

Article

High-Fatality Escalation Pathways in Hazardous Chemical Accidents: A Hierarchical Configurational Analysis for Process Safety

Jingwen Zhang ^{1,2}, Yanan Li ², Yuhao Wang ³ and Bing Feng ^{1,*}¹ School of Management, Wuhan Institute of Technology, Wuhan 430205, China; zjwchris@163.com² School of Mechanical Engineering, Henan Polytechnic Institute, Nanyang 473000, China; lyn_03200@163.com³ School of Joint Design and Innovation, Xi'an Jiaotong University, Xi'an 710049, China; 11018689@mail.polimi.it

* Correspondence: fengbing@wit.edu.cn

Abstract

Accidents in process industries continue to cause severe casualties, and a small number of events account for a large share of fatalities. This study proposes a Topic–Hierarchy Coincidence Analysis (T-H CNA) framework to identify condition combinations associated with high-fatality outcomes by integrating BERTopic, Human Factors Analysis and Classification System (HFACS), and Coincidence Analysis (CNA). The framework is applied to 121 Chinese investigation reports of serious-or-above chemical accidents from 2015 to 2025. BERTopic is used to extract 25 causal semantic themes from accident-cause texts, which are then mapped by expert classification onto eight second-level HFACS categories (Fleiss' $\kappa = 0.7118$). On this basis, CNA identifies minimally sufficient configurations (MSCs) and traces their cross-level transmission pathways. Six three-condition MSCs are obtained, with consistency values ranging from 0.750 to 0.923; the broadest pathway covers 28.3% of high-fatality cases. Deficient organizational climate and failure to correct known problems recur as the main upstream endpoints and remain stable under stricter fatality thresholds. Although operational errors appear in 88.43% of cases, they do not enter any sufficient configuration. The results indicate that high-fatality outcomes are more closely associated with coupled upstream organizational and supervisory failures than with terminal errors alone, supporting upstream-oriented process safety governance.

Keywords: chemical accident; high-fatality outcomes; BERTopic; HFACS; Coincidence Analysis; configurational causation; process safety



Academic Editor: Alpaslan Atmanli

Received: 11 May 2026

Revised: 5 June 2026

Accepted: 24 June 2026

Published: 26 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

In recent years, severe chemical accidents continue to occur in major industrial countries worldwide. Although chemical accidents are not highly frequent overall, once they occur, they can rapidly cause substantial casualties and serious property losses [1,2]. In terms of outcome distribution, such accidents exhibit a pronounced unevenness, in which a small number of events account for the bulk of casualties and losses [2,3]. For example, typical scenarios such as chemical releases in confined spaces, runaway reactions, and cascading explosions show extremely high fatality rates, far exceeding those of routine accidents in other industrial sectors [4].

The core task of safety science research lies not only in explaining how accidents have occurred, but also in identifying and mitigating the systemic risks capable of triggering

catastrophic events, so as to reduce casualties, environmental damage, and socioeconomic losses. In the context of chemical accidents, this task points to a question that still requires deeper examination: under similar institutional constraints and compliance requirements, why do some serious chemical accidents remain limited to a few fatalities, whereas others result in multiple deaths and catastrophic consequences [5]? The formation of severe accident outcomes is more likely the result of the coupling of multiple conditions under specific circumstances [6]. Identifying the condition combinations associated with high-fatality outcomes in serious chemical accidents has become an urgent scientific issue in chemical safety management [7,8]. In the process industry, frontline operators are not only at the forefront of system operation, but also the most direct victims of accident disasters. Protecting their lives is not only a basic requirement of occupational health and process safety management, but also a central concern for the sustainable development of the chemical industry.

Accident analysis is an important part of process safety research [9]. Theories of accident causation have undergone a paradigmatic shift from single-factor linear models, such as the Heinrich domino model and the Swiss Cheese Model [10], to systems-based models, such as STAMP, AcciMap, and the 24Model [11,12]. Subsequent research generally recognizes that, in complex sociotechnical systems, accidents result from the combined effects of failures across multiple levels, including organization, supervision, preconditions, and execution. However, this theoretical consensus has long been constrained in empirical practice by bottlenecks in the large-scale processing of accident text data and in coding efficiency. A large amount of accident information is recorded in investigation reports as unstructured narrative text, and traditional manual coding is not only costly, but also subject to coder bias and variation in inter-rater consistency [13].

In recent years, NLP and text mining methods have been gradually introduced into safety science to ease the bottleneck in large-sample coding [14,15]. Existing studies have used keyword co-occurrence, TF-IDF-LDA, and BERT-based pretrained models to mine accident causes from investigation reports [16]. Among these methods, BERTopic is built on Transformer-based pretrained embeddings and uses a modular architecture consisting of embedding, UMAP dimensionality reduction, HDBSCAN clustering, and c-TF-IDF reweighting to generate topics with strong contextual sensitivity and semantic interpretability [17]. Compared with classical LDA, it shows clear advantages in both topic coherence and topic diversity [18,19]. Recent studies in civil aviation and construction further demonstrate that BERTopic can reveal fine-grained risk scenarios that conventional bag-of-words models fail to capture [18,20,21].

The core advantage of BERTopic lies in its ability to uncover fine-grained causal semantic clusters under unsupervised conditions that are difficult to capture through manual review. It transforms the traditional implicit coding process, which is largely guided by expert judgment, into a data-driven and explicit inductive paradigm [22], thereby improving the efficiency of identifying causal themes in large-scale accident texts. BERTopic is particularly well suited to chemical accident reports, which are often characterized by dense technical terminology and highly homogeneous sentence patterns. However, the method stops at topic description rather than causal explanation. The resulting topic set is essentially a collection of discrete labels without hierarchical structure or configurational relationships [23]. This limitation makes it necessary to embed a theoretical framework and a causal method on top of topic mining to support further analysis.

The Human Factors Analysis and Classification System (HFACS), which was developed on the basis of the Swiss Cheese Model, classifies accident causes into four levels: organizational influences, unsafe supervision, preconditions for unsafe acts, and unsafe acts [24]. It provides a standardized tool for structurally mapping active failures and latent

conditions [24]. HFACS has been widely validated in safety science [25–27]. In recent years, revised frameworks such as HFACS-OGI for chemical and oil and gas settings and HFACS-CSMEs for the Chinese context have been proposed, further improving its domain applicability [28,29].

HFACS mainly serves the functions of causal classification and hierarchical organization. It cannot by itself explain how different factors work together, nor can it directly identify which combinations of factors lead to specific accident outcomes. In existing research, scholars usually apply additional methods after completing HFACS coding, such as frequency statistics, chi-square tests, Bayesian networks, and association rule mining [30–32]. These methods can help identify high-frequency factors, statistical associations, or probabilistic dependencies, but they still show clear limitations in explaining complex accident causation. First, they are better suited to evaluating the importance of individual factors than to identifying systemic mechanisms, whereas the causes of chemical accidents are often highly systemic. Second, they tend to focus only on the outcome itself and are less able to identify transmission mechanisms across hierarchical levels at the same time.

Over the past decade, the configurational perspective that has emerged in management and the social sciences has emphasized the co-occurrence of conditions, the coexistence of multiple pathways, and causal asymmetry. Configurational comparative methods (CCMs) are well suited to identifying complex causal structures. Within this family of methods, Qualitative Comparative Analysis (QCA) has been introduced into safety science and accident analysis [6,33], which confirms the transferability of configurational thinking to safety research. However, QCA uses a top-down maximization and minimization algorithm, and the resulting solution set does not inherently guarantee freedom from redundancy [34]. In addition, it can model only one outcome at a time and therefore cannot analyze causal chain structures [35]. Yet the latter is precisely a core feature of research on hierarchical transmission mechanisms in safety science, including studies based on HFACS and STAMP.

Coincidence Analysis (CNA) was proposed in 2009. Like QCA, it shares the INUS view of causation, but it adopts a bottom-up search strategy. The minimally sufficient conditions (MSCs) identified by CNA are inherently free of redundant conditions within each configuration and can therefore be interpreted directly in terms of INUS causation. More importantly, among INUS-discovery methods, CNA is distinctive in its ability to simultaneously model multiple outcomes and identify both common-cause and causal-chain structures [35,36]. This property makes CNA particularly suitable for hierarchical assumptions such as those in HFACS. Upstream levels can function not only as distal causes of the final outcome, namely high-fatality outcomes, but also as proximate causes of intermediate levels. Within a single analytical framework, CNA can simultaneously identify configuration-to-outcome relationships and cross-level transmission pathways [37].

In recent years, CNA has accumulated relatively mature applications in public health and implementation science [37–39], but its use in safety science, especially in research on the causes of accidents, is still at an early stage. At the same time, the growing emphasis on hierarchical structure analysis in safety science aligns well with the strengths of CNA.

To address the issues outlined above and respond to the practical need to identify high-fatality outcomes in chemical accidents, this study proposes a Topic-Hierarchy Coincidence Analysis framework (T-H CNA) for accident analysis. The framework integrates three stages: BERTopic-based data-driven topic mining, HFACS-based theoretical mapping, and CNA-based configurational causal identification. BERTopic is used to ease the bottleneck of coding scalability in large-sample accident text analysis. HFACS is used to provide a hierarchical theoretical framework for interpreting the topics identified through data-driven

analysis. CNA is then used to identify configurational patterns associated with high-fatality outcomes. Their complementary relationships are shown in Figure 1.

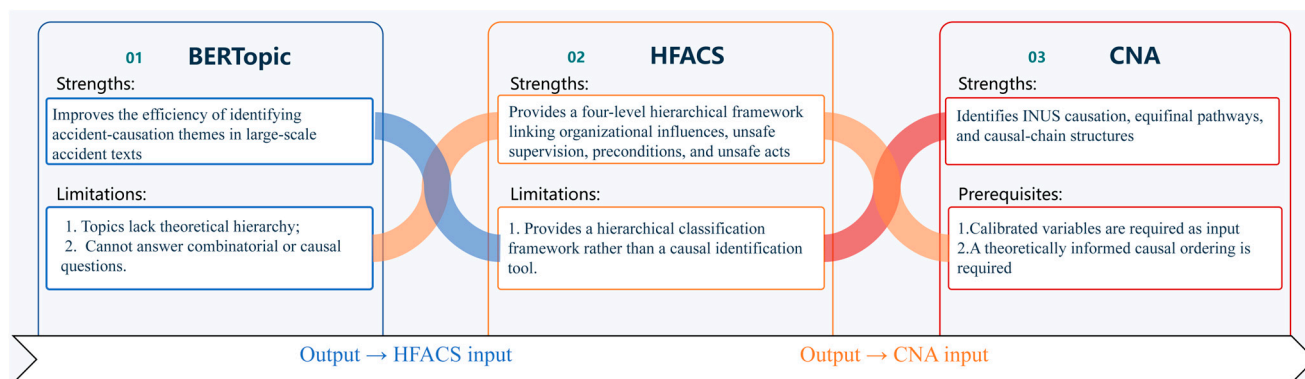


Figure 1. Complementary roles of BERTopic, HFACS, and CNA in the T-H CNA framework.

The study makes two contributions. Methodologically, it provides a reproducible workflow that combines natural language processing, hierarchical theoretical mapping, and configurational causal analysis, offering a transferable approach for the structured identification of causal factors in accident texts from high-risk industries. Practically, the recurrent configurations identified provide empirical support for upstream-oriented safety governance—shifting regulatory priorities toward organizational climate and corrective-action closure to reduce the risk of high-fatality outcomes and protect frontline operators.

2. Materials and Methods

2.1. Construction of the Research Framework

This study takes chemical accident investigation reports from 2015 to 2025 as the base corpus for text mining analysis. As shown in Figure 2, the construction of the accident causation configuration analysis framework consists of three modules.

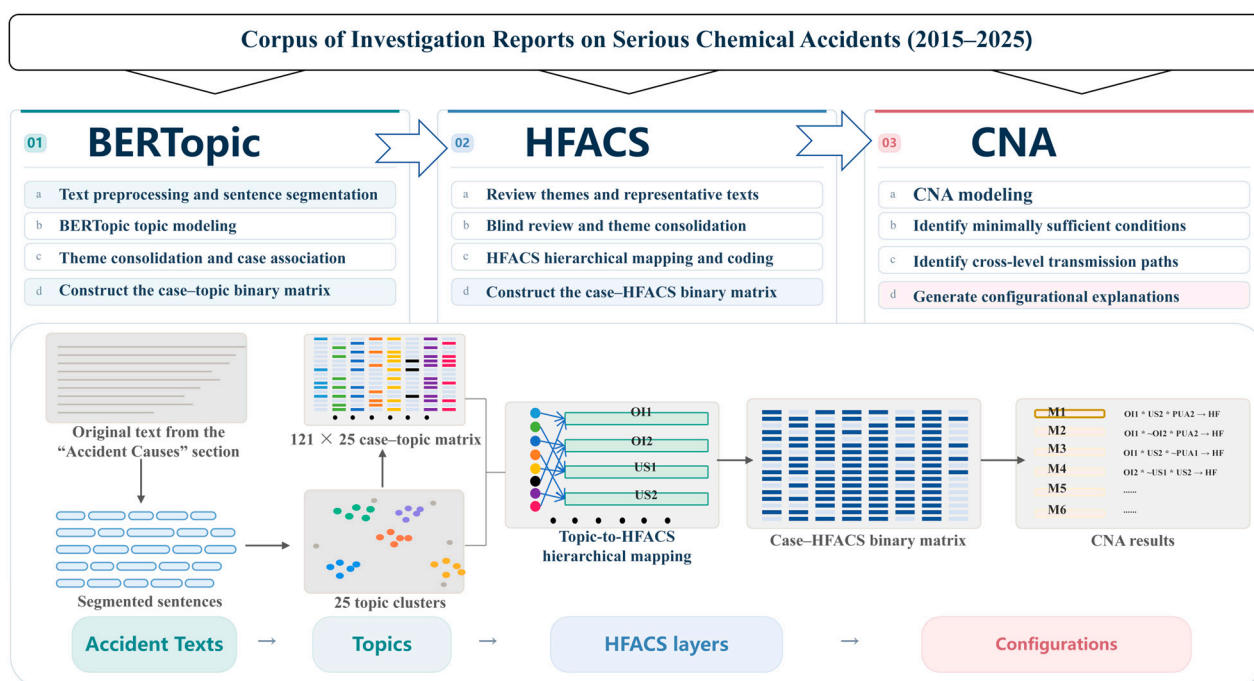


Figure 2. Overall workflow of the T-H CNA framework. * indicates logical conjunction (AND), whereas ~ indicates the negation of a condition.

(1) Accident causation topic mining based on BERTopic. The accident cause texts in the investigation reports are first subjected to preprocessing and sentence segmentation. BERTopic is then applied to perform unsupervised topic mining and clustering, from which the associations between cases and topics are summarized and a topic–case binary matrix is constructed.

(2) Inductive mapping based on HFACS. The output of the previous stage is reviewed by experts, and topics are consolidated according to the HFACS framework. Scattered topics are mapped and coded onto HFACS levels, so that algorithmic labels are transformed into theory-driven condition variables, on the basis of which a case-HFACS condition binary matrix is constructed.

(3) Causal pathway identification based on CNA. The binary matrix is converted into configurational input, and CNA is applied to identify, under theoretical constraints, the condition combinations associated with severe outcomes. Cross-level causal configuration pathways are then extracted, yielding a configurational explanation of accident occurrence.

2.2. BERTopic-Based Mining of Accident-Causation Themes

2.2.1. Data Collection and Preprocessing

The research data are mainly drawn from official investigation reports published on the websites of the Ministry of Emergency Management of China and provincial emergency management departments. These reports are typically based on detailed investigations conducted by government-organized expert panels, and their accident causes sections provide substantial descriptions and analyses of technical defects, human errors, and management failures, thereby offering an authoritative and standardized corpus for this study [40]. To obtain as many relevant texts as possible, both targeted search and expanded search strategies are used during data collection. Systematic searches are first conducted on official platforms, followed by supplementary screening through the website of the China Chemical Safety Association and search engines such as Baidu and Google. In total, 121 investigation reports on chemical accidents classified as serious or above are collected, covering severe accident cases from major provinces and various types of chemical enterprises in China over the past decade. Because the dataset consists exclusively of Chinese accident investigation reports, the empirical configurations identified in this study should be interpreted within comparable institutional and industrial frameworks. Differences in accident-reporting standards, enforcement practices, industrial structure, and safety management systems may affect the occurrence and documentation of causal conditions. The classification of accident severity in China is shown in Figure 3.

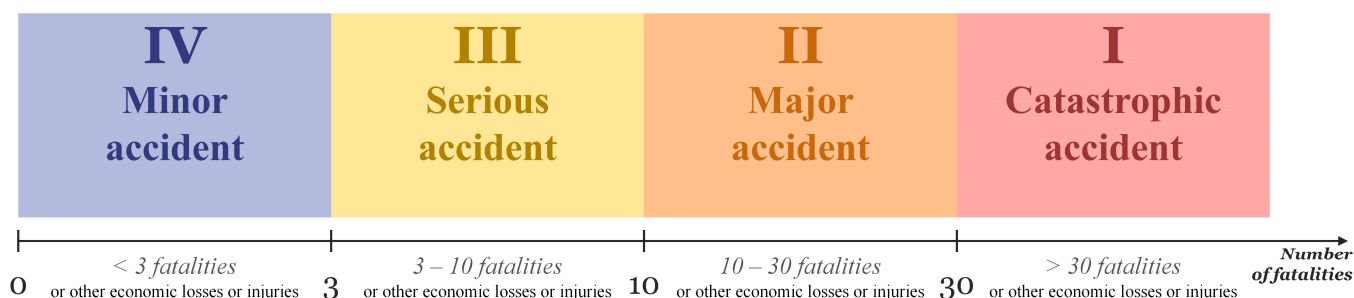


Figure 3. Accident categories.

In line with the research objective, the complete text of the accident causes section is extracted from each investigation report. Referring to the analytical boundary commonly adopted in organizational analysis in the chemical industry, descriptions related to external government causes are removed from the original texts to ensure logical consistency in the

subsequent analysis [32]. At the same time, information on accident timing, geographic location, and losses is recorded, resulting in a text corpus of several hundred thousand Chinese characters.

The thematic orientation of accident reports is usually complex, and direct topic modeling may lead to semantic ambiguity and feature sparsity [14]. Following the recommendations of Grootendorst for short-text topic modeling and the application of Sun et al. to scientific texts, this study adopts a sentence segmentation strategy to divide the original texts into semantic units [17,41]. Periods, semicolons, exclamation marks, question marks, and line breaks are used as the main delimiters to split compound paragraphs into relatively independent semantic fragments. After segmentation, the original line-number index of each sentence is recorded simultaneously to facilitate later traceability. The segmented text set is denoted as $R = \{r_1, r_2, \dots, r_{n(R)}\}$, where $n(R)$ is the total number of segmented sentences.

Building on sentence segmentation, the HanLP toolkit is further used for word segmentation and text cleaning to preserve, as far as possible, the terminological characteristics of the chemical safety domain. During the text denoising stage, this study uses the HIT stopword list as the baseline and integrates the “Common Chemical Names” lexicon from the Sogou Lexicon to construct a domain-specific stopword dictionary. At the same time, in view of the writing characteristics of accident investigation reports, the cleaning procedure specifically removes high-frequency function words in official documents, expressions of abstract concepts, institutional titles, and adverbs that make no substantive semantic contribution. In addition, the “Chemical Safety Production Lexicon” from the Sogou Lexicon is incorporated as a dictionary to preserve the integrity of technical terms. Finally, valid sentences with a text length of no less than five characters are retained as the final corpus and imported into BERTopic for topic modeling. The cleaned text set is denoted as $S = \{s_1, s_2, \dots, s_{n(S)}\}$, where $n(S)$ denotes the total number of sentences after cleaning.

2.2.2. BERTopic Topic Modeling

Semantic vectorization uses the BAAI/bge-base-zh-v1.5 sentence embedding model, which performs well on Chinese semantic benchmarks [42]. During vectorization, the original sentences are retained for semantic embedding, and the text set R is used as the input so that the encoder can capture contextual information as fully as possible. At the stage that links dimensionality reduction and clustering, UMAP is used to reduce the dimensionality of the high-dimensional BGE embeddings. Since UMAP involves stochastic initialization, this study fixes `random_state = 42` to improve the reproducibility of the dimensionality reduction results [43]. HDBSCAN is then used to perform unsupervised clustering on the reduced embeddings. This algorithm can identify density-based structures without prespecifying the number of clusters and can assign low-density samples to noise [44]. Given that accident-causation texts contain many low-frequency but substantively meaningful niche topics, this study fixes the minimum number of samples in HDBSCAN at `min_samples = 2` to reduce the risk that potentially valid small clusters are excessively assigned to noise.

To improve the discriminative power and interpretability of topic keyword representations, this study develops a staged keyword optimization procedure. Based on the text set S , CountVectorizer is used to generate the term-frequency matrix, with `min_df` fixed at 5 and `max_df` dynamically constrained according to the size of the corpus. BERTopic then applies class-based TF-IDF (c-TF-IDF) to combine the texts within each topic cluster into a class document and to weight the representativeness of terms at the topic level. Finally, Maximal Marginal Relevance (MMR) is used to re-rank candidate keywords for each topic

in terms of diversity [45], thereby improving the distinctiveness and information richness of the keywords while preserving topic representativeness.

2.2.3. Model Evaluation Metrics and Parameters

To comprehensively evaluate the topic modeling results, this study conducts quantitative assessment from two complementary dimensions, namely topic quality and clustering structure. First, topic coherence, denoted as C_V , is used to measure the semantic consistency among keywords within the same topic, and it is calculated on the basis of the top 10 keywords for each topic. A higher C_V value indicates stronger co-occurrence consistency and semantic relatedness among the keywords within a topic, and usually suggests better topic interpretability [46].

Second, topic diversity, denoted as TD , is used to measure the degree of distinction across topics. A higher TD value indicates less semantic overlap and stronger differentiation among topics. Suppose that there are T topics in total, and that the top k keywords are retained for each topic. Let $|U|$ denote the number of unique terms in the union of keywords across all topics. Topic diversity is then defined as [17]:

$$TD = \frac{|U|}{T \times k} \quad (1)$$

$T \times k$ represents the total number of keywords. A higher value indicates that fewer keywords are shared across topics, clearer topic boundaries, and stronger overall differentiation. Finally, the noise ratio is recorded as an auxiliary robustness indicator to reflect the proportion of samples classified as noise during clustering.

Since the modeling results of BERTopic are relatively sensitive to the number of neighbors in UMAP (`n_neighbors`) and the minimum cluster size in HDBSCAN (`min_cluster_size`), this study conducts a grid search over these two key parameters to ensure the quality of topic extraction. Using the cached BGE embeddings as the common input, a complete BERTopic model is trained under different combinations of `min_cluster_size` and `n_neighbors`, and metrics including topic coherence, topic diversity, number of topics, and noise ratio are recorded. In the process of selecting the optimal parameters, a coarse search is first conducted with a step size of 5 to identify the parameter range with relatively good performance, after which a fine-grained search is carried out within that range with a step size of 2.

To balance semantic consistency within topics and differentiation across topics, this study further defines the product of C_V and TD as a composite evaluation score, denoted as S_B , that is:

$$S_B = C_v \times TD \quad (2)$$

To avoid excessive fragmentation in the clustering results, a noise ratio of no more than 20% is used as a screening criterion to exclude parameter combinations that do not meet this requirement. The remaining combinations are then ranked in descending order according to the S_B score, and the top five parameter sets are selected for subsequent expert evaluation.

Although C_V and TD can assess topic quality statistically, they do not directly reflect the semantic suitability of the topics for chemical accident causation [20]. This study therefore adds an expert evaluation stage to combine statistical results with domain knowledge. The expert panel is provided with summary materials for the top five candidate models, including the top keywords, representative sentences, and topic distribution statistics for each topic. The experts then rate the candidate models on a 7-point Likert scale in terms of semantic interpretability, coherence, and distinctiveness. The mean expert score is used to determine the final parameter setting of the BERTopic model in this study.

2.3. Mapping of Causal Variables Based on the HFACS Framework

2.3.1. Selection of the HFACS Framework

The topic set T provides data-driven causal semantic clusters, but these topics remain parallel and discrete labels. This study therefore introduces HFACS as the theoretical framework, which helps bring domain knowledge into the theoretical calibration [25]. The main purpose of this study is not to reconstruct a complete system control structure, but to further transform the text-based topic results into discrete hierarchical condition variables for configurational analysis. For this purpose, HFACS is better suited to serve as a middle-range theoretical bridge between BERTopic topic results and CNA condition modeling, because it provides a clear hierarchical structure, relatively stable second-level categories, and a mature basis for industry-specific revision.

Given the distinctive characteristics of the chemical industry, this study mainly adopts the HFACS-CSMEs framework revised for the Chinese chemical industry by Wang et al. [29] and supplements it with the application suggestions of Theophilus et al. on HFACS in the oil, gas, and chemical sectors [28], to determine the final hierarchical correspondence.

2.3.2. Expert Classification and Its Validation

This study uses Expert Classification to complete the hierarchical mapping of BERTopic-derived topics to HFACS. Expert classification can jointly consider the semantic information of topic keywords, representative accident sentences, and the contextual features of the chemical industry. It is therefore more suitable for mapping tasks that are guided by an explicit theoretical framework [31]. For this purpose, this study invites five experts with long-term experience in industrial engineering, chemical engineering, and management science and engineering to support the subsequent analysis.

The classification procedure is as follows. First, the expert panel is provided with the full information for the BERTopic topics, together with a reference table of HFACS levels. The experts then independently judge the HFACS level and subcategory code corresponding to each topic. To improve the robustness of the results, the mapping is completed independently by the experts under blind review conditions.

To quantitatively assess inter-expert classification agreement, this study uses Fleiss' Kappa coefficient, which has been widely used in HFACS-related research [47,48]. The formula is as follows [47]:

$$\kappa = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e} \quad (3)$$

$\bar{P}$ is the mean proportion of pairwise agreement among raters across all topics, and $\bar{P}_e$ is the expected proportion of agreement under chance conditions. A κ value closer to 1 indicates higher agreement among the experts in topic classification, whereas a lower κ value suggests that the topic boundaries may be more ambiguous and require further review. Based on this indicator, this study evaluates the HFACS classification results and then determines the final classification according to the majority opinion of the experts. In this study, the expert classification yields a Fleiss' κ of 0.7118, and the corresponding analysis is presented in detail in the Results section. For topics on which disagreements remain, a consensus discussion is conducted until a final agreement is reached.

2.3.3. Construction of the Case–HFACS Binary Matrix

After the HFACS classification of topics is completed, this study further converts the previously constructed case–topic binary matrix into a case–HFACS dimension binary matrix for subsequent analysis. If an accident case contains at least one topic that belongs to a given second-level HFACS category, the corresponding dimension for that case is coded as 1; otherwise, it is coded as 0. Through this aggregation, the original topic variables

are compressed into condition variables with theoretical meaning. At the same time, this procedure effectively reduces dimensional fragmentation and better meets the requirement of an interpretable logical structure in the subsequent analysis.

On this basis, this study aims to examine which factors are more likely to lead to High-Fatality (HF) outcomes in chemical accidents classified as serious or above.

$$HF_i = 1 \text{ if } d_i > 3; \text{ else } HF_i = 0 \quad (4)$$

where d_i denotes the number of deaths in case i . Under the Chinese accident classification system (*Regulation on the Reporting, Investigation, and Handling of Production Safety Accidents*), three deaths constitute the statutory boundary between an “ordinary” accident and a “serious” accident. However, the criteria for “serious-or-above” classification are not based on fatalities alone: an accident can also reach this tier through other criteria. The sample therefore contains heterogeneity around this threshold: some cases result in only a handful of fatalities, whereas others lead to larger losses of life. Setting $d_i > 3$ as the high-fatality cutoff separates accidents whose fatalities meaningfully exceed the statutory boundary from those at or below it. To verify that the configurational findings are not artefacts of this single calibration, two stricter cutoffs ($d_i > 4$ and $d_i > 5$) are further examined in the threshold sensitivity analysis.

2.4. Identification of Causal Pathways Based on CNA

This study further uses Coincidence Analysis (CNA) to identify the condition combinations associated with high-fatality outcomes in serious-or-above chemical accidents and to examine their cross-level configurational relationships. CNA emphasizes the configurational nature of conditions and hierarchical causal chains [49], and is therefore suitable for analyzing complex structures in accident causation.

The logical core of CNA lies in identifying INUS conditions in the formation of an outcome, namely an insufficient but necessary part of an unnecessary but sufficient condition [50]. Its typical Boolean form can be expressed as follows:

$$(X_1 * X_2) + (X_3 * X_4) \leftrightarrow Y \quad (5)$$

In this expression, $*$ denotes the conjunctive operator, which refers to the joint presence of multiple conditions within the same configuration. The symbol $+$ denotes the disjunctive operator, which refers to multiple alternative configurational pathways leading to the same outcome [50].

The input data for CNA are derived from the case-HFACS dimension binary matrix constructed in the preceding section. Considering the need to balance solution complexity and interpretability in CNA, this study restricts the analytical variables to the Level II subcategories of the HFACS framework. This choice of granularity follows the practice of Patterson and Shappell in empirical HFACS research. It preserves the discriminative capacity of causal variables in cross-level transmission analysis while avoiding overly complex configurations and insufficient co-occurrence cases caused by excessive variable subdivision [51]. This study conducts the analysis using the *cna* package in R 4.6.0. Since all variables in the analytical matrix obtained in Section 2.3 are binary, this study applies crisp-set CNA. The *configTable()* function is used to convert the original case–variable matrix into a compressed configuration table. This table merges cases with identical condition–value combinations and records their frequencies.

The core parameters of CNA include two solution-quality thresholds, namely consistency (con.) and coverage (cov.). Their formulas are as follows [52]:

$$\text{Consistency } (X \rightarrow Y) = |X \cap Y| / |X| \quad (6)$$

$$\text{Coverage } (X \rightarrow Y) = |X \cap Y| / |Y| \quad (7)$$

$|X \cap Y|$ denotes the number of cases in which both the condition combination X and the outcome Y are present. $|X|$ and $|Y|$ denote the total number of cases satisfying the condition combination X and the total number of cases in which the outcome Y occurs, respectively. Following the empirical settings used by Baumgartner [52], Whitaker et al. [37], and related exploratory studies, this study fixes the consistency threshold at $\text{con} = 0.75$ and sets the minimum coverage requirement for MSCs at 0.10. (Most CNA studies do not impose a lower bound on MSC coverage. We adopt one here to mitigate the influence of isolated cases.) This setting retains causal pathways with substantive explanatory value while reducing the risk of spurious relationships caused by accidental co-occurrence.

To balance theoretical constraints with empirical search, this study runs both a theoretically constrained model and an unconstrained model. The theoretically constrained model is specified according to the hierarchical logic of HFACS. In the presentation of the final solutions, pathways that pass the consistency test in the hierarchically constrained CNA model ($\text{con} \geq 0.75$) and are robustly reproduced in the unconstrained CNA model are treated as the main empirical patterns reported in this study.

In addition, two benchmark analyses were conducted on the same HFACS conditions. First, logistic regression was used to estimate the net association between each condition and the high-fatality outcome. Second, csQCA was used as a configurational benchmark to compare truth-table minimization results with CNA-derived minimally sufficient configurations.

3. Results

3.1. Descriptive Statistics

The final text corpus includes 2 catastrophic accidents, 11 major accidents, and 108 serious accidents. In terms of accident type, explosions, deflagrations, and related accidents account for 65 cases, representing 53.7% of the sample, and constitute the most common accident type. Poisoning and asphyxiation accidents account for 43 cases, representing 35.5%, ranking second. In terms of geographic distribution, Shandong, Hebei, and Hubei have the highest accident frequencies, with 15, 11, and 10 cases, respectively. The annual and monthly distributions of accidents are relatively balanced, with an average of approximately 11 accidents per year. Accidents occur relatively more frequently in June and April. Additional descriptive statistics are shown in Figure 4.

3.2. BERTopic Modeling Results

3.2.1. Model Parameter Selection

Following the two-stage search strategy described in Section 2.2.3, this study first conducts a coarse-grained scan of `min_cluster_size` and `n_neighbors` with a step size of 5. The results show that parameter settings with relatively high composite scores S_B are concentrated within `min_cluster_size` $\in [10, 20]$ and `n_neighbors` $\in [30, 40]$. A fine-grained search is then conducted within this range with a step size of 2. The number of topics, noise ratio, topic coherence C_V , and topic diversity TD are recorded simultaneously. Parameter settings with excessively high noise ratios are excluded, and the remaining settings are ranked in descending order according to S_B . The top five candidate parameter settings are then submitted to the expert panel for independent evaluation. The experts rate each

setting on a 7-point Likert scale in terms of semantic interpretability, topic coherence, and topic distinctiveness. The topic evaluation metrics are presented in Table 1.

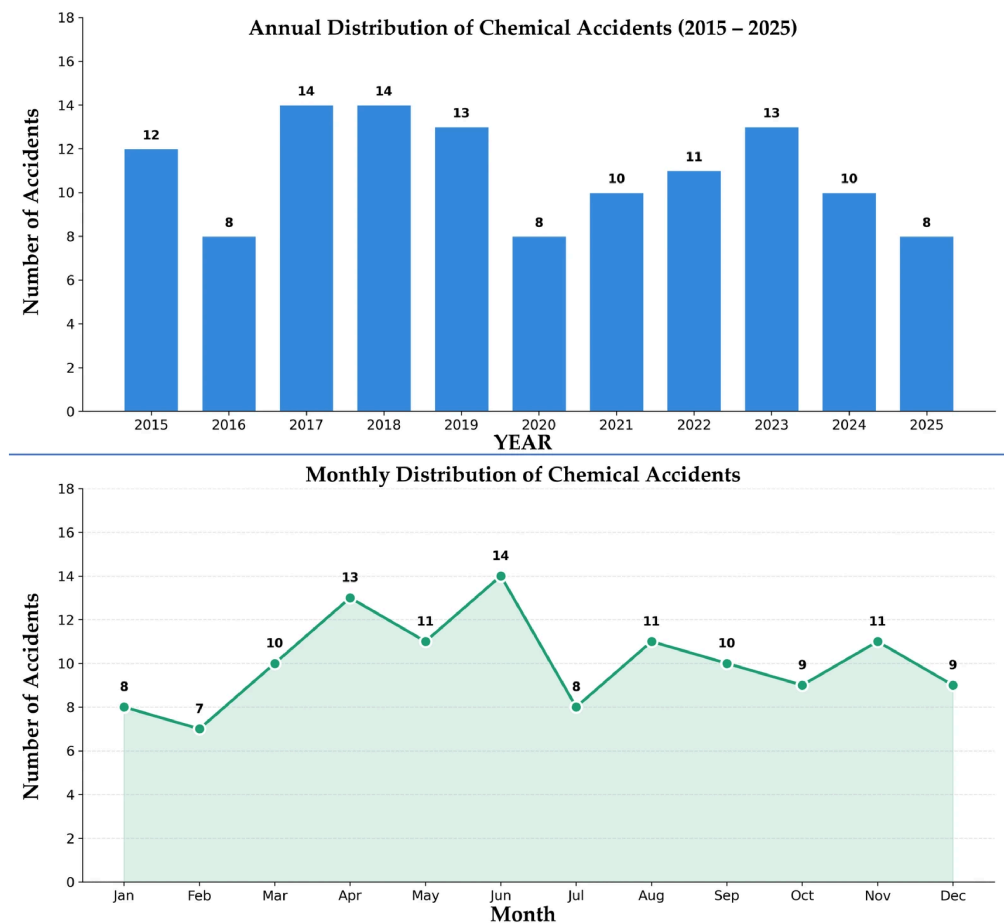


Figure 4. Temporal distribution of serious-or-above chemical accidents in China from 2015 to 2025.

Table 1. Top five candidate BERTopic parameter settings and quantitative evaluation metrics.

Rank	min_cluster_Size	n_neighbors	Number of Topics	Noise Ratio (%)	C_V	TD	S_B
1	17	32	17	19.00%	0.558	0.806	0.450
2	10	32	40	13.60%	0.485	0.785	0.381
3	15	32	25	17.40%	0.487	0.780	0.380
4	13	35	33	18.80%	0.492	0.770	0.379
5	10	37	17	13.00%	0.455	0.829	0.377

In the blind expert evaluation, the third-ranked parameter setting receives the highest overall score across the three semantic dimensions. Compared with the other parameter settings, its topic results show better semantic interpretability and an appropriate topic granularity. It is therefore selected as the final parameter setting. Under this setting, 25 topics are identified, with a noise ratio of 17.40%, $C_V = 0.487$, and $TD = 0.780$. This setting provides a reasonable balance among topic granularity, semantic boundary clarity, and domain interpretability.

3.2.2. Topic Identification Results

Under the final parameter setting, the BERTopic model identifies 25 non-noise topics. As shown in Table 2, these topics are named according to their representative sentences and

top keywords. The topic-naming procedure was conducted through collective discussion within the research team. For each topic, the team jointly reviewed the representative keywords, representative sentences, and corresponding original accident-report contexts, and then discussed an appropriate topic name on a topic-by-topic basis. For topic names with disagreements, the team traced the relevant sentences back to the original accident reports, discussed their meanings in depth, and then determined the final topic names. The 25 topics cover multiple aspects, including safety training and education, responsibility implementation, work management, equipment and facilities, project construction, emergency response, and risk prevention and control. Together, they provide a relatively systematic representation of the core problem structure in the accident-cause texts.

Table 2. BERTopic-derived accident-causation topics and representative keywords.

Topic ID	Topic Name	Top Representative Keywords
0	Inadequate Job-Specific Operational Training	operation, training, job position, safety management, system, work safety, operating procedures, equipment/unit
1	Erroneous Emergency Response Decision-Making	explosion, production, poisoning, space, material, reaction, decomposition, inhalation, formation
2	Improper Hot Work Procedures	hot work, ignition source, approval, maintenance, on-site, safety measures, work permit, processing, application
3	Perfunctory Safety Education and Training	training, safety education, safety training, formalism, work safety, education and training, assessment
4	Deficient Safety Responsibility System	safety responsibility system, deficiency, work safety, safety responsibility, all staff, safety management duties
5	Inadequate Confined-Space Hazard Identification	space, hazardous factors, confined space, preventive measures, identification, control, hazard, on-site
6	Ineffective Safety Supervision	absence, defect, merely nominal, supervision, work safety management, process, safety management, change
7	Improper Emergency Rescue	rescue, escalation, casualties, accident escalation, emergency rescue, any person, adoption, protective measures, wearing
8	Illegal Hazardous Chemical Storage	hazardous chemicals, illegal, storage, operation, acquisition, safety facilities, filing, permit
9	Deficient On-Site Equipment Safety Management	on-site safety management, maintenance, special equipment, on-site, installation, upkeep, safety equipment, absence
10	Inadequate Emergency Planning and Drills	drills, emergency rescue, emergency plan, rescue, on-site, emergency rescue plan, accident emergency response
11	Deficient Construction Review and Qualification Control	construction, gatekeeping, review, qualification, decision-making, installation, design, completion, acceptance
12	Improper Safety Qualification Review	availability, qualification, signing, safety production conditions, safety management agreement, construction, work safety
13	Noncompliant Project Planning and Construction	construction, permit, planning, engineering, fire protection, commencement, acquisition, license
14	Illegal Storage of Waste and Scrap Materials	waste, scrap, environment, storage, solids, pollution, protection
15	Improper Production System Operation	cutoff, dry burning, pressure, pipeline, piping, system, operation, mis-entry
16	Deficient Work Permit Management	maintenance, permit, plan, handling, work permit, construction, repair, oversight failure

Table 2. *Cont.*

Topic ID	Topic Name	Top Representative Keywords
17	Failure to Identify and Rectify Hazards	inspection, remediation, safety hazards, grading, rectification, formalism, control, accident hazards
18	Failure to Correct Employee Violations	supervision and correction, rules and regulations, operating procedures, employees, work safety rules and regulations, safe operating procedures
19	Inadequate Safety Review during Trial Operation	trial production, trial run, review, equipment/unit, testing, availability, product, safety production conditions
20	Failure of Responsible Personnel to Fulfill Safety Responsibilities	director, work safety, actual, safety responsibility system, legal representative, principal person in charge
21	Noncompliant Engineering Quality	construction, violation, illegality, engineering quality, conduct, design, engineering, relocation
22	Poor Safety Awareness and Disregard for Regulations	weak, safety awareness, disregard, law, leakage, work safety awareness, work safety, principal person in charge
23	Unauthorized Process Changes	unauthorized, change, process, operation, simple, control, function, automation
24	Inadequate Risk Identification and Control Implementation	identification, control, prevention and control, prevention, inspection, circulation, evaluation, petrochemical

In terms of topic content, these 25 topics are not isolated or scattered. One group of topics mainly reflects deficiencies in safety training, education, and safety awareness, including Topics 0 and 3. Together, these topics indicate that, at the time of accident occurrence, frontline personnel have insufficient capacity to identify hazards and perform standardized operations.

Another group of topics mainly reflects failures in responsibility implementation and routine management, such as Topics 4, 6, 17, and 24. These topics are more closely related to problems in system implementation, routine supervision, hazard rectification, and responsibility transmission within enterprises. This suggests that, in many accident cases, risks do not simply arise from a single moment at the accident site, but are closely associated with long-standing management laxity, inadequate inspection, and insufficient implementation of risk prevention and control measures.

A third group of topics mainly concerns defects in equipment, facilities, and site conditions, such as Topics 9, 14, and 21. From different perspectives, these topics reflect deficiencies in equipment integrity, facility configuration, material storage, and engineering quality at accident sites. This indicates that some accident risks originate from unsafe production conditions themselves.

In addition, another group of topics focuses on emergency response, mainly including Topics 1, 7, and 10. These topics suggest that, in some cases, the expansion of accident consequences does not arise entirely from the initial event itself, but is closely related to misjudgment after the event, improper remedial actions, and inadequate emergency preparedness.

The topic-similarity heatmap of the 25 topics is shown in Figure 5. Given the high homogeneity of the underlying text corpus, the topics show a medium-to-high level of similarity, with the lower bound of similarity remaining at approximately 0.60. This pattern is common in topic modeling within a single domain. Nevertheless, identifiable boundaries remain among different topics, and no large-scale high overlap is observed. This indicates that the final model achieves a reasonable level of topic differentiation. This result also suggests that the 25 topics identified in this study are not the product of

mechanical fragmentation, but form a relatively stable semantic structure around several core causal modules.

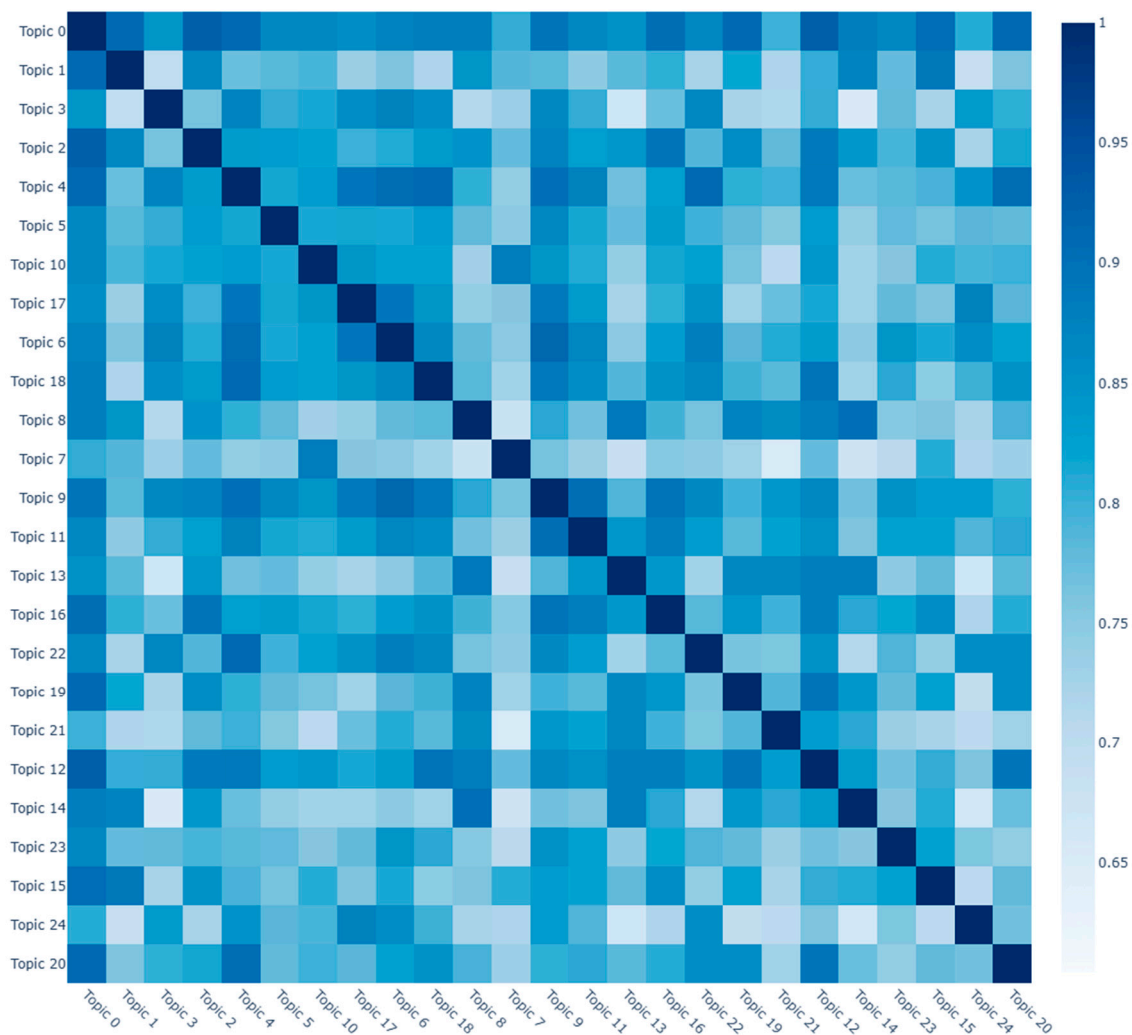


Figure 5. Similarity heatmap of BERTopic-derived accident-causation topics.

On this basis, following the procedure described in Section 2.2.4, this study traces the segmented sentences back to their corresponding accident cases according to the original line-number indices. A 121×25 case–topic binary matrix is then constructed, with the 25 topics as columns and the 121 cases as rows. This matrix provides a data-driven representation of the causal topic combinations involved in each accident case and serves as the original data for the subsequent analysis.

3.3. Results of HFACS Framework Mapping

3.3.1. Validation of Expert Classification

Based on the blind evaluation results of the five experts, this study classifies the identified topics into the hierarchical categories of HFACS. The final classification results show that $\kappa = 0.7118$ ($z = 28.19$, $p < 0.01$). According to the criteria proposed by Landis and Koch, a κ value within the range of 0.61–0.80 indicates substantial agreement. This result indicates that, under independent blind evaluation conditions, the expert panel forms a relatively stable consensus on the HFACS hierarchical assignment of the BERTopic-derived topics. For a small number of topics with classification disagreements, such as Topic 20 and Topic 10, the expert panel reaches a final consensus after collective review, based on the HFACS classification table and the hierarchical orientation reflected in the topic

keywords. For disputed topics that lack direct corresponding entries in the table, such as Topic 21, the expert panel further refers to their higher-level categories and completes the final classification after discussion.

3.3.2. Mapping Structure of the HFACS Hierarchy

The 25 BERTopic-derived topics are systematically mapped onto the four first-level categories and eight second-level subcategories of HFACS. The overall mapping structure is shown in Figure 6.

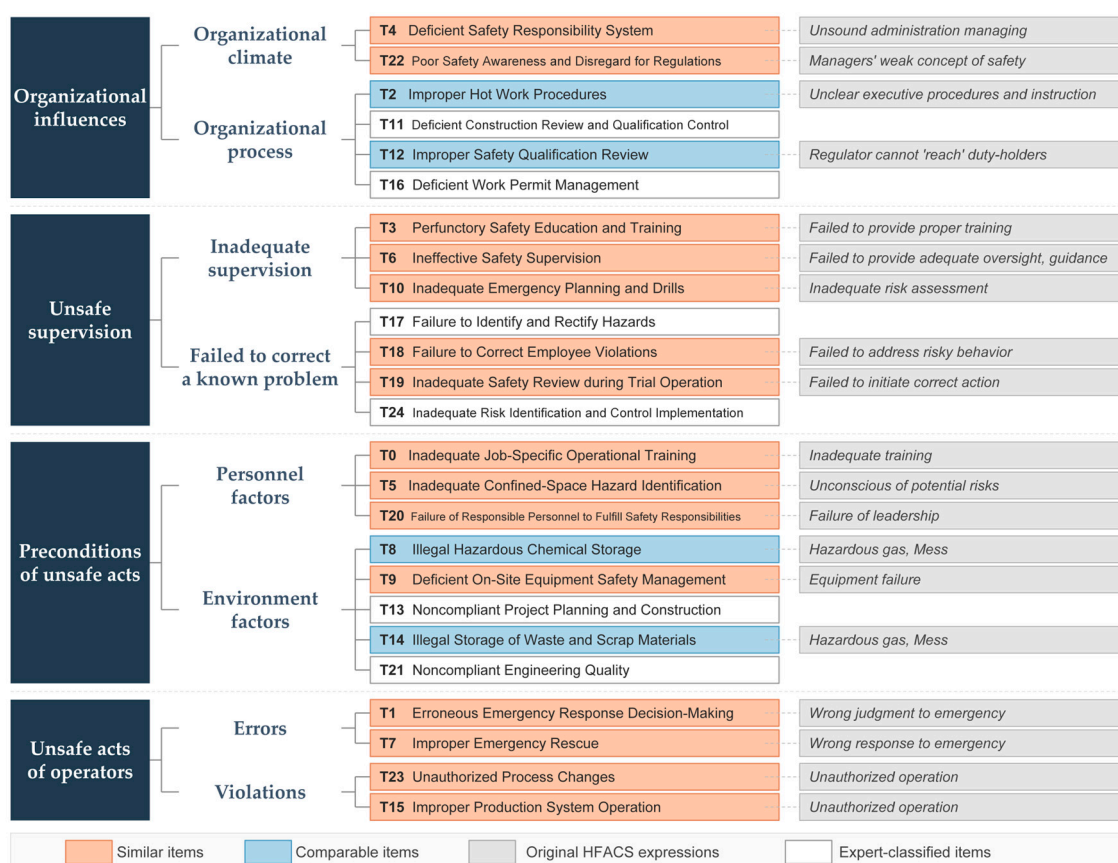


Figure 6. Mapping of BERTopic-derived accident-causation topics to the HFACS hierarchy.

Most topics find direct or approximate counterparts within existing HFACS subcategories, indicating that the BERTopic-derived results have good structural compatibility with the HFACS framework [29]. Three types of correspondence emerge during the mapping process. “Similar items” indicate that the topic semantics are highly similar to the typical expressions of HFACS subcategories and can be directly aligned. “Comparable items” indicate that the topics do not correspond one-to-one with the original HFACS expressions but can be assigned by analogy based on their semantic core. “Expert-classified items” indicate that the topic content extends beyond the classical examples listed under HFACS subcategories and is theoretically classified by experts in light of the chemical industry context.

In the correspondence analysis with the ten second-level categories of HFACS-CSMEs, resource management under Organizational influences and task assignment and management under Unsafe supervision do not form suitable corresponding items [29]. For the former, resource management accounts for the lowest proportion in the revised HFACS-CSMEs study, and its related manifestations are mostly low-frequency or absent items, suggesting that its distribution in the chemical industry context may be relatively limited.

For the latter, task assignment and management mainly involves directive information on internal task allocation within enterprises. Such information is often insufficiently disclosed in the cause-summary sections of accident investigation reports, and therefore does not show sufficiently concentrated features in the topic extraction results of this study.

All four first-level categories receive clear topic support. The four topics under Unsafe acts of operators are notably fewer than those under the upstream levels. This distribution is generally consistent with the hierarchical structure of HFACS, which extends from terminal unsafe acts to upstream latent conditions. Organizational influences and Unsafe supervision together cover 13 topics, accounting for more than half of all topics. This indicates that the causal focus is more concentrated in upstream factors such as organizational systems and supervisory management.

3.3.3. Construction of the Case–HFACS Dimension Binary Matrix

After the classification of topics into the HFACS hierarchy is completed, this study further compresses the 121×25 case–topic binary matrix into a 121×9 case–HFACS second-level dimension binary matrix according to the rules described in Section 2.3.3. Combined with the high-fatality variable HF, a binary analytical matrix containing 121 cases, 8 condition variables, and 1 outcome variable is finally formed for subsequent CNA modeling. To facilitate matrix computation, model output presentation, and subsequent reporting of results, this study assigns unified codes to the HFACS second-level dimensions and the outcome variable, as shown in Table 3.

Table 3. Coding and distribution of HFACS condition variables and the high-fatality outcome.

Code	Variable	Value = 1, <i>n</i> (%)	Value = 0
OI1	Organizational climate	71 (58.68)	50
OI2	Organizational process	66 (54.55)	55
US1	Inadequate supervision	90 (74.38)	31
US2	Failed to correct a known problem	70 (57.85)	51
PUA1	Personnel factors	91 (75.21)	30
PUA2	Environment factors	70 (57.85)	51
UAO1	Errors	107 (88.43)	14
UAO2	Violations	33 (27.27)	88
HF	High-fatality outcome	60 (49.59)	61

Among these variables, UAO1 has the highest frequency, appearing in 107 cases, which may be related to the greater visibility of direct behavioral errors in accident investigation reports. It is followed by PUA1 (75.21%) and US1 (74.38%). By contrast, UAO2 has the lowest frequency, appearing in only 33 cases. For the outcome variable, 60 cases are coded as $HF = 1$, accounting for 49.59% of the sample, which is broadly comparable to the 61 cases coded as $HF = 0$.

3.4. Identification of Causal Configurational Pathways Based on CNA

HFACS hierarchically constrained cs-CNA and unconstrained cs-CNA are run separately under the threshold of $\text{con} \geq 0.75$. This section reports the pathways reproduced across the constrained and unconstrained CNA models as the main empirical patterns.

3.4.1. Minimally Sufficient Condition (MSCs) Combinations for High-Fatality Outcomes

When solving for $HF = 1$, six three-condition minimally sufficient condition combinations are identified. Considering that a balance between explanatory power and parsimony is more useful for revealing cross-level causal structures [53], this section mainly reports

the MSCs obtained under the default settings of the R package. The results are shown in Table 4.

Table 4. MSCs for high-fatality outcomes.

No.	Configuration	$n(X \cap HF)/n(X)$	Con.	Cov.	Exhaustiveness	Faithfulness
M1	OI1 * US2 * PUA2 $\rightarrow$ HF	17/22	0.773	0.283	1.000	0.938
M2	OI1 * $\sim$ OI2 * PUA2 $\rightarrow$ HF	12/13	0.923	0.200	1.000	0.938
M3	OI1 * US2 * $\sim$ PUA1 $\rightarrow$ HF	6/7	0.857	0.100	1.000	0.938
M4	$\sim$ OI2 * $\sim$ US1 * US2 $\rightarrow$ HF	6/7	0.857	0.100	1.000	0.938
M5	OI1 * $\sim$ OI2 * $\sim$ US1 $\rightarrow$ HF	6/8	0.750	0.100	1.000	0.938
M6	$\sim$ US1 * US2 * PUA2 $\rightarrow$ HF	6/8	0.750	0.100	0.933	0.933

Note: * indicates logical conjunction (AND), whereas $\sim$ indicates the negation of a condition (i.e., the corresponding variable is coded as 0).

In terms of condition composition, all six MSCs are three-condition combinations, and all of them are cross-level configurations that include at least one condition from Organizational influences or Unsafe supervision. No configuration is composed only of conditions from Preconditions of unsafe acts or Unsafe acts of operators. Among them, OI1, namely organizational climate, appears in four pathways, M1, M2, M3, and M5, and serves as the main causal endpoint within Organizational influences. US2, namely failed to correct a known problem, also appears in four pathways, M1, M3, M4, and M6, and serves as the main causal endpoint within Unsafe supervision.

At the level of Preconditions of unsafe acts, PUA2, namely environment factors, appears in positive form in three MSCs, M1, M2, and M6. By contrast, PUA1, namely personnel factors, appears only in negated form, $\sim$ PUA1, in M3. This indicates that, in the causal configurations leading to high-fatality outcomes, physical working environment conditions have greater configurational relevance than individual personnel conditions.

At the pathway level, M1 has the highest coverage (cov = 0.283), making it the pathway that explains the broadest set of high-fatality cases. M2 has the highest consistency (con = 0.923), suggesting that when this condition combination is present, a high-fatality outcome is more likely to occur.

Figure 7 reports the threshold sensitivity results. Of the 121 cases, the high-fatality group (HF = 1) comprises 60 cases at deaths > 3, 38 at deaths > 4, and 26 at deaths > 5; the HF = 0 counts are 61, 83, and 95, respectively. When the fatality threshold was raised to deaths > 4 and deaths > 5, the exact Boolean expressions changed, but several condition families remained recurrent, especially OI1, US2, and PUA2. This suggests that the findings are not merely an artifact of the baseline fatality cutoff. Under the stricter threshold, both M1 and M2 re-emerge with additional conditions appended, indicating that their core condition combinations remain present but become embedded in larger configurations.

The logistic regression benchmark is shown in Table 5, identifies PUA2 as the only significant positive predictor of high-fatality accidents. This result supports the relevance of environmental preconditions in explaining severe accident consequences. The QCA benchmark further confirms that the same HFACS matrix contains configurational structure; however, the complex solution generated four equivalent models with many long conjunctions, most involving six to eight conditions, making effective interpretation difficult.

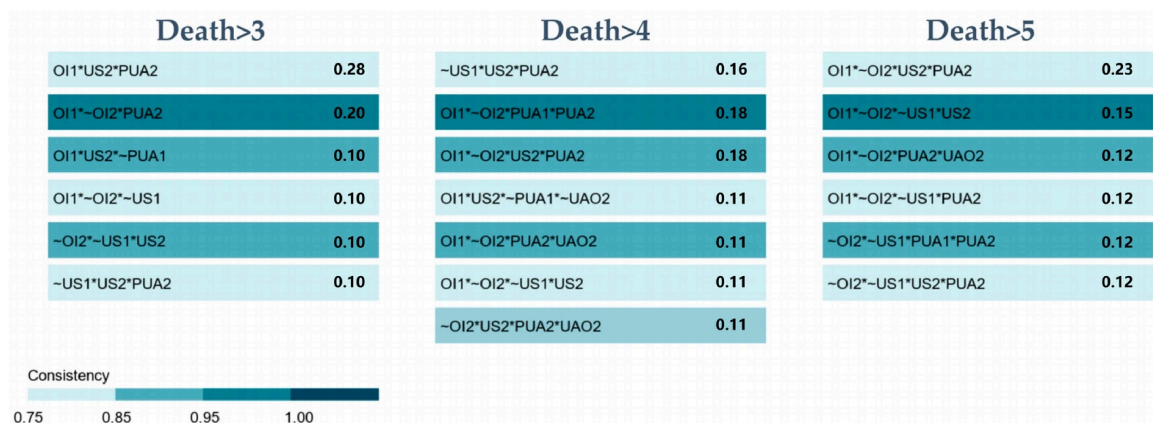


Figure 7. Threshold Robustness of CNA-Derived Sufficient Configurations. * indicates logical conjunction (AND), whereas ~ indicates the negation of a condition.

Table 5. Logistic regression results for high-fatality accidents (death > 3).

Predictor	Beta	SE	OR	95% CI for OR	<i>p</i>
OI1	0.418	0.411	1.52	0.68–3.40	0.309
OI2	−0.768	0.421	0.46	0.20–1.06	0.068
US1	0.202	0.452	1.22	0.50–2.97	0.655
US2	0.731	0.407	2.08	0.94–4.61	0.073
PUA1	0.266	0.483	1.3	0.51–3.36	0.582
PUA2	1.327	0.445	3.77	1.58–9.01	0.003
UAO1	−0.174	0.657	0.84	0.23–3.05	0.791
UAO2	−0.74	0.466	0.48	0.19–1.19	0.113

3.4.2. Causal Transmission Pathways

The CNA search further identifies vertical transmission relationships among the second-level dimensions. The high-fitting pathways are mainly concentrated among Unsafe supervision, Preconditions of unsafe acts, and Unsafe acts of operators. Table 6 presents all pathways with coverage greater than 0.5.

The results show that US1, US2, and OI1 have strong directional relationships with PUA1 (T1, T4, and T5), indicating that conditions from the supervision level and the organizational influence level converge on the same precondition. PUA1 is further linked to UAO1, namely operational errors, with relatively high consistency (con = 0.868) (T2). Meanwhile, the US level also shows a surface pathway that bypasses the precondition level and directly transmits to UAO1 (T3). Among the six condition variables across the three upstream levels, namely OI, US, and PUA, five are identified as upstream conditions linked to UAO1 in the CNA results.

Table 6. Causal transmission pathways identified by CNA.

No.	Path	Con.	Cov.
T1	US1 → PUA1	0.778	0.769
T2	PUA1 → UAO1	0.868	0.738
T3	US1 → UAO1	0.867	0.729
T4	US2 → PUA1	0.800	0.615
T5	OI1 → PUA1	0.775	0.604
T6	PUA2 → UAO1	0.571	0.570
T7	OI1 → UAO1	0.859	0.570
T8	US2 → UAO1	0.843	0.551
T9	OI2 → UAO1	0.818	0.505

When the MSCs are synthesized into Atomic Solution Formulas (ASFs) and Complex Solution Formulas (CSFs), the maximum faithfulness value is 0.644, which is relatively low. This reflects the coexistence of multiple alternative causal pathways in serious-or-above chemical accidents, where the same severe outcome can be triggered by several equivalent and non-nested condition combinations [53]. Therefore, following existing applied studies, this study does not use a single ASF or CSF as the final unit of interpretation.

4. Discussion

4.1. Configurational Pathways of High-Fatality Outcomes

The MSC-level patterns indicate that, in this sample, high-fatality outcomes in serious-or-above chemical accidents are not well explained by terminal behavioral factors alone, but are more consistently associated with cross-level coupling involving organizational and supervisory conditions. The threshold robustness analysis in Figure 7 supports this interpretation at the condition-family level: although the exact Boolean expressions change after raising the fatality cutoff, OI1 and US2 remain recurrent across alternative thresholds. Organizational climate and failed corrective supervision can therefore be interpreted as recurrent upstream condition families rather than artifacts of a single outcome definition.

It should be noted that, in CNA, when a negated condition enters a minimally sufficient pathway, this indicates that the absence of the corresponding condition is a non-redundant component of that configuration. Its meaning must therefore be interpreted in relation to the overall configuration rather than in isolation according to its literal meaning [50].

M1 has the highest coverage and is the most representative dominant causal pathway for explaining high-fatality accident cases. It combines deficient organizational climate at the macro level, failure to correct known hazards at the meso level, and adverse physical environmental conditions at the micro level. In this context, severe outcomes are consistent with the accumulation of latent structural deficiencies in the upstream system, rather than by isolated individual errors. The 2019 “3·21” catastrophic explosion accident at Tianjiayi Chemical in Xiangshui, Jiangsu, provides a typical empirical case for this configurational pathway. The investigation report shows that the enterprise had long maintained a distorted development philosophy that prioritized economic benefits over safety. After environmental protection and emergency management departments at multiple levels repeatedly pointed out hazards related to the storage of nitration waste, the enterprise management did not implement substantive rectification and instead deliberately concealed the problem by falsifying records. As a result, a large amount of highly explosive nitration waste accumulated over a long period in an old solid-waste warehouse without temperature-control facilities. The terminal event, namely spontaneous combustion caused by heat accumulation in the waste, can therefore be interpreted as being consistent with accumulated systemic risk, which eventually led to catastrophic consequences.

M2 has the highest consistency. The presence of $\sim$ OI2 reveals a problematic organizational feature. $\sim$ OI2 means that the enterprise shows no obvious deficiencies at the level of organizational processes, whereas the simultaneous presence of OI1 indicates that problems already exist in the underlying organizational climate. In this context of separation between formal processes and actual organizational conditions, high-fatality outcomes are triggered once adverse working environmental conditions (PUA2), are also present. Traditional regulatory approaches that focus on the completeness of institutional documents and formal compliance in assessments have little discriminatory power for this type of accident. They may even allow hazards to accumulate when inspections remain at the level of superficial compliance. Effective review should therefore reach the substantive operating conditions of the enterprise, including the organizational climate.

US2 appears in M1, M3, M4, and M6, suggesting that hazard accumulation is a recurrent precondition in most pathways leading to high-fatality outcomes. This structure indicates that the cause of severe consequences sometimes lies not in the failure to identify problems, but in the failure to resolve them, which allows hazards to accumulate. M3 further reinforces this interpretation. Even when the individual condition of frontline personnel shows no obvious abnormality ($\sim$ PUA1), high-fatality outcomes are more frequently observed when deficient organizational climate and failed corrective supervision coexist. In other words, in serious-or-above chemical accidents, accident escalation may result from the coupled failure of organization and supervision, rather than from frontline operators themselves.

A comparison between US1 and US2 further clarifies their difference. In the three MSCs of M4, M5, and M6, US1 appears in negated form ($\sim$ US1). Severe consequences may still occur when US2 is present or when other adverse conditions coexist. This means that US1 and US2 are not synonymous. The former can be regarded as a process indicator, reflecting whether supervisory actions are carried out, and can be directly observed through external review. The latter can be regarded as an outcome indicator, reflecting whether the supervisory loop actually closes the identified risk exposure. It is difficult to capture through superficial compliance checks and plays a more important role in the identified pathways.

At the level of Preconditions of unsafe acts, PUA2 appears in three pathways, namely M1, M2, and M6. This is related to the inherent hazards of chemical production environments and also suggests that improving the working environment may be more effective in reducing severe consequences than merely emphasizing personnel conditions. The benchmark analyses also provide convergent support for this interpretation.

4.2. Discussion of Hierarchical Transmission

From the perspective of cross-level transmission, PUA1 receives multiple upstream inputs and transmits to UAO1, thus playing a hub role in this sample. However, it does not enter any MSC leading to high-fatality outcomes in positive form. It should therefore be viewed more as a channel through which upstream organizational and supervisory failures affect the operational level. It externalizes latent conditions, but it is not the key switch that determines whether an accident escalates into a high-fatality outcome. The factors that truly drive consequence escalation remain the upstream deficiencies, which also indirectly supports the HFACS assumption regarding unsafe acts.

UAO1 mainly acts as a terminal receiving endpoint. At the transmission level, it is the common convergence point of multiple upstream pathways, but it does not enter any configuration in positive form among the MSCs for high-fatality outcomes. Together with the logistic benchmark in Table 5, where neither UAO1 nor UAO2 is identified as a significant positive predictor, these results indicate that operational errors are more likely to be behavioral manifestations at the accident endpoint after upstream organizational and supervisory failures are transmitted across levels, rather than core sufficient conditions for explaining severe consequences. It is worth noting that UAO1 appears in 107 of the 121 cases, accounting for 88.43% of the sample, making it the most frequent variable. However, this frequency advantage does not translate into core explanatory power in the configurational sense. The high frequency of terminal operational errors more likely reflects their high visibility in accident investigation narratives, rather than necessarily indicating that they have the strongest independent causal status. It should be acknowledged that the near-universal presence of UAO1 may limit its capacity to enter any discriminating MSC. Nevertheless, the logistic benchmark in Table 5, in which UAO1 is likewise non-significant, provides independent support for the same interpretation.

In an accident system involving hierarchical transmission, judging causal importance solely on the basis of frequency can easily underestimate the structural role of upstream organizational and supervisory factors and overestimate the independent explanatory power of terminal behavioral factors.

4.3. Governance Implications

First, at the organizational level, the findings suggest that governance should give greater attention not only to whether safety systems exist, but also to whether these systems are embedded in enterprise decision-making processes. In practice, some enterprises may have relatively complete safety management documents, yet safety requirements can still be weakened under production pressure, schedule pressure, and cost constraints. Therefore, enterprises should incorporate safety performance indicators into management assessment systems, such as the reduction in recurring hazards, the implementation of accountability for overdue rectification, and the effectiveness of closed-loop management for major hazards. In chemical process settings, these indicators can be tied to the closure of PHA/HAZOP recommendations, management-of-change follow-up actions, overdue maintenance items, and repeated abnormal-operation records.

Second, the focus of supervision is to strengthen the closed loop of problem rectification. The effectiveness of supervision does not depend on the number of inspections, but on whether known problems are removed from the worksite in a timely manner. Therefore, enterprises should establish mechanisms such as escalation for overdue rectification, re-verification and closure of major hazards, and mandatory review of recurring problems. For hazards involving hazardous chemical storage, reaction units, hot work, and confined-space operations, risk elimination should not be replaced merely by stating that the issue has been included in a rectification plan. Compared with inspection coverage, indicators such as the average closure cycle of major hazards, the overdue rectification rate, and the recurrence rate of repeated problems can better reflect the actual control capacity of the supervision level. For regulatory agencies, follow-up inspections should verify field evidence of corrective actions, especially for repeated findings, overdue rectification items, and deficiencies in safety-critical barriers such as alarms, interlocks, emergency isolation, and ventilation.

Third, prevention at the level of work preconditions should focus on weakening the amplifying effect of environmental conditions on accident consequences. The high-risk nature of chemical accidents means that the production site itself is a critical component of safety. Therefore, governance resources should be prioritized for improving the production environment, such as reducing the amount of high-risk materials, implementing zoned storage, and strengthening online monitoring. Compared with measures that rely solely on personnel vigilance, these measures are usually more stable and easier to verify.

Finally, the focus at the behavioral level should not be limited to training and punishment, but should pay greater attention to outcome indicators, such as the effectiveness of training and whether operational error rates decrease. Repeated operational errors should not be simply attributed to individual negligence. Instead, the organizational, supervisory, and environmental conditions behind them should be traced in reverse.

5. Conclusions

This study proposed a Topic–Hierarchy Coincidence Analysis (T-H CNA) framework to identify configurational pathways associated with high-fatality outcomes in serious-or-above chemical accidents. By integrating BERTopic, HFACS, and CNA, the framework transformed unstructured accident-cause narratives from 121 Chinese investigation reports into hierarchical causal conditions and then identified cross-level sufficient configurations.

The results show that 25 causal semantic themes can be consolidated into eight HFACS categories, from which six three-condition minimally sufficient configurations for high-fatality outcomes were obtained.

The key finding is that the sufficient configurations for high-fatality outcomes are not formed by terminal operational errors alone. Instead, the identified configurations commonly include upstream organizational or supervisory conditions. Deficient organizational climate and failure to correct known problems repeatedly appear as upstream endpoints in the identified configurations and remain stable under stricter fatality thresholds. By contrast, operational errors appear in 88.43% of cases but do not enter any sufficient configuration, indicating that high frequency does not necessarily imply configurational sufficiency. These results support a shift in process safety governance from treating frontline errors as isolated causes toward identifying and interrupting upstream condition combinations that are sufficient for high-fatality escalation.

This study also has several limitations. First, the sample of 121 cases remains moderate at the granularity of HFACS second-level dimensions. The maximum faithfulness at the ASF and CSF levels is only 0.644. Therefore, this study treats the robustly reproduced MSCs and transmission pathways as the main findings and adopts a conservative interpretation of the strength of evidence. Second, the binarization of variables leads to the loss of continuous information on the number of fatalities and direct economic losses. Future research may further extend this analysis by using multi-value or fuzzy-set CNA. Third, reliance on a single corpus source may introduce narrative bias. Future studies may integrate in-depth news reports, law enforcement records, litigation materials, insurance claim data, and investigation reports from other regions to improve the external validity of the findings. In addition, the T-H CNA framework provides room for incorporating other topic extraction methods and different theoretical hierarchical frameworks, thereby supporting further exploration of the boundaries of safety science research.

Author Contributions: Conceptualization, J.Z. and Y.L.; methodology, J.Z.; software, J.Z., Y.L. and Y.W.; formal analysis, J.Z., Y.L. and Y.W.; investigation, B.F.; resources, J.Z. and Y.W.; data curation, J.Z. and Y.L.; writing—original draft preparation, J.Z. and Y.L.; writing—review and editing, J.Z. and B.F.; visualization, Y.W.; supervision, B.F.; project administration, B.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study B.F. are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: The authors thank the providers of the publicly available accident investigation reports used in this study. Gemini 3.1 Pro was used only for language polishing and grammar refinement during manuscript preparation. It was not used for research design, data collection, data analysis, interpretation of results, or the generation of scientific content. The author reviewed and edited all AI-assisted outputs and takes full responsibility for the content of the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CNA	Coincidence Analysis
HFACS	Human Factors Analysis and Classification System
BERTopic	Bidirectional Encoder Representations from Transformers Topic Modeling
MSC	Minimally Sufficient Condition

References

- Wang, H.; Wei, L.; Wang, K.; Wang, R.; Duo, Y.; Yang, G. Current status, challenges, and future pathways of chemical industrial park safety in China. *J. Loss Prev. Process Ind.* **2024**, *87*, 105233. [\[CrossRef\]](#)
- Wang, H.; Wei, L.; Wang, K.; Duo, Y.; Chen, C.; Zhang, S.; Su, M.; Zeng, T. Exploring human factors of major chemical accidents in China: Evidence from 160 accidents during 2011–2022. *J. Loss Prev. Process Ind.* **2024**, *89*, 105279. [\[CrossRef\]](#)
- Zhou, F.; Liu, X.; Wang, F. Size–frequency distribution characteristic of fatalities due to workplace accidents and industry dependency. *Mathematics* **2025**, *13*, 2021. [\[CrossRef\]](#)
- Fouzi, N.F.R.; Aziz, H.A.; Yaakub, N. Systematic review of chemical safety and chemical security risk management approach. *Process Saf. Environ. Prot.* **2024**, *183*, 676–686. [\[CrossRef\]](#)
- De Leo, F.; Elia, V.; Gnoni, M.G.; Tornese, F. Integrating Safety-I and Safety-II approaches in near miss management: A critical analysis. *Sustainability* **2023**, *15*, 2130. [\[CrossRef\]](#)
- Wang, Y.; Zheng, C.; Wang, L. Causal factors and configuration paths for chemical occupational accidents: Grounded theory and fsQCA analysis of 154 cases in China. *Front. Public Health* **2026**, *13*, 1713532. [\[CrossRef\]](#) [\[PubMed\]](#)
- Chen, C.; Reniers, G. Chemical industry in China: The current status, safety problems, and pathways for future sustainable development. *Saf. Sci.* **2020**, *128*, 104741. [\[CrossRef\]](#)
- Chen, C.; Reniers, G.; Khakzad, N. Cost-benefit management of intentional domino effects in chemical industrial areas. *Process Saf. Environ. Prot.* **2020**, *134*, 392–405. [\[CrossRef\]](#)
- Tauseef, S.M.; Abbasi, T.; Abbasi, S.A. Development of a new chemical process-industry accident database to assist in past accident analysis. *J. Loss Prev. Process Ind.* **2011**, *24*, 426–431. [\[CrossRef\]](#)
- Reason, J. *Managing the Risks of Organizational Accidents*; Ashgate: Aldershot, UK, 1997.
- Leveson, N.G. *Engineering a Safer World: Systems Thinking Applied to Safety*; MIT Press: Cambridge, MA, USA, 2011.
- Wu, Y.; Fu, G.; Han, M.; Jia, Q.; Lyu, Q.; Wang, Y.; Wu, Z. Comparison of the theoretical elements and application characteristics of STAMP, FRAM, and 24Model: A major hazardous chemical explosion accident. *J. Loss Prev. Process Ind.* **2022**, *80*, 104880. [\[CrossRef\]](#)
- Marucci-Wellman, H.R.; Corns, H.L.; Lehto, M.R. Classifying injury narratives of large administrative databases for surveillance—A practical approach combining machine learning ensembles and human review. *Accid. Anal. Prev.* **2017**, *98*, 359–371. [\[CrossRef\]](#) [\[PubMed\]](#)
- Single, J.I.; Schmidt, J.; Denecke, J. Knowledge acquisition from chemical accident databases using an ontology-based method and natural language processing. *Saf. Sci.* **2020**, *129*, 104747. [\[CrossRef\]](#)
- Niu, Y.; Fan, Y.; Ju, X. Critical review on data-driven approaches for learning from accidents: Comparative analysis and future research. *Saf. Sci.* **2024**, *171*, 106381. [\[CrossRef\]](#)
- Macêdo, J.B.; Moura, M.D.C.; Aichele, D.; Lins, I.D. Identification of risk features using text mining and BERT-based models: Application to an oil refinery. *Process Saf. Environ. Prot.* **2022**, *158*, 382–399. [\[CrossRef\]](#)
- Grootendorst, M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv* **2022**, arXiv:2203.05794. [\[CrossRef\]](#)
- Cao, Y.; Qu, Z.; Wu, S.; Chen, Y.; Skitmore, M.; Ma, X.; Wang, J. Analyzing OSHA construction accident reports using BERTopic topic modeling for thematic insights. *Buildings* **2025**, *16*, 10. [\[CrossRef\]](#)
- Egger, R.; Yu, J. A topic modeling comparison between LDA, NMF, Top2Vec, and BERTopic to demystify Twitter posts. *Front. Sociol.* **2022**, *7*, 886498. [\[CrossRef\]](#) [\[PubMed\]](#)
- Nanyonga, A.; Joiner, K.; Turhan, U.; Wild, G. Semantic topic modeling of aviation safety reports: A comparative analysis using BERTopic and PLSA. *Aerospace* **2025**, *12*, 551. [\[CrossRef\]](#)
- Zhang, Y.; Wang, W.; Mi, L.; Sun, G.; Qiao, L.; Tao, M.; Wang, L. Uncovering the organizational vulnerability toward construction project accidents: BERTopic-based text mining analysis. *J. Constr. Eng. Manag.* **2025**, *151*, 04025139. [\[CrossRef\]](#)
- Qiang, J.; Qian, Z.; Li, Y.; Yuan, Y.; Wu, X. Short text topic modeling techniques, applications, and performance: A survey. *IEEE Trans. Knowl. Data Eng.* **2022**, *34*, 1427–1445. [\[CrossRef\]](#)
- He, Y.; Wang, B.; Cui, L. Configuration analysis of coal mine accident causation: Uncovering pathways, substitutions, and provincial heterogeneity in China. *Reliab. Eng. Syst. Saf.* **2026**, *269*, 112122. [\[CrossRef\]](#)
- Wiegmann, D.A.; Shappell, S.A. Human error analysis of commercial aviation accidents: Application of the Human Factors Analysis and Classification System (HFACS). *Aviat. Space Environ. Med.* **2001**, *72*, 1006–1016. [\[PubMed\]](#)
- Madigan, R.; Golightly, D.; Madders, R. Application of Human Factors Analysis and Classification System (HFACS) to UK rail safety of the line incidents. *Accid. Anal. Prev.* **2016**, *97*, 122–131. [\[CrossRef\]](#) [\[PubMed\]](#)
- Patterson, J.M.; Shappell, S.A. Operator error and system deficiencies: Analysis of 508 mining incidents and accidents from Queensland, Australia using HFACS. *Accid. Anal. Prev.* **2010**, *42*, 1379–1385. [\[CrossRef\]](#) [\[PubMed\]](#)
- Xia, N.; Zou, P.X.W.; Liu, X.; Wang, X.; Zhu, R. A hybrid BN-HFACS model for predicting safety performance in construction projects. *Saf. Sci.* **2018**, *101*, 332–343. [\[CrossRef\]](#)

28. Theophilus, S.C.; Esenowo, V.N.; Arewa, A.O.; Ifelebuegu, A.O.; Nnadi, E.O.; Mbanaso, F.U. Human factors analysis and classification system for the oil and gas industry (HFACS-OGI). *Reliab. Eng. Syst. Saf.* **2017**, *167*, 168–176. [\[CrossRef\]](#)
29. Wang, J.; Fan, Y.; Gao, Y. Revising HFACS for SMEs in the chemical industry: HFACS-CSMEs. *J. Loss Prev. Process Ind.* **2020**, *65*, 104138. [\[CrossRef\]](#)
30. Zarei, E.; Khan, F.; Abbassi, R. A dynamic human-factor risk model to analyze safety in sociotechnical systems. *Process Saf. Environ. Prot.* **2022**, *164*, 479–498. [\[CrossRef\]](#)
31. Jia, Q.; Fu, G.; Xie, X.; Xue, Y.; Hu, S. Enhancing accident cause analysis through text classification and accident causation theory: A case study of coal mine gas explosion accidents. *Process Saf. Environ. Prot.* **2024**, *185*, 989–1002. [\[CrossRef\]](#)
32. Wang, J.; Fan, Y.; Niu, Y. Routes to failure: Analysis of chemical accidents using the HFACS. *J. Loss Prev. Process Ind.* **2022**, *75*, 104695. [\[CrossRef\]](#)
33. Meng, Q.; Liu, W.; Zhang, Z.; Chong, H.Y. Multiple configurational effects and analyses on construction workers' unsafe behaviors: A qualitative approach of fuzzy-set qualitative comparative analysis. *Int. J. Occup. Saf. Ergon.* **2025**, *32*, 356–373. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Swiatczak, M.D. Different algorithms, different models. *Qual. Quant.* **2022**, *56*, 1913–1937. [\[CrossRef\]](#)
35. Baumgartner, M.; Ambühl, M. Causal modeling with multi-value and fuzzy-set Coincidence Analysis. *Political Sci. Res. Methods* **2020**, *8*, 526–542. [\[CrossRef\]](#)
36. Baumgartner, M. Detecting causal chains in small-N data. *Field Methods* **2013**, *25*, 3–24. [\[CrossRef\]](#)
37. Whitaker, R.G.; Sperber, N.; Baumgartner, M.; Thiem, A.; Cragun, D.; Damschroder, L.; Miech, E.J.; Slade, A.; Birken, S. Coincidence analysis: A new method for causal inference in implementation science. *Implement. Sci.* **2020**, *15*, 108. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Huang, R.; Remer, S.L.; Knapp, L.K.; Cohen, M.E.; Rosen, M.A.; Hall, B.L.; Wick, E.C.; Ko, C.Y. Hospital configurations leading to successful implementation of enhanced recovery programs. *Ann. Surg.* **2025**. *Epub ahead of print.* [\[CrossRef\]](#) [\[PubMed\]](#)
39. Akerstrom, M.; Severin, J.; Miech, E.J.; Wikström, E.; Roczniowska, M. Reducing sickness absence among public-sector healthcare employees: The difference-making roles of managerial and employee participation. *Int. Arch. Occup. Environ. Health* **2024**, *97*, 341–351. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Jing, S.; Liu, X.; Gong, X.; Tang, Y.; Xiong, G.; Liu, S.; Xiang, S.; Bi, R. Correlation analysis and text classification of chemical accident cases based on word embedding. *Process Saf. Environ. Prot.* **2022**, *158*, 698–710. [\[CrossRef\]](#)
41. Sun, B.; Lu, S.; Pan, X.; Han, L.; Tu, J. Multi-stakeholders cooperative decision analysis in underground logistics system development: A methodology combining topic modeling and system dynamics. *Comput. Ind. Eng.* **2025**, *210*, 111548. [\[CrossRef\]](#)
42. Xiao, S.; Liu, Z.; Zhang, P.; Muennighoff, N. C-Pack: Packaged resources to advance general Chinese embedding. *arXiv* **2023**, arXiv:2309.07597. [\[CrossRef\]](#)
43. Li, C.; Tian, Q.; Gao, L.; Liu, H. Public feedback analysis on multi-stage emergency management policies using BERTopic-SKEP integrated model. *Sci. Rep.* **2025**, *15*, 684. [\[CrossRef\]](#) [\[PubMed\]](#)
44. McInnes, L.; Healy, J.; Astels, S. hdbscan: Hierarchical density based clustering. *J. Open Source Softw.* **2017**, *2*, 205. [\[CrossRef\]](#)
45. Carbonell, J.; Goldstein, J. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Melbourne, Australia*; ACM: New York, NY, USA, 1998; pp. 335–336. [\[CrossRef\]](#)
46. Röder, M.; Both, A.; Hinneburg, A. Exploring the space of topic coherence measures. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining (WSDM'15), Shanghai, China*; ACM: New York, NY, USA, 2015; pp. 399–408. [\[CrossRef\]](#)
47. Fleiss, J.L. Measuring nominal scale agreement among many raters. *Psychol. Bull.* **1971**, *76*, 378–382. [\[CrossRef\]](#)
48. Ergai, A.; Cohen, T.; Sharp, J.; Wiegmann, D.; Gramopadhye, A.; Shappell, S. Assessment of the Human Factors Analysis and Classification System (HFACS): Intra-rater and inter-rater reliability. *Saf. Sci.* **2016**, *82*, 393–398. [\[CrossRef\]](#)
49. Misangyi, V.F.; Greckhamer, T.; Furnari, S.; Fiss, P.C.; Crilly, D.; Aguilera, R. Embracing causal complexity: The emergence of a neo-configurational perspective. *J. Manag.* **2017**, *43*, 255–282. [\[CrossRef\]](#)
50. Baumgartner, M.; Falk, C. Boolean difference-making: A modern regularity theory of causation. *Br. J. Philos. Sci.* **2023**, *74*, 171–197. [\[CrossRef\]](#)
51. Lenné, M.G.; Salmon, P.M.; Liu, C.C.; Trotter, M. A systems approach to accident causation in mining: An application of the HFACS method. *Accid. Anal. Prev.* **2012**, *48*, 111–117. [\[CrossRef\]](#) [\[PubMed\]](#)

52. Baumgartner, M.; Ambühl, M. Optimizing consistency and coverage in configurational causal modeling. *Sociol. Methods Res.* **2023**, *52*, 1288–1320. [[CrossRef](#)]
53. Parkkinen, V.-P.; Baumgartner, M. Robustness and model selection in configurational causal modeling. *Sociol. Methods Res.* **2023**, *52*, 176–208. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.