emphasize the most representative terms for interpretability; n_neighbors (10–15) in UMAP for balancing local and global data structures during dimensionality reduction; n_components (3–5) to set the dimensionality of reduced embeddings. These configurations resulted in topics ranging from 6 to 43. Figure **3** illustrates topic coherence and diversity scores across multiple experiments, providing a quantitative assessment of topic quality. Coherence scores peaked at 32 topics, achieved 0.898, while diversity score was highest at 9 topics, with 0.967. Variations in the number of top words had no impact on the scores.

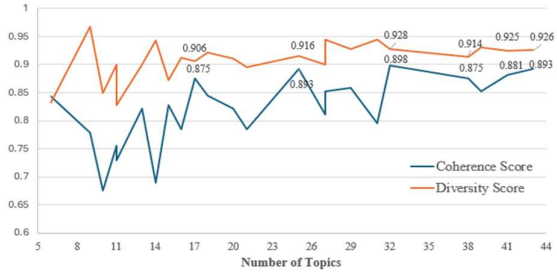

Figure 3. Coherence and Topic Diversity Scores.

As highlighted earlier, the resulting topics should be evaluated through human judgment [30]. Consequently, the topics and their top words were manually reviewed, focusing on experiments with higher coherence and topic diversity scores, such as those with 17, 25, 32, 38, 41, and 43 topics. The evaluation process was guided by the main causes of hazards identified in oil pump stations and pipelines in the literature, including operator error, inadequate pipe management, corrosion, insufficient maintenance, ignored alarms, fire, explosions, and equipment issues [31], [32]. This review assessed whether each cluster contained distinct topics and how well-separated these topics were, or if overlapping top words indicated a lack of distinction among them.

Following this review, 41 topics were initially selected because they exhibited less overlap. However, further analysis revealed several groups of similar topics that required merging. Topics 1, 2, 3, 12, 19, and 21 were found to be closely related and grouped together, as were topics 0, 11, 22, 24, 31, and 33, topics 8, 9, 10, 16, 18, 34, 35, and 36, topics 4, 29, 30, and 39, topics 7, 17, and 27, topics 6, and 14, topics 20, 23, and 28, topics 26, and 38, and topics 13 and 25. These similar topics were merged, and additional experiments were conducted to optimize the coherence and topic diversity scores, resulting in 13 topics, achieving coherence and diversity scores of 80.0% and 92.4%, respectively. Table 1 presents the final topics, including the top frequent words (1-3 grams) associated with each topic and their corresponding suggested labels. Cause topic generation is based on topic words and manual induction. Thus, the topic's formation needs knowledge, experience, and analysis. This approach ensured a robust and interpretable topic modelling outcome while maintaining high performance metrics.

It is worth noting that this model resulted in 71 outliers, stemming from several factors. Some data points contain very short descriptions and general statements, providing insufficient semantic information for clustering. Additionally, the dataset exhibited high variability in text length across data points, further contributing to the outliers. Zero-shot BERTopic was applied to explore potential new topics for documents that did not align with predefined topics. Experiments were conducted with cosine similarity values ranging from 0.5 to 0.95. Values between 0.5 and 0.8 did not yield new topics, while values above 0.8 introduced new topics that closely resembled the predefined ones. Based on these findings, a cosine similarity value of 0.8 was selected, which effectively reduced the number of outliers from 71 to just 32 documents. The frequency of topics is presented in Table 1.

Table 1. The Cause topics and their corresponding labels and Frequency.

Topic	Top Words	Category	Frequency
0	valve, operate valve, motor operate valve, operate, error, close omit, main oil line, completely omit, oil line pump	Omitted or wrong operations	43
1	slop tank, slop tank pump, failure, reinjection, reinjection pump, switch, interlock, flow, block, slop tank block	Pump failure / malfunctions.	200
2	valve, pass, open, valve skid, skid, check valve, check, open pass, valve fail open, safety valve fail, fail open pass	Valve errors or failures.	754

3	maintenance, closure, close maintenance case, blind leave close, maintenance case block, leave close maintenance	Inadequate maintenance	58
4	emergency, emergency shutdown, error false signal, alarm may fail, shutdown valve, error, level, faulty level alarm	Safety system failure	60
5	multiple, multiple simultaneous, multiple drain operation, multiple drain, simultaneous, simultaneous thermal	Simultaneous operations	77
6	roof, complete, complete loss, gauge lead unintended, surge valve gauge, valve gauge, valve gauge lead, valve	Relief devices error/ failure	35
7	insulation, corrosion, corrosion insulation, insulation cui, corrosion insulation cui, cui, cui tank wall, insulation low	Corrosion	87
8	external fire, fire external, external fire external, fire external fire, fire, external	External fire	110
9	underway leak heating, also source, heating coil may, leak heating, leak heating coil, event underway leak,	Leakage	39
10	mode, mistakenly put, put back, put, pump leave manual, tank pump leave, mode mistakenly put, manual mode	Human errors	5
11	pipe, bore, pipe rupture within, pipe rupture, rupture, large bore pipe, large bore, large, bore pipe rupture	Pipe failure or error.	50
12	transmitter fail, transmitter, failure level, failure level transmitter, level transmitter, level transmitter fail	Transmitter failure /error	24

The identified topics were validated against the 11 main causes identified manually by Kilincet al. [29], including omitted operations, fires, inadequate maintenance, operational failures, out-of-range, wrong equipment, incorrect operations, misplaced equipment, equipment failures, safety system failures, and simultaneous operations. The validation confirmed a strong alignment with traditional methods, while the model also provided more granular insights, such as pump malfunctions, pipe rapture, valve failures (the most frequent category), and relief device failures, which align with categories like equipment failure or wrong equipment. Additionally, the model revealed extra causes, such as human errors, leakage, and corrosion, expanding the depth and detail of the traditional HAZOP analysis. Identifying main causes for the entire plant provides valuable insights, but the HAZOP team evaluates nodes individually. Therefore, it is crucial to pinpoint the key causes for each node and determine which risk factors are included, excluded, and prioritized. Figure 4 illustrates the topics and their frequency for each node, highlighting the most critical causes. For instance, in Node 1 (Pig Receiving), the primary causes include valve failures, pipe ruptures, and omitted operations. Notably, Node 1 accounts for 11 out of the 13 identified causes, excluding pump malfunctions and relief device failures.

In addition to those identified main causes for each node, prioritizing them based on likelihood (e.g., 1 to 8, where 1 indicates an event has not yet occurred in the industry and 8 indicates frequent occurrence at the facility) ensures that critical issues are addressed first. High-frequency topics within each likelihood category highlight recurring or critical hazards, guiding the implementation of effective safeguards.

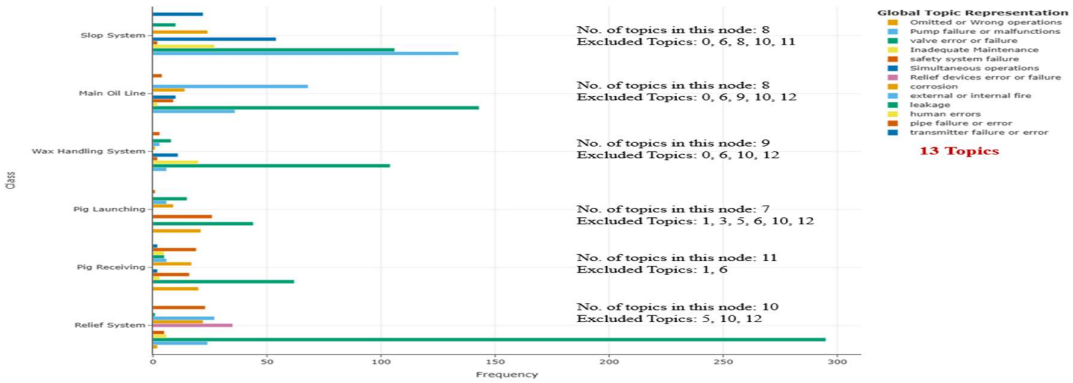

Figure 4. The included and excluded topics associated with each P&ID's node

For instance, in node 2 (Pig Launching), as shown in Figure 5**,** the frequency distribution of topics across probability categories "2," "3," "4," "5," "6," and "7" identifies specific areas requiring mitigation strategies. Topic 9 (leakage) has the highest frequency in category "7," indicating a high probability of occurrence and making it a priority for risk assessment. In contrast, Topic 0 (omitted or wrong operations), Topic 2 (valve errors or failures), and Topic 4 (safety systems failure) are the most frequent topics in categories "4," "3," and "2," respectively. Although these categories represent lower likelihoods, their recurrence across multiple deviations within the same node could cumulatively escalate the overall risk. This emphasizes the need to evaluate whether the frequency of such hazards justifies earlier intervention to prevent escalating if left unaddressed.

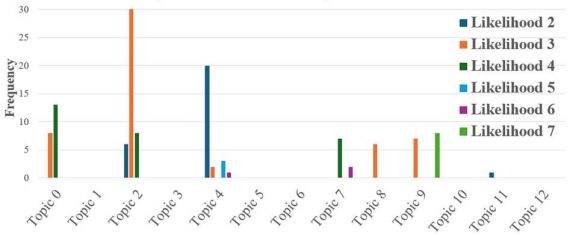

Figure 5. Categorization of topics by likelihood for Node 2 (Pig Launching).

The BERTopic model was compared to the LDA model to evaluate its effectiveness. LDA was selected as it has been extensively applied in HAZOP studies. Since LDA requires a predefined number of topics, a range from 6 to 43 topics (matching BERTopic's output) was tested. Various hyperparameters were explored, including passes (10–100), iterations (1000–5000), and alpha and beta values (0.01). Multiple experiments were conducted to identify the optimal number of topics (K), determined using perplexity and coherence metrics, as smaller perplexity and higher coherence typically indicate better topic recognition [28]. Figure 6 shows perplexity was lowest at 42 topics, while coherence peaked at 12 topics with a score of 45%. However, low and fluctuating coherence scores indicated LDA's challenges in producing stable and interpretable results. Topic diversity was highest (89%) at 8 topics. A detailed review of 42, 12, 8, and 41 topics revealed overlapping clusters, complicating the selection of optimal topics.

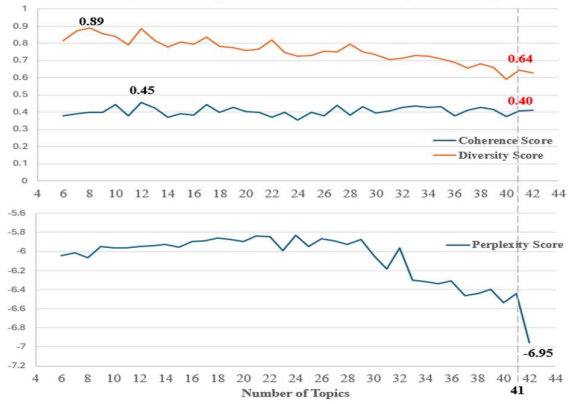

Figure 6. LDA Model Evaluation Metrices.

5 Conclusion

This paper briefly reviewed current developments of intelligent systems in HAZOP studies, emphasizing the potential of utilizing historical HAZOP data to streamline the process. The review reveals several gaps including inadequate incorporation of all HAZOP elements in classification models, prediction of countermeasures based on these interconnected elements, and categorization of causes, consequences, and countermeasures by node, likelihood, severity, and risk levels. This study applied the BERTopic model to categorize causes into 15 risk factors, achieving a coherence score of 80% and a topic diversity score of 92.4%, which outperforms LDA's 45.4% coherence and 88.8% diversity. The model identified key topics by node and likelihood. This approach can be extended to consequences and countermeasures, generating a prepopulated HAZOP table to assist workshop participants in verifying or identifying additional issues.

The main limitation of this study is its focus on oil pump stations, using data from a single HAZOP report, which may affect the model's generalizability. Expanding the dataset is recommended to improve model performance. Moreover, hyperparameter tuning for BERTopic was time-consuming, requiring multiple trials for high coherence and diversity. Optimization techniques are suggested to streamline this process.

References

[1] A. Jensen and T. Aven, "Hazard/threat identification: Using functional resonance analysis method in conjunction with the Anticipatory Failure Determination method," *Proc Inst Mech Eng O J Risk Reliab*, vol. 231, no. 4, pp. 383–389, Aug. 2017, doi: 10.1177/1748006X17698067.

[2] F. Joubert, E. Steyn, and L. Pretorius, "Using the HAZOP Method to Conduct a Risk Assessment on the Dismantling of Large Industrial Machines and Associated Structures: Case Study," *J Constr Eng Manag*, vol. 147, no. 1, Jan. 2021, doi: 10.1061/(asce)co.1943-7862.0001942.

[3] M. F. Chia and P. K. Naraharisetti, "HAZOP using Stateflow software: Methodology and case study," *Process Safety and Environmental Protection*, vol. 179, pp. 137–156, Nov. 2023, doi: 10.1016/j.psep.2023.09.005.

[4] J. Single, J. Schmidt, and J. Denecke, "Computer-aided hazop studies: Knowledge representation and algorithmic hazard identification," in *WIT Transactions on the Built Environment*, WITPress, 2019, pp. 55–66. doi: 10.2495/SAFE190061.

[5] E. Elhosary and O. Moselhi, "Automation for HAZOP study: A state-of-the-art review and future research directions," *Journal of Information Technology in Construction*, vol. 29, pp. 750–777, Sep. 2024, doi: 10.36680/j.itcon.2024.033.

[6] H. J. Pasman and W. J. Rogers, "How can we improve HAZOP, our old work horse, and do more with its results? an overview of recent developments," *Chem Eng Trans*, vol. 48, pp. 829–834, 2016, doi: 10.3303/CET1648139.

[7] J. I. Single, J. Schmidt, and J. Denecke, "Ontology-based support for hazard and operability studies," *International Journal of Safety and Security Engineering*, vol. 10, no. 3, pp. 311–319, Jun. 2020, doi: 10.18280/ijsse.100302.

[8] J. Oeing, T. Holtermann, W. Welscher, C. Severins, M. Vogel, and N. Kockmann, "preHAZOP: Graph-Based Safety Analysis for Early Integration into Automated Engineering Workflows," *Chem Ing Tech*, vol. 95, no. 7, pp. 1083–1095, Jul. 2023, doi: 10.1002/cite.202200222.

[9] H. Yan, S. Wu, L. Yuan, S. Wang, and Y. Cao, "A New HAZOP Automated Method for the Oil and Gas Complex Equipment," *Chemistry and Technology of Fuels and Oils*, vol. 59, no. 4, pp. 872–879, Sep. 2023, doi: 10.1007/s10553-023-01592-8.

[10] P. Gao and W. Li, "Integration of HAZOP and Bayesian network in city gas explosion emergency response processes," *Emergency Management Science*

and Technology, vol. 2, no. 1, pp. 1–9, 2022, doi: 10.48130/emst-2022-0019.

[11] J. Janošovský, M. Danko, J. Labovský, and Ľ. Jelemenský, "Software approach to simulation-based hazard identification of complex industrial processes," *Comput Chem Eng*, vol. 122, pp. 66–79, Mar. 2019, doi: 10.1016/j.compchemeng.2018.05.021.

[12] X. Feng, Y. Dai, X. Ji, L. Zhou, and Y. Dang, "Application of natural language processing in HAZOP reports," *Process Safety and Environmental Protection*, vol. 155, pp. 41–48, Nov. 2021, doi: 10.1016/j.psep.2021.09.001.

[13] A. Ekramipooya, M. Boroushaki, and D. Rashtchian, "Application of natural language processing and machine learning in prediction of deviations in the HAZOP study worksheet: A comparison of classifiers," *Process Safety and Environmental Protection*, vol. 176, pp. 65–73, Aug. 2023, doi: 10.1016/j.psep.2023.06.004.

[14] L. Peng, D. Gao, and Y. Bai, "A study on standardization of security evaluation information for chemical processes based on deep learning," *Processes*, vol. 9, no. 5, 2021, doi: 10.3390/PR9050832.

[15] F. Wang and W. Gu, "Intelligent HAZOP analysis method based on data mining," *J Loss Prev Process Ind*, vol. 80, Dec. 2022, doi: 10.1016/j.jlp.2022.104911.

[16] A. Ekramipooya, M. Boroushaki, and D. Rashtchian, "Predicting possible recommendations related to causes and consequences in the HAZOP study worksheet using natural language processing and machine learning: BERT, clustering, and classification," *J Loss Prev Process Ind*, vol. 89, Jul. 2024, doi: 10.1016/j.jlp.2024.105310.

[17] F. Wang, W. Gu, Y. Bai, and J. Bian, "A method for assisting the accident consequence prediction and cause investigation in petrochemical industries based on natural language processing technology," *J Loss Prev Process Ind*, vol. 83, Jul. 2023, doi: 10.1016/j.jlp.2023.105028.

[18] M. Grootendorst, "BERTopic: Neural topic modeling with a class-based TF-IDF procedure," Mar. 2022, [Online]. Available: http://arxiv.org/abs/2203.05794

[19] M. Imran and N. Almusharraf, "Google Gemini as a next generation AI educational tool: a review of emerging educational technology," *Smart Learning Environments*, vol. 11, no. 1, p. 22, May 2024, doi: 10.1186/s40561-024-00310-z.

[20] B. Li, Y. Esfandiari, M. N. Schmidt, T. S. Alstrøm, and S. U. Stich, "Synthetic data shuffling accelerates the convergence of federated learning under data heterogeneity," *Transactions on Machine Learning Research*, Apr. 2024, doi: https://doi.org/10.48550/arXiv.2306.13263.

[21] G. Liu, M. Boyd, M. Yu, S. Z. Halim, and N. Quddus, "Identifying causality and contributory factors of pipeline incidents by employing natural language processing and text mining techniques," *Process Safety and Environmental Protection*, vol. 152, pp. 37–46, Aug. 2021, doi: 10.1016/j.psep.2021.05.036.

[22] E. Aytaç and M. Khayet, "A Topic Modeling Approach to Discover the Global and Local Subjects in Membrane Distillation Separation Process," *Separations*, vol. 10, no. 9, p. 482, Sep. 2023, doi: 10.3390/separations10090482.

[23] L. McInnes, J. Healy, N. Saul, and L. Großberger, "UMAP: Uniform Manifold Approximation and Projection," *J Open Source Softw*, vol. 3, no. 29, p. 861, Sep. 2018, doi: 10.21105/joss.00861.

[24] R. Zhou, W. Gao, D. Ding, and W. Liu, "Supervised dimensionality reduction technology of generalized discriminant component analysis and its kernelization forms," *Pattern Recognit*, vol. 124, p. 108450, Apr. 2022, doi: 10.1016/j.patcog.2021.108450.

[25] M. Khayet, E. Aytaç, and T. Matsuura, "Bibliometric and sentiment analysis with machine learning on the scientific contribution of Professor Srinivasa Sourirajan," *Desalination*, vol. 543, p. 116095, Dec. 2022, doi: 10.1016/j.desal.2022.116095.

[26] D. Bretsko, A. Belyi, and S. Sobolevsky, "Comparative Analysis of Community Detection and Transformer-Based Approaches for Topic Clustering of Scientific Papers," in *Computational Science and Its Applications – ICCSA 2023: 23rd International Conference, Proceedings, Part I*, Athens, Greece, Jul. 2023, pp. 648–660. doi: 10.1007/978-3-031-36805-9_42.

[27] J. Kim and S. Lee, "Technology Opportunity Analysis for Creating Innovative Solutions: Applying Semi-supervised Topic Modelling on Patent Data," in *2024 PICMET conference*, IEEE, Aug. 2024, pp. 1–9. doi: 10.23919/PICMET64035.2024.10653159.

[28] H. Xu *et al.*, "Cause analysis of hot work accidents based on text mining and deep learning," *J Loss Prev Process Ind*, vol. 76, p. 104747, May 2022, doi: 10.1016/j.jlp.2022.104747.

[29] M. O. , Kilinc, G. A. , Ciftcioglu, and M. N. , Kadirgan, "SIS Application at a Petroleum Crude Oil Pipeline Pump Station After HAZOP Study," MARMARA University, ISTANBUL, 2020.

[30] C. Doogan and W. Buntine, "Topic Model or Topic Twaddle? Re-evaluating Semantic Interpretability Measures," in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2021, pp. 3824–3848. doi: 10.18653/v1/2021.naacl-main.300.

[31] S. Sendy and F. T. Basaria, "A Comparative Analysis of Hazard Analysis Methods for Sustainable Energy Development," *E3S Web of Conferences*, vol. 388, p. 01037, May 2023, doi: 10.1051/e3sconf/202338801037.

[32] A. Waqar, I. Othman, N. Shafiq, and M. S. Mansoor, "Evaluating the critical safety factors causing accidents in downstream oil and gas construction projects in Malaysia," *Ain Shams Engineering Journal*, vol. 15, no. 1, p. 102300, Jan. 2024, doi: 10.1016/j.asej.2023.102300.