aerospace

MDPI

Article

Knowledge Graph-Based Causal Analysis of Aviation Accidents: A Hybrid Approach Integrating Retrieval-Augmented Generation and Prompt Engineering

Xinyu Xiang, Xiyuan Chen * and Jianzhong Yang

Department of Safety Science and Engineering, Civil Aviation University of China, Tianjin 300300, China; 2023091091@cauc.edu.cn (X.X.); jzyang2004@126.com (J.Y.)
* Correspondence: xy-chen@cauc.edu.cn

Abstract

The causal analysis of historical aviation accidents documented in investigation reports is important for the design, manufacture, operation, and maintenance of aircraft. However, given that most accident data are unstructured or semi-structured, identifying and extracting causal information remain labor intensive and inefficient. This gap is further deepened by tasks, such as system identification from component information, that require extensive domain-specific knowledge. In addition, there is a consequential demand for causation pattern analysis across multiple accidents and the extraction of critical causation chains. To bridge those gaps, this study proposes an aviation accident causation and relation analysis framework that integrates prompt engineering with a retrieval-augmented generation approach. A total of 343 real-world accident reports from the NTSB were analyzed to extract causation factors and their interrelations. An innovative causation classification schema was also developed to cluster the extracted causations. The clustering accuracy for the four main causation categories—Human, Aircraft, Environment, and Organization—reached 0.958, 0.865, 0.979, and 0.903, respectively. Based on the clustering results, a causation knowledge graph for aviation accidents was constructed, and by designing a set of safety evaluation indicators, "pilot—decision error" and "landing gear system malfunction" are identified as high-risk causations. For each high-risk causation, critical combinations of causation chains are identified and "Aircraft operator—policy or procedural deficiency/pilot—procedural violation/Runway contamination → pilot—decision error → pilot procedural violation/32 landing gear/57 wings" was identified as the critical causation combinations for "pilot—decision error". Finally, safety recommendations for organizations and personnel were proposed based on the analysis results, which offer practical guidance for aviation risk prevention and mitigation. The proposed approach demonstrates the potential of combining AI techniques with domain knowledge to achieve scalable, data-driven causation analysis and strengthen proactive safety decision-making in aviation.

Keywords: aviation accident; causation analysis; airworthiness; LLM; RAG; prompt engineering; knowledge graph

Check for updates

Academic Editor: Miroslav Kelemen

Received: 3 December 2025
Revised: 21 December 2025
Accepted: 22 December 2025
Published: 24 December 2025

Copyright: © 2025 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) license.

1. Introduction

In the past few years, the aviation industry has been recovering from the influence of the pandemic. The global passenger traffic steadily increased from 1.8 billion (2020) to

4.2 billion (2023), and the number of flight departures for scheduled commercial operations continued to rise from 22.47 million (2020) to 35.25 million (2023) according to International Civil Aviation Organization 2024 Safety Report [1]. Although 2023 was considered the safest year with only one fatal accident and 72 fatalities, the number of fatal accidents sharply increased to 9 in 2024, which caused 238 fatalities [2]. Given the great number of passengers and aircrafts in the air and the catastrophic consequence of aircraft accidents, safety management studies in the aviation industry are crucial. The historical accident investigation reports are valuable data sources for aircraft causal analysis. These reports are composed of thorough accident information, insight investigation, and reliable conclusions. However, these reports are mostly unstructured or semi-structured in a lengthy passage, and identifying the key information requires strong domain knowledge. A precise and automated method to extract and analyze the causation and relationship in accident investigation reports is lacking.

Traditional natural language processing (NLP) and machine learning (ML) methods, namely, support vector machine [3], long-short term memory (LSTM) [4], convolution neural networks [5] and recurrent neural networks (RNNs) [6], are widely applied for accident analysis and prevention in the aviation industry. However, such methods often rely heavily on extensive data labeling and feature engineering, which require significant human intervention. Moreover, their effectiveness is limited by the quality and quantity of available data. Most of these models are highly specialized and lack flexibility and robustness, which limit their capability to satisfy real-world operational demands.

The emergence of large language models (LLMs) has introduced a new paradigm for problem solving in the field of NLP. By encoding vast amounts of pretraining corpora into model parameters through sophisticated pretraining procedures, LLMs inherently possess strong capabilities in reading comprehension, text generation, and information extraction. They have demonstrated outstanding performance across various NLP tasks and offer new tools and methods for causation analysis in aviation accidents. However, hallucination and lack of specific domain knowledge are two major obstacles limiting the broad application of LLMs. Two mainstream training-free approaches are used to address these limitations: prompt engineering (PE) and retrieval-augmented generation (RAG). By designing prompts using techniques such as few-shot learning or chain of thought (CoT) reasoning [7], LLMs can generate highly structured and concise content. RAG has been shown to enhance the performance of LLMs by incorporating external knowledge segments that are relevant to the query of the user [8]. The relationship analysis between causations across different accidents is also important. Knowledge graphs (KGs), as an advanced tool for large-scale data analysis, excel in representing complex relationships and have been widely applied in accident investigation and analysis [9,10].

Despite the achievements of existing techniques and methods in the field of aviation accident analysis, current research still encounters several limitations: (1) The low utilization rate of textual content from accident reports. Many studies focus only on a single category of contributing factors, such as human or aircraft-related causes, while the impact of environmental and organizational factors is ignored. Moreover, they often fail to consider the interactions among these factors. (2) Inadequate capacity to process domain knowledge intensive tasks. When dealing with intensive-domain knowledge required tasks, such as system identification from component information, current methods have shortcomings in either efficiency or accuracy. (3) The neglect of causation relationships across different accidents. Effective approaches to uncover the relations between causations across different accidents are lacking.

To address these issues, this study proposes a comprehensive causation analysis method for aviation accidents that integrates RAG with KGs. LLMs and the RAG frame-

work are employed to perform the initial extraction of accident causes and their relationships from investigation reports. A top-down taxonomy is designed to categorize causes into four major types and 126 subtypes, as well as to divide causal relationships into three categories. This taxonomy is embedded into the prompt templates through PE, which allows the LLMs to further cluster the initially extracted causes and relationships. The clustering results are then batch-injected into the KG using Cypher queries via Python (3.10) scripts. Several risk-related parameters are defined to enable the quantitative analysis of causes and relationships. Finally, targeted safety recommendations are provided based on the analytical outcomes.

2. Literature Review

Research on aviation accidents typically involves two main stages: key information extraction and subsequent analysis. This section first reviews the evolution of key information extraction methods in this domain, which range from early statistical models and rule-based approaches to more recent semi-automated techniques that incorporate NLP. Among these methods, the application of LLMs and related technologies, such as RAG, represents a promising direction for aviation accident analysis. Thereafter, post-analysis methods for deep causation and correlation exploration, including KGs, are briefly introduced. These technological advancements have significantly enhanced the efficiency and quality of aviation accident analysis.

2.1. Key Information Extraction Methods in Aviation Accident Analysis

Key information extraction is essentially a named entity recognition (NER) task, and its development can be categorized based on the evolution of pretrained models (PMs).

2.1.1. Traditional NER Methods

NER methods prior to PMs can broadly be categorized into ML and early deep learning approaches. Although the deep learning methods of that period lacked contextualized understanding and relied on static word embeddings, they still led to significant advancements. Brooker evaluated the effectiveness of Bayesian belief networks for aviation risk prediction, which highlighted their limitations due to expert judgment biases and hidden common causes [11]. Moreover, resilience engineering was demonstrated to offer a more reliable path for improving aviation safety. Zhang and Wang applied TF-IDF to classify the risk features of aviation safety systems [12]. Zhang et al. proposed a sequential deep learning approach using word embedding and LSTM to predict aviation accident outcomes from NTSB investigation report texts and event sequences [13]. V. de Vries combined text mining with random forest (RF) classifier algorithms, which validated its effectiveness for the classification of occurrence reports [14]. Wang et al. proposed a novel entity extraction method for aviation safety reports by combining knowledge-enhanced embeddings, domain-specific dictionaries, and bilinear attention networks [15]. Olive and Basora proposed an autoencoding neural network-based framework to detect and characterize anomalies in aircraft trajectory data [16]. Liu and Yang proposed an integrated evaluation approach by first applying four methods—two traditional ML algorithms (HMM and CRF) and two deep learning algorithms (Bi-LSTM and Bi-LSTM-CRF)—to perform entity recognition. Then, they combined the results using an ensemble model with a voting mechanism to determine the validity of each identified entity [9]. Although traditional NER models, which are typically built on RNN/LSTM combined with CRF, have achieved notable success, they are constrained by poor parallelism, limited capability to capture long-range dependencies [17], and a heavy reliance on data annotation and manual feature engineering.

2.1.2. LLM-Related Methods

Transformer introduced a novel encoder–decoder architecture along with the attention mechanism [18], which formed the foundation for subsequent PMs that effectively address the aforementioned limitations. As one of the most representative PMs, Bidirectional Encoder Representations from Transformers (BERT) introduced a powerful pretraining method for deep bidirectional language representations, which enabled state-of-the-art performance on multiple NLP tasks through simple fine-tuning without task-specific architectures [19]. Other variants of BERT, such as the Robustly Optimized BERT Pretraining Approach [20] and Sentence-BERT [21], have significantly advanced the development of the NLP field. One representative application is the Aviation-BERT-NER system proposed by Chandra et al., it can identify named entities critical for aviation safety analysis and has achieved strong results on real-world NTSB narratives [22]. Although these new methods have significantly progressed in NER and aviation accident analysis, they also introduce challenges such as high computational resource consumption, highly complex and time-consuming model training, and great demands for large-scale and high-quality training data. These limitations hinder their further application in the field of aviation accident analysis. Unlike BERT, which is primarily designed for encoding tasks such as classification or NER, most LLMs adopt a larger-scale architecture with billions of parameters and are trained on massive corpora. Thus, LLMs have superior generative capabilities and conversational fluency. In addition to language generation, LLMs excel at few-shot [23] and zero-shot [7] learning. As a result, they can be applied to new tasks with minimal or no task-specific training.

The application of LLMs generally follows three main paradigms: PE, RAG, and fine-tuning. Fine-tuning methods such as LoRA [24] adapt LLMs to more complex downstream tasks by further training. However, they require substantial computational resources and their effectiveness heavily depends on the quality of the training dataset. By contrast, PE and RAG do not involve modifying model parameters; both aim to improve model performance by enriching the input queries. Representative PE techniques include few-shot learning [23] and CoT prompting [25]. RAG is a framework that involves retrieving relevant passages from an external knowledge base and injecting them into the prompt of the model. It supplements contextual information and improves response accuracy. Kıcıman et al. evaluated the causal reasoning capabilities of LLMs on tasks such as pairwise causal discovery, counterfactual reasoning, and event causality, which achieved high scores and highlighted the need for future methods that integrate LLMs with traditional causal inference techniques [26]. Such causal inference capability has boosted the application of LLMs in aviation accident analysis. Liu et al. proposed the HFACS-CoT+ prompt framework by combining CoT with the HFACS, which significantly enhanced the logical reasoning capabilities of LLMs in human factor analysis [27]. Chen et al. combined LLMs with PE, few-shot learning, and a self-judgment mechanism to automatically identify cause-related entities and relations [28]. Ren et al. proposed a RACI model—a RAG-aided causal identification approach combining LLMs—to extract key causation information from unstructured aviation accident reports and achieved strong performance [29]. All the above-mentioned studies highlight the potential of LLM-based methods in aviation accident analysis. However, current research lacks a sufficient and detailed classification hierarchy for the four primary categories of aviation accident causation. This absence of a standardized reference introduces a challenge for LLMs when classifying causation factors, particularly those related to the aircraft. Given the inherent complexity of an aircraft, which follows a strict hierarchical structure encompassing systems, subsystems, components, and parts, directly employing LLMs for causation classification often results in mismatched granularity or incorrect hierarchical alignment.

2.2. Post-Analysis Methods in Aviation Accident Analysis

After extracting the causations and relations of each accident, the causations and relations of different accidents need to be integrated for deeper accident analysis. KGs are popular for their outstanding capabilities in visualizing large-scale data and performing complex correlation analysis. They are widely used in the field of aviation accident analysis for cross-accident causation analysis, key causation pattern mining, and hazard factor evaluation. Qu et al. proposed an end-to-end entity linking method for constructing a civil aviation emergency KG, which addressed challenges such as complex entity structures and error propagation in traditional two-step linking approaches [30]. Xu et al. proposed a KG-based correlation analysis model for aviation accident causation. By constructing a KG and applying multi-dimensional topological metrics, the model identifies key causal factors and potential risk patterns across different flight phases [31]. Recently, with the development of LLMs, LLM-based KG Question Answering has emerged as a research hotspot. By leveraging the powerful language understanding capabilities of LLMs, users can more easily perform complex KG query tasks, which facilitates deeper research of KGs. Graph-based Retrieval-Augmented Generation (GraphRAG) is an advanced extension of the conventional RAG framework that incorporates graph-structured knowledge to enhance contextual retrieval and reasoning [32]. Instead of treating retrieved documents as independent text chunks, GraphRAG organizes information as nodes and edges, representing entities and their semantic relationships. During retrieval, the model leverages both textual similarity and graph topology to locate semantically related knowledge paths, thereby enabling more interpretable and logically coherent responses. However, its limitations lie in the high cost of graph construction and maintenance, scalability challenges for large knowledge graphs, and potential noise propagation through poorly connected or inaccurately extracted relations.

2.3. Research Gaps

Although existing studies have significantly progressed in the field of aviation accident analysis, several research gaps remain. (1) Early approaches based on statistical models and rule-based systems rely heavily on manual processing and perform poorly when handling large volumes of unstructured textual data. (2) Methods based on pre-trained models require extensive high-quality datasets and considerable computational resources. Meanwhile, (3) LLM-based methods encounter challenges such as hallucination and overly divergent causation type classification.

To address these issues, this study proposes a comprehensive causation-type framework for aviation accident analysis. By integrating PE with RAG, the proposed method enables automatic and efficient causation identification and clustering. Based on the clustering results, a causation KG is constructed, from which risk indicators are derived to identify high-risk causations and causation combinations. Finally, targeted safety recommendations are proposed. This study offers aviation practitioners a more advanced and reliable tool for causation analysis and risk management.

3. Methodology

This section first outlines the framework and schema design of the proposed method, followed by a detailed presentation of the algorithm and validation strategy. This part covers the entire process from accident report processing to KG construction.

3.1. Overall Framework

The proposed method comprises two main stages, namely, single-accident analysis and cross-accident analysis, As shown in Figure 1, the single-accident analysis begins with

accident reports and involves two sequential steps: (1) preliminary causation and relation analysis, which includes preliminary causation identification and relation analysis, and (2) causation standardization, achieved through the clustering of causations and nodes replacement. The standardized causation data generated from this stage are then integrated into the cross-accident analysis, which consists of graph construction and graph analysis. Finally, the insights obtained from the graph-based analysis contribute to the generation of results and suggestions for improving aviation safety.

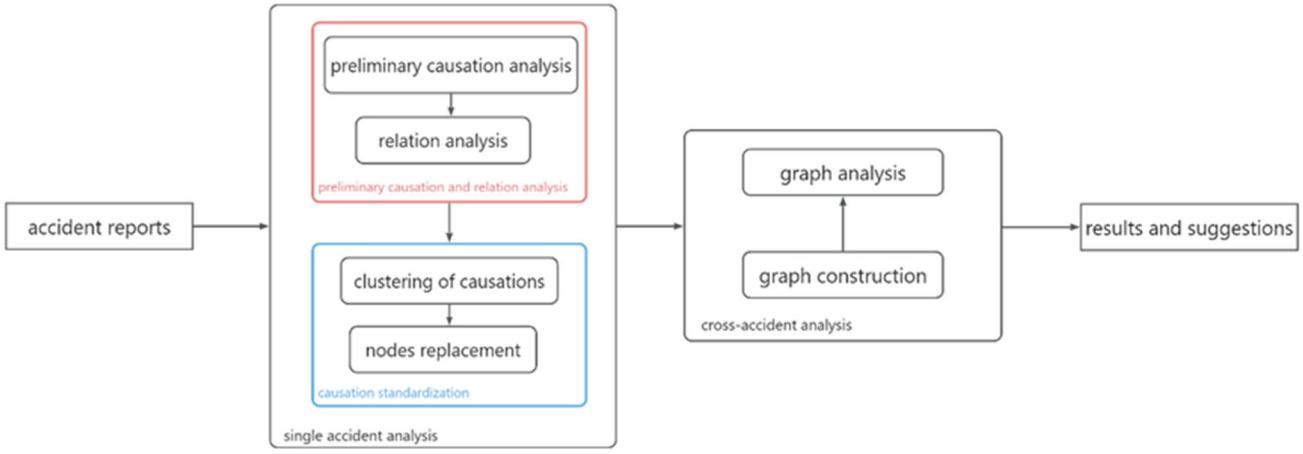


Figure 1. Methodology framework.

In the single-accident analysis stage, as illustrated in Figure 2, each accident report is firstly subjected to individual causation and relation analysis to extract the sentence-level causations and their interrelations to establish preliminary knowledge triples, after which the sentence-level causations will be clustered to establish the standardized knowledge triples. Once this process is completed for all reports, all the standardized triples will be collectively injected into a KG, where designed metrics will be applied to support cross-accident correlation analysis.

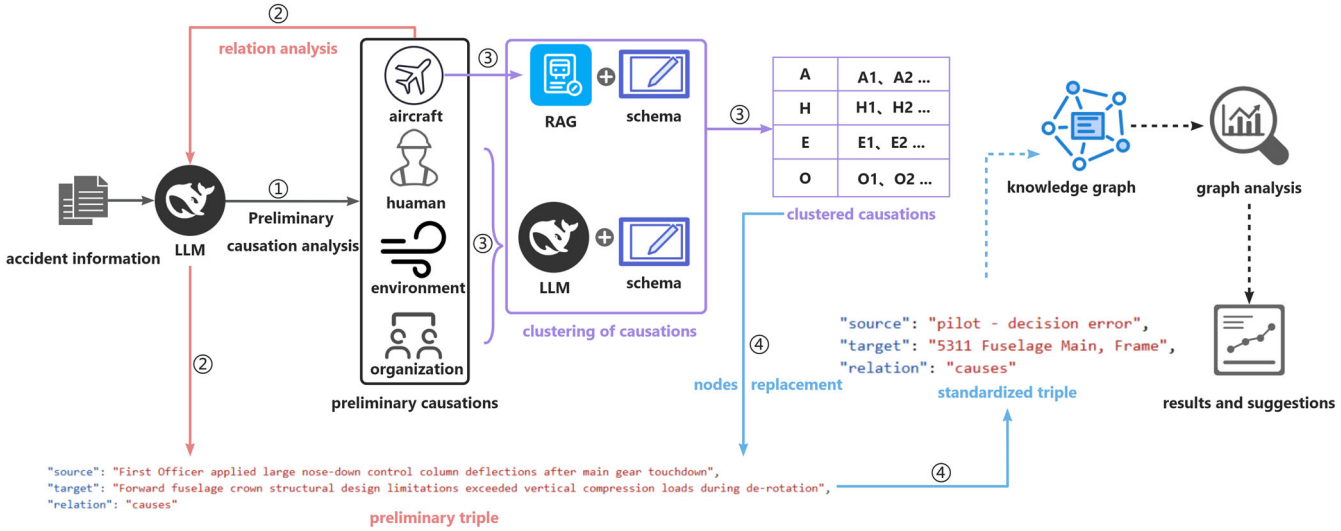


Figure 2. Overall framework of accident analysis.

Specifically, each accident report is initially embedded into a predefined prompt template which is then injected to the LLM to identify causation-related sentence pieces and classify them into four categories: human, aircraft, environment, and organization (black line marked number 1 as shown in Figure 2). The classified causation-related sentence

pieces will then be injected to the LLM for relation analysis (red line marked number 2 as shown in Figure 2) to identify the relations among different causations. With every two causations as the nodes and the relation between them as relation, preliminary knowledge triples are established. However, such triples are not suitable for KG construction because one causation may have various representations, which will hinder cross-accident correlation analysis. Therefore, a schema enhanced clustering phase using the LLM is introduced for causations (purple line marked number 3 in Figure 2) (it is not necessary for relation identification since the relation types are strictly defined in previous relation analysis). Nevertheless, purely LLM based clustering for aircraft-related causations has been proved to be inaccurate, owing to the limited domain-specific knowledge of most current LLMs. To mitigate this limitation, RAG is additionally employed to bridge the domain knowledge gap and generate more accurate clustered aircraft-related causations. Finally, the standardized knowledge triples are obtained by replacing the raw causation sentence in the preliminary triples with their corresponding clustered expressions (blue line marked number 4 as shown in Figure 2). This standardization allows for the integration of different accident reports into a unified KG based on consistent causation entities and relationship types. Subsequently, risk assessment parameters are designed based on domain-specific knowledge to identify high-risk causations and causation combinations. Targeted safety recommendations are then formulated according to the risk evaluation results.

It should be noted that all LLM-related steps in this paper utilize the DeepSeek-V3.1 API with the following parameters: --temperature 0.7, --top_k 20, --top_p 0.95, --min_p 0, --frequency_penalty 1.0, --presence_penalty 1.5, --reptition_penalty 1.1. Such parameters are also applied to every LLM call in later processes.

3.2. Dataset and Schema

To ensure data completeness and account for structural variations under different regulatory frameworks, we selected aviation accident investigation reports published on the official NTSB website that meet the following criteria: (1) the accident occurred on or after 1 January 2000; (2) the operation falls under FAR Part 121: Air Carrier; and (3) the investigation status is marked as "Completed." From this set, we further filtered reports that contain complete information in the following sections: "Factual Narrative," "Analysis Narrative," "Probable Cause," "Highest Injury," and "Damage Level." A total of 348 reports met all criteria and were selected as the final dataset for this study. Figure 3 illustrates the distribution of aircraft damage across all reports.

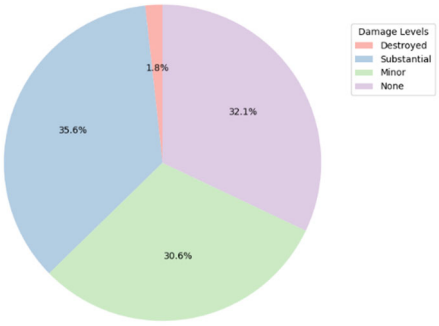


Figure 3. Damage distribution.

It is worth noting that these accident reports are based on the investigators' interpretations of each incident, which may differ from the actual ground truth of what occurred.

Specifically, 32.1% of the accidents resulted in no damage to the aircraft, 30.6% involved minor damage, and 35.6% led to substantial damage. Only 1.8% of the cases resulted in the aircraft being destroyed. Figure 4 presents the distribution of human injuries across

all accident reports. Most accidents did not result in any injuries. Specifically, 7.6% of the accidents led to minor injuries, 15.7% resulted in serious injuries, and 2.0% caused fatal injuries.

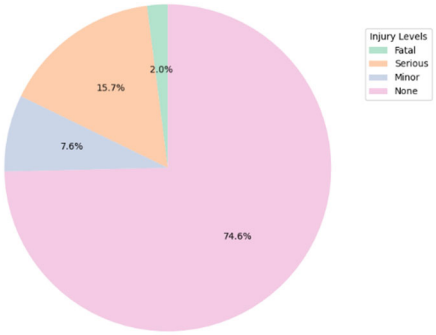


Figure 4. Injury distribution.

This study introduces a domain-informed and original classification scheme for categorizing aviation accident causal factors. Building upon the conventional Human–Machine–Environment–Organization framework [33], the proposed system further refines each category into specific subtypes to enhance analytical precision.

For human-related factors, as shown in Figure 5, we referred the NTSB Aviation Coding Manual [34] and adopted a dual-dimensional classification based on "Personnel Type" pluses "Factor Type." The Personnel Type dimension includes six categories: pilot, cabin crew, maintenance personnel, air traffic controller, ground handling personnel, and other personnel. The Factor Type dimension also comprises six categories: decision error, procedural violation, skill deficiency, physiological or psychological issue, situational awareness failure, and other unsafe behavior. This taxonomy emphasizes the roles of key personnel involved in accidents and the types of unsafe behavior. The inclusion of "other personnel" and "other unsafe behavior" ensures a closed classification loop. For organizational-related factors, a similar dual-dimensional framework is applied as shown in Figure 6, which combines "Organization Type" and "Factor Type." The Organization Type includes six entities: aircraft operator, airport operator, air traffic control unit, manufacturer (aircraft/component), regulatory authority, and other organizations. The corresponding Factor Type includes process deviation, coordination breakdown, oversight failure, policy or procedural deficiency, resource management failure, and other organizational causes. Similarly, the combination of "other organizations" and "other organizational cause" forms a complete and flexible classification structure for organizational causes. The classification of environmental factors is relatively straightforward, including categories such as bird strike, runway contamination, and others. A detailed breakdown is provided in Figure 7. The classification of aircraft-related factors is more complex. However, given that aircraft systems are categorized under a strict technical framework in real-world aviation practice, this study adopts the Joint Aircraft System/Component (JASC) Code issued by the Federal Aviation Administration (FAA) on 27 October 2008, as shown in Figure 8. This schema enables us to assign various mechanical and structural failures to their corresponding systems. The use of this classification has strong practical relevance, as aircraft design, operation, and maintenance activities are typically organized by systems. To facilitate further research and reference, all subcategories within the four main causation types have been assigned specific codes. For aircraft-related factors, only systems represented in the dataset are included. In addition, the consequence is categorized into 2 main types (Damage for aircraft and Injury for human) and 8 subtypes, as shown in Figure 9.

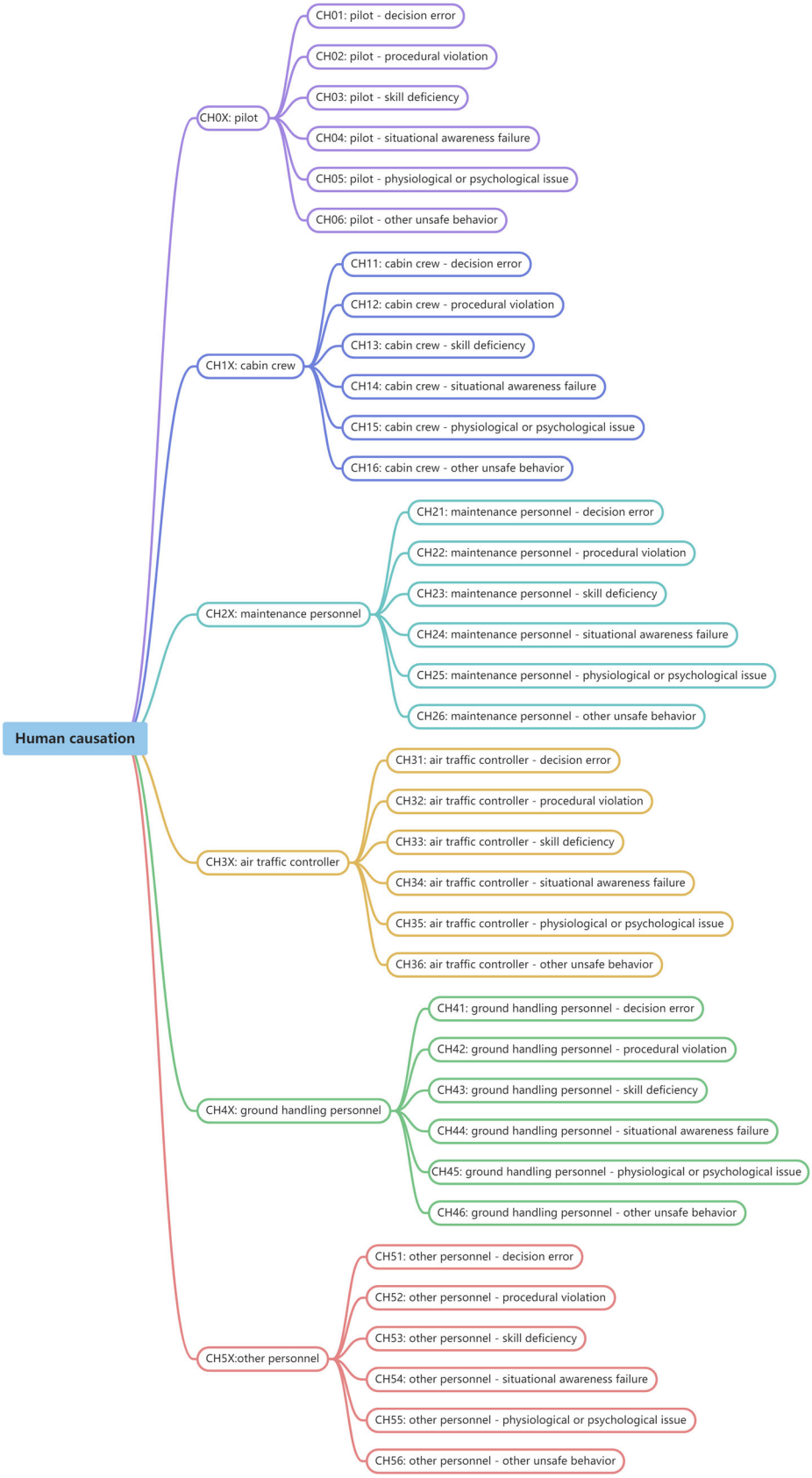


Figure 5. Human causation schema.

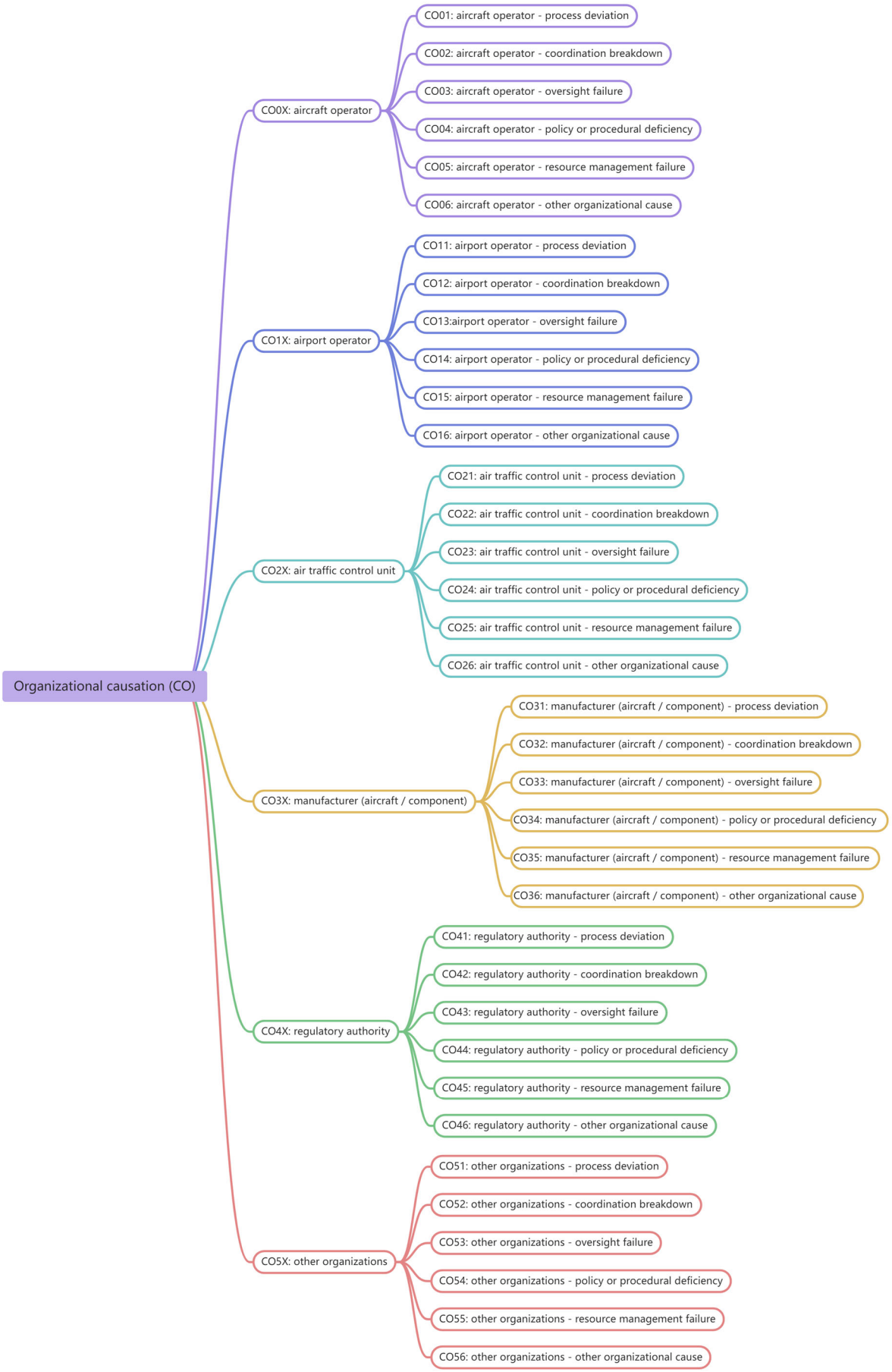


Figure 6. Organization causation schema.

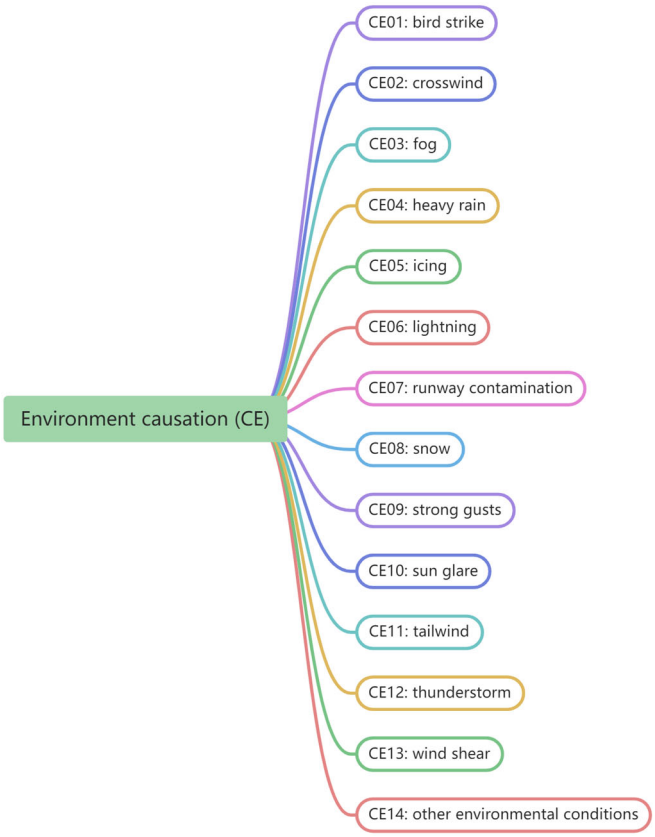


Figure 7. Environment causation schema.

3.3. Algorithm for Single-Accident Analysis

The single accident analysis which aims to generate standardized causation and relation triples contains two main stages: preliminary causation and relation analysis (Figure 10) and causation standardization (Figure 11).

In preliminary causation and relation analysis, as shown in Figure 10, the accident description is embedded into a predefined causation analysis prompt template to generate an initial causation analysis prompt. Considering the inherent randomness in the output of LLMs, a trial mechanism is introduced to ensure output reliability. Specifically, the LLM-generated output must conform to a predefined format—the inclusion of causation factors from all four major categories: Human, Aircraft, Environment, and Organization. Only outputs that meet this structural requirement are accepted. The output from the preliminary causation analysis is then indexed with the following mapping rules:

"1x": human causations,

"2x": aircraft causations,

"3x": environmental causations,

"4x": organizational causations.

"x" could be any number between 0 and 9 since each causation type might have several causations in an accident.

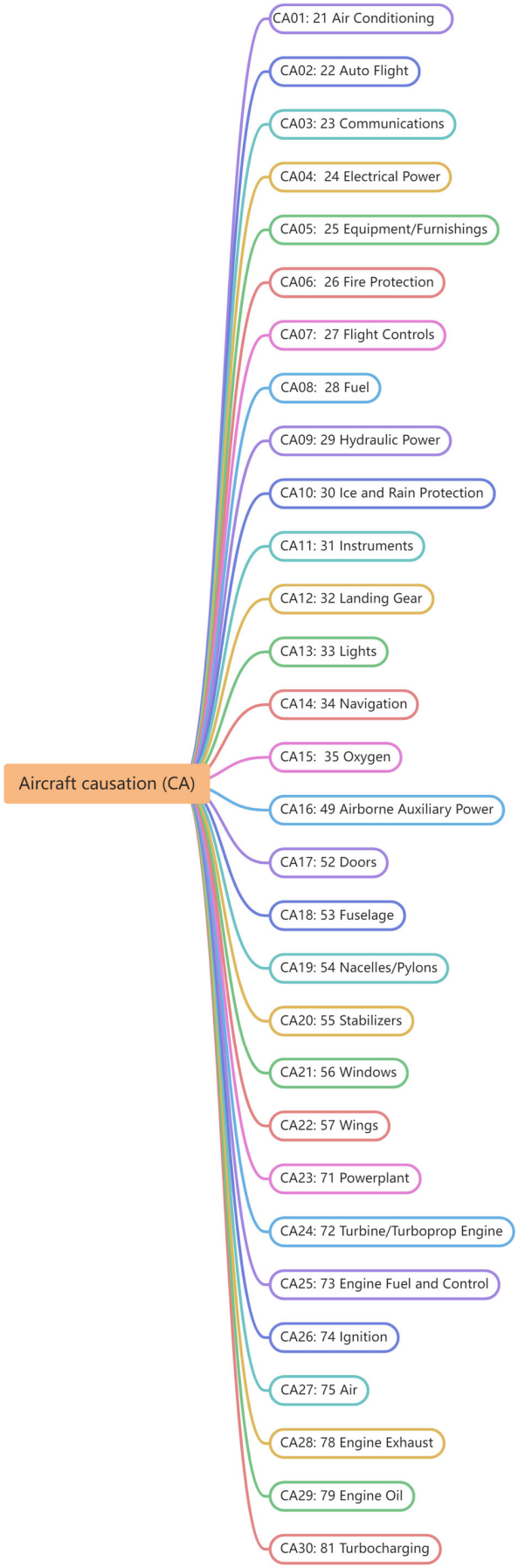


Figure 8. Aircraft causation schema.

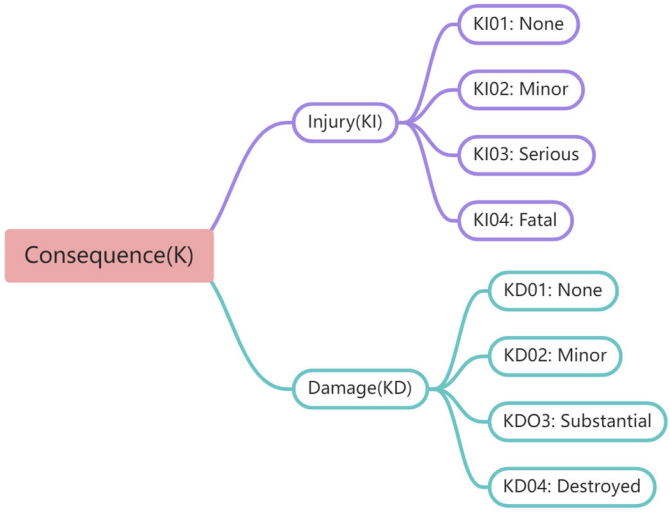


Figure 9. Consequence schema.

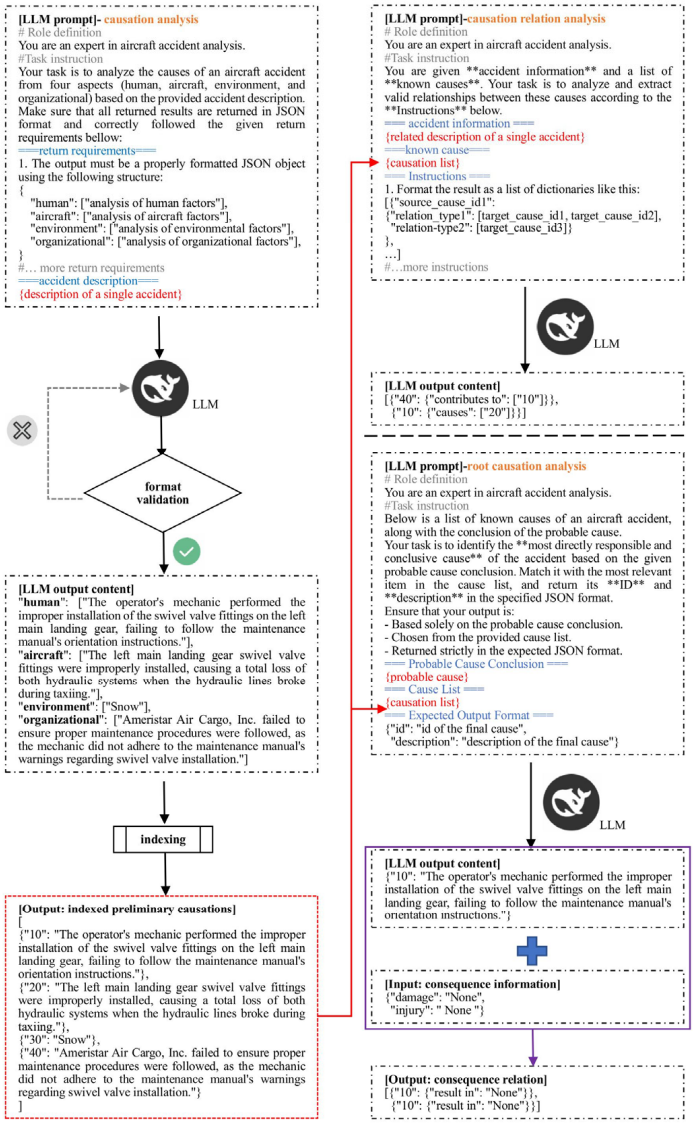


Figure 10. Workflow of preliminary causation and relation analysis (detailed prompts are given in Appendix A).

These indexed causation results serve as the foundation for two subsequent tasks: causation relation analysis and direct causation analysis. Causation relation analysis aims to identify the relationships between all indexed causation factors. First, all indexed causations are paired without duplication. To ensure contextual completeness, each causation pair is combined with the relevant portion of the original accident description to construct a causation relation analysis prompt. This prompt is then submitted to an LLM, which generates the causal relationship between the paired causations. Direct causation analysis aims to identify the most immediate causation of an accident. To achieve this goal, all indexed causations, along with the "Probable Cause" section from the accident report, are integrated into a prompt template. The prompt is then processed by an LLM, which outputs the direct causation along with its corresponding index. Finally, the identified direct causation is paired with the related consequence to establish the consequence relation.

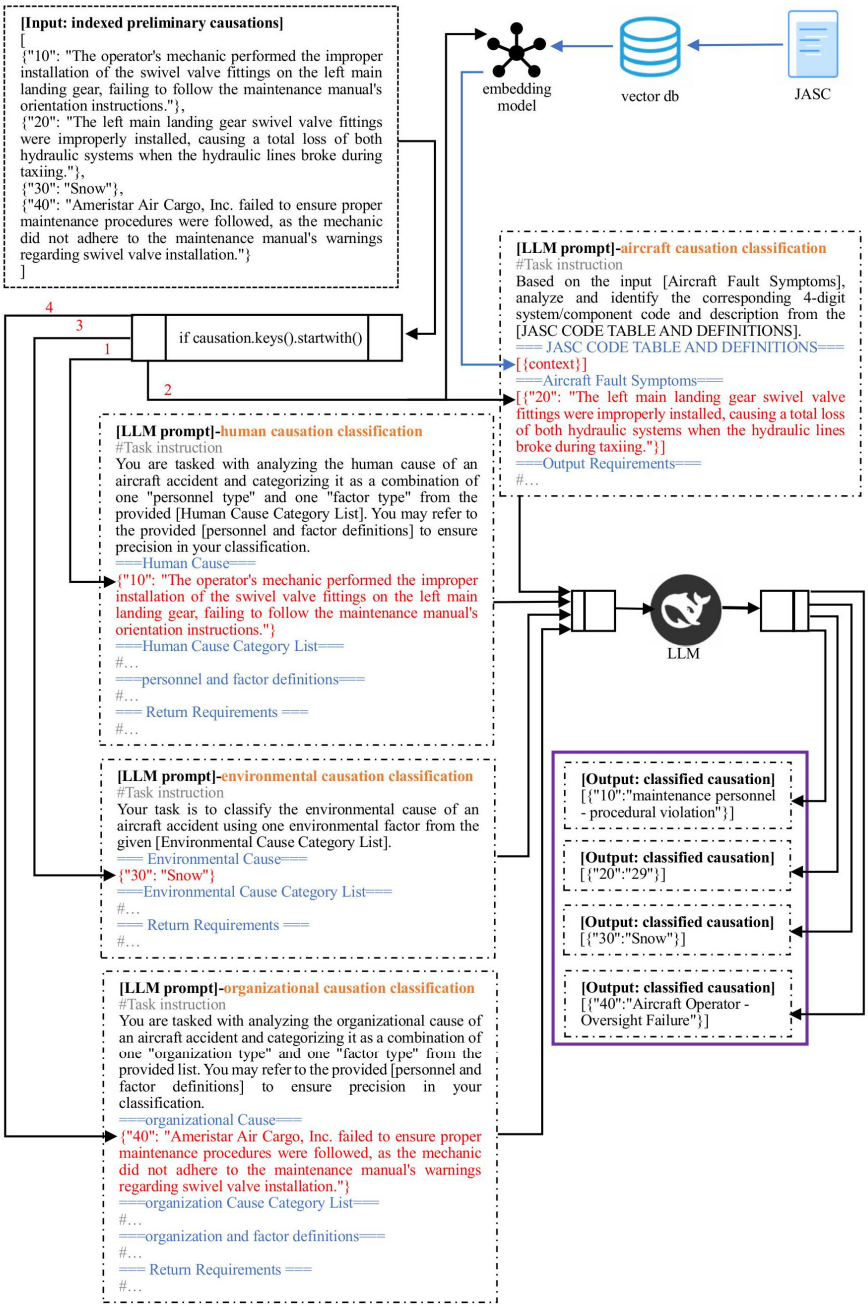


Figure 11. Workflow of causation standardization (detailed prompts are given in Appendix A).

The indexed preliminary causation results serve as the input for the causation clustering task, as shown in Figure 11, and each causation factor is first routed according to its assigned index. If the causation type falls under aircraft-related factors, then an embedding model will be employed to fetch several relevant knowledge fragments from an external knowledge base. These retrieved fragments are subsequently integrated into the corresponding prompt template to construct the task-specific input. For causation factors of other types, the indexed causations are directly combined with their respective clustering templates to form the task prompts. All prompts are subsequently processed by the LLM, which generates the final clustered causations. With the clustered causations and their relations, standardized triples are established.

All the aforementioned tasks are implemented in Python, and Table 1 presents the pseudocode outlining the implementation details of these methods.

Table 1. Pseudo-code for causation and relation analysis of a single accident.

Pseudo-Code for Causation and Relation Analysis of a Single Accident
input: factual narrative, analysis narrative, probable cause Output: standardized triples 1. Extract and classify preliminary causations using LLM based on the accident narratives 2. Verify and refine causation outputs until the required format is satisfied 3. For each causation pair, analyze causal relations using the contextual narratives 4. Construct preliminary knowledge triples (source–relation–target) 5. Identify the direct causation based on the probable cause statement and preliminary causations. 6. Extend triples with consequence information to form complete knowledge relations 7. Cluster similar causations to obtain standardized causation representations 8. For the clustering of aircraft-related causations, enrich classification with domain knowledge (e.g., JASC codes) 9. Establish standardized triples by replacing the causations in preliminary triples with its clustered expressions 10. Return all standardized triples

3.4. Graph Construction

After completing the causation and relationship analysis for all accident reports, a KG is constructed to enable deeper analysis of high-risk causation factors and their interrelations. This graph serves as a unified framework to integrate all identified causation entities and relationships across multiple accidents. The KG consists of two types of data: nodes and edges. The nodes are composed of two main categories: causation and consequence.

Each causation node contains three key attributes:

- **description**: the clustered causation or consequence entity,
- **count**: the number of occurrences of the entity across all accident cases,
- **tag**: the primary category of the causation, which is classified as one of Human, Aircraft, Environment, or Organization.

Consequence nodes follow a similar structure.

The edges represent two types of relationships: causation relations and consequence relations. Each edge also includes a count field, which quantifies the number of times a specific entity-to-entity relationship occurs across the dataset. This structured representation provides a solid foundation for subsequent correlation analysis, risk identification, and safety decision support.

NeoDash is a powerful graph visualization tool that not only supports basic rendering of entities and relations but also enables the visualization of node and edge weight parameters, such as using node size and edge thickness to represent quantitative values. In this study, the count attributes of nodes and edges were appropriately normalized and scaled to enhance interpretability. Using Cypher queries, we automated the construction of the aviation accident causation KG, as illustrated in Figure 12.

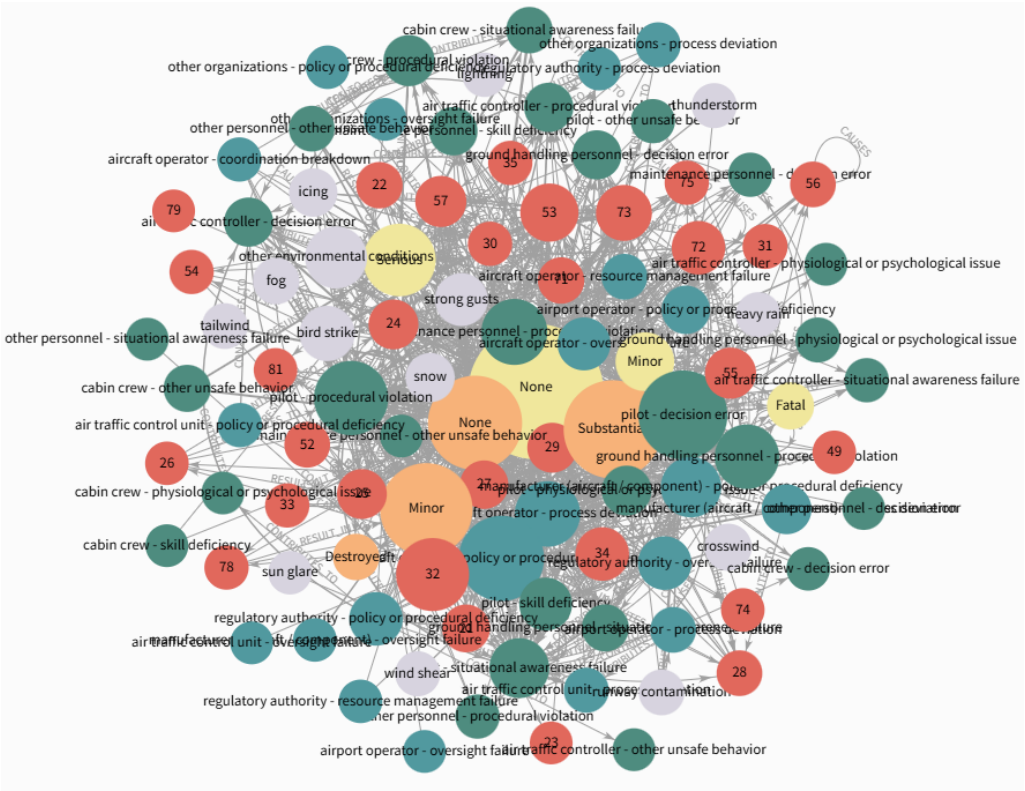


Figure 12. Causation and relation graph of all accidents.

4. Results

4.1. Evaluations

To make sure the outcome of our method is of reasonable accuracy, evaluations over different tasks are necessary. According to Figure 1, the evaluation of the analysis results primarily focuses on two tasks: preliminary causation and relation analysis, as well as causation standardization.

Specifically, for the preliminary causation and relation analysis task, the assessment is conducted in two stages—preliminary causation analysis and relation analysis—where the results from both steps are manually verified against the original accident descriptions to determine their consistency with the source material. In the preliminary causation analysis evaluation, each identified causation description is cross-checked with the original text, and the accuracy metric for this stage is derived by calculating the ratio of verified causations to the total number of identified causations. Similarly, during the relation analysis phase, every set of discovered links is validated against the original content, and the final accuracy metric is obtained by comparing the number of verified links to the total number of identified links.

In order to support the evaluation of preliminary causation and relation analysis, we created a validation dataset comprising 50 random accident reports, each annotated with causation and relational analysis results. In addition, to verify the robustness of our method over different LLMs, we used three different LLMs, which are Deepseek-V3.1(around 671 B),

GPT-4.1(thousands of billions) and Qwen-MAX(thousands of billions), as the candidate of the base model in our method. The evaluation results are shown in Table 2.

Table 2. Evaluation results of preliminary causation and relation analysis.

Task	Parameter	DeepSeek-V3.1	GPT-4.1	Qwen-MAX
Preliminary causation analysis	Total number	169	170	166
	Accuracy	0.952	0.957	0.954
Relation analysis	Total number	118	120	122
	Accuracy	0.983	0.984	1.000

As shown in Table 2, in the preliminary causation analysis, all evaluated LLMs achieved similar accuracy rates of around 95%, with only minor variations in the total number of identified causations (ranging from 166 to 170). Similar results were observed in the relation analysis, where all LLMs consistently achieved high accuracy rates exceeding 98%, and the total number of identified links (relationships between causations) ranged from 118 to 122. The evaluation results for both the preliminary causation and relation tasks demonstrate that the identified causations and links are highly consistent with the original accident descriptions. Furthermore, the method exhibits strong robustness across different LLMs, and such high accuracy ensures the reliability of the data for subsequent analysis.

To assess the reliability of the causation entity clustering results, we conducted a manual evaluation involving domain experts. A total of 200 causation instances were randomly selected covering all four major categories. Two evaluation metrics were designed for this purpose: a_c—accuracy of causation type c (as shown in Equation (1)) and a_{total}—the average accuracy of all types of causations (as shown in Equation (2)).

$$a_c = n_{cp}/N_c \tag{1}$$

where N_c denotes the total number of causation instances in category c, and n_{cp} represents the number of correctly classified instances in that category.

$$a_{total} = \sum_{c \in C} n_{cp}/N \tag{2}$$

where C = {Human, Aircraft, Environment, Organization}, and N is the total number of evaluated causation instances across all categories.

As shown in Figure 13, the Aircraft category exhibits the lowest clustering accuracy at 0.865, whereas the Environment category achieves the highest accuracy at 0.979. The remaining two categories, Human and Organization, both exceed 0.90 in accuracy. Several factors contribute to the relatively low accuracy of the Aircraft category, with the granularity of the external knowledge base being the most significant. Although the JASC code provides a comprehensive structure of aircraft systems, its descriptions—especially at the subsystem level—are often vague, which leads to information granularity mismatches. For instance, report CHI08IA292 mentions a failure of the K106 electrical relay, but the JASC code lacks such detailed component-level records. As a result, LLMs have difficulty identifying the corresponding system. In addition to knowledge base granularity, other factors also influence the performance of RAG-based aircraft causation clustering, including (but not limited to) the size of the LLM, the choice of embedding model, and the number of retrieved knowledge segments. Overall, average clustering accuracy across all categories reaches 0.926, which indicates that the clustering results are generally reliable and agree well with expert judgment.

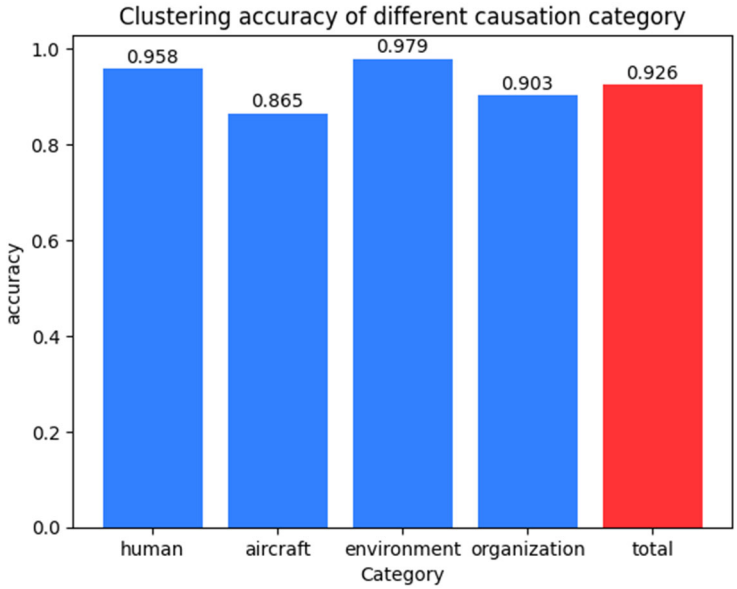


Figure 13. Clustering accuracy of different causation categories.

4.2. Result of Causation Analysis

To facilitate the significance analysis of each causation within the overall KG, three evaluation metrics were defined: importance, degree, and weighted degree. As shown in Equations (3)–(5), the importance of a node refers to the number of occurrences of the node as a proportion of the total number of all causations. The degree of a node refers to the total number of edges connected to it, which reflects its direct connectivity. By contrast, the weighted_ degree incorporates the weights of the edges, which captures the quantity and strength of connections in a node. The results are shown in Figures 14–16.

$$importance(x) = C_n(x)/N \tag{3}$$

$$iodegree = indegree + outdegree \tag{4}$$

$$weighted_iodegree = w_indegree + w_outdegree \tag{5}$$

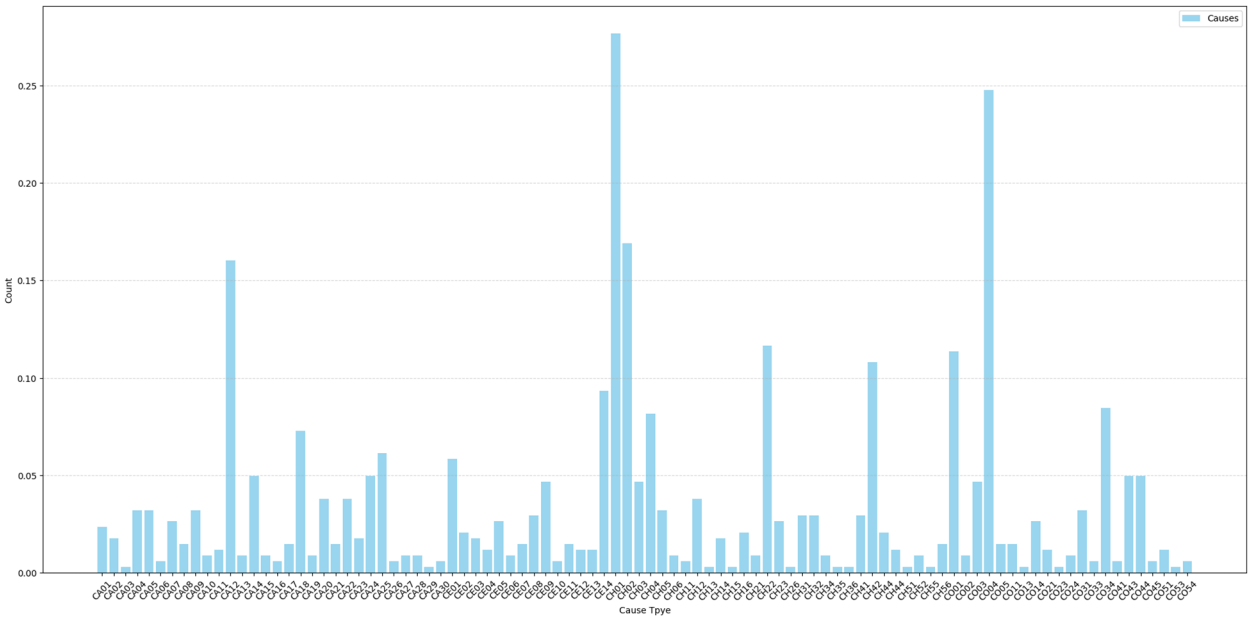


Figure 14. Importance of different causations.

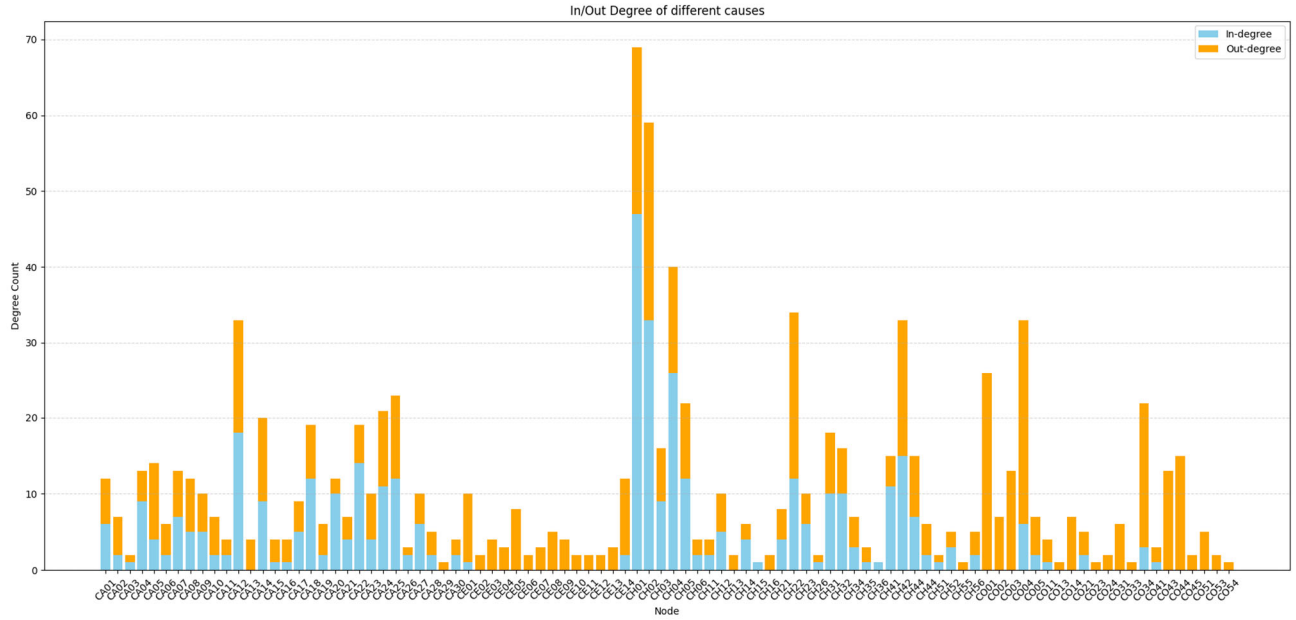


Figure 15. I/O degree of different causations.

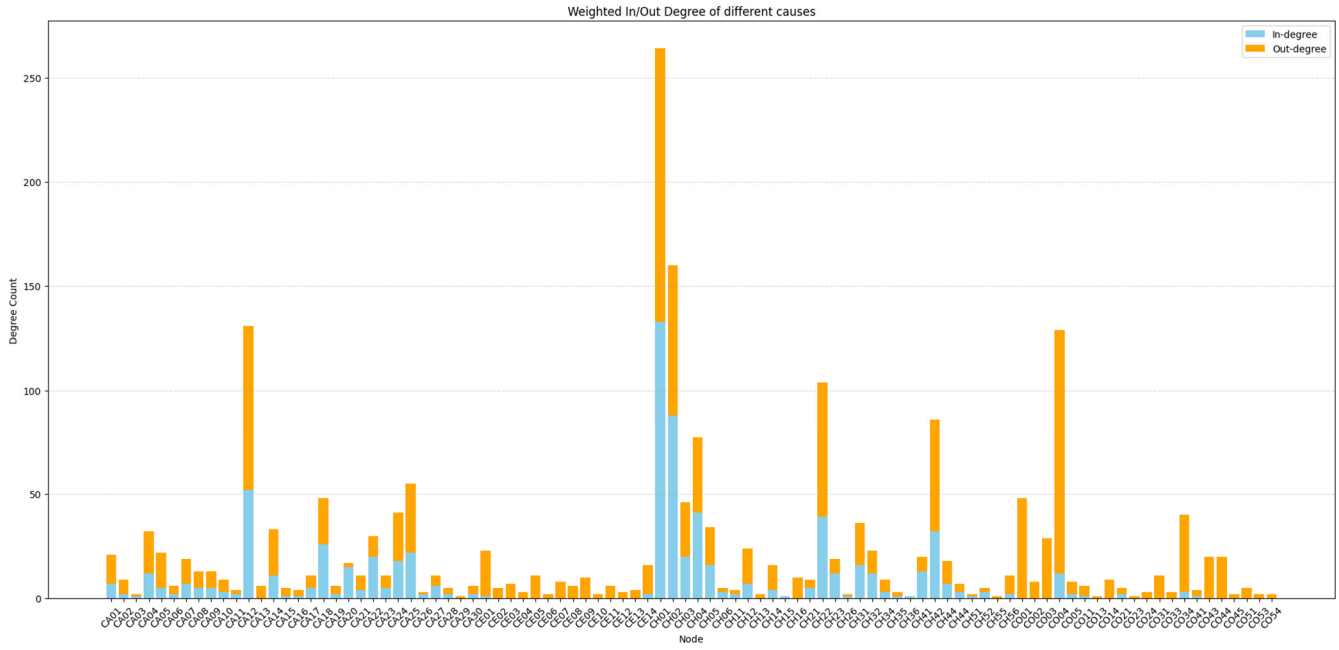


Figure 16. Weighted I/O degree of different causations.

As shown in Figure 14, the top four causation factors in terms of importance are CH01, CO04, CH02, and CA12. This result indicates that these causes appear most frequently across all analyzed accidents. Using an importance score threshold of 0.1 to define high-impact causations, CH contains the largest number of high-impact causes, with four entries (CH01, CH02, CH22, CH42). CO follows with two entries (CO01, CO04), while CA has only one (CA12), and CE has none. The results indicate that human causations, especially pilot-related causations, have the highest frequency of occurrence among all aviation accidents.

As shown in Figure 15, CH01 exhibits the highest number of incoming and outgoing relation types. Using a degree count of 30 as the threshold for identifying high-connectivity causation factors, the top-ranked causes include CH01, CH02, CH04, CH22, CA12, CH42, and CO04. Figure 16 reveals a similar trend: when considering the weighted number of causation-related connections, CH01 and CH02 remain the two highest-scoring causation

factors. Given that causation codes starting with CH0X correspond to pilot-related factors, pilot-related causes exert the greatest influence on the propagation of aviation accidents. In addition, CA12 and CO04 exhibit a relatively high number of connections, which suggests that they play significant roles in the propagation of aviation accidents.

To further identify key causal factors, we computed the contribution of each causation type to the consequences (denoted as *Kc* as shown in Equation (6)). *Kc* is defined as the sum of edge weights between each causation node and the eight types of consequence nodes.

$$K_c = \sum_{d \in KD} count_{cd} + \sum_{i \in KI} count_{ci} \tag{6}$$

where $count_{cd}$ is the relation number of node c and damage consequence node d, and $count_{ci}$ is the relation number of node c and injury consequence node i.

The results in Figure 17 show that only CA and CH causation types have direct connections to consequence nodes, while CE and CO do not contribute directly. In particular, the first three causations with the highest consequence contribution score are CH01, CA12, and CH42.

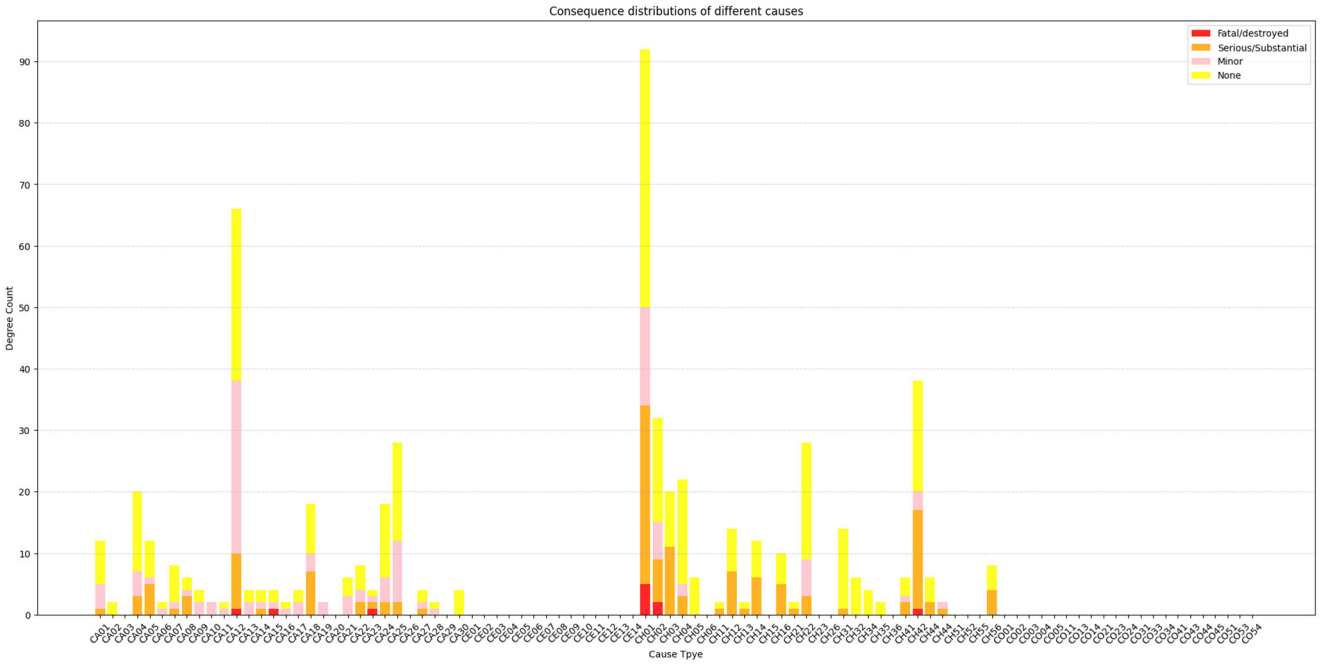


Figure 17. Consequence distributions of different causations.

Synthesizing the evaluation of multiple dimensions, we conclude that the top three causations in terms of overall importance are CH01, CA12, and CH42.

4.3. Result of Correlation Analysis

To quantitatively assess the positional role of an entity within the causation propagation chain, two metrics were introduced: *Active Closeness* (denoted as AC_i in Equation (7)) and *Passive Closeness* (denoted as PC_i in Equation (8)) [9]. AC_i measures the extent to which an entity i can directly or indirectly trigger other entities within the KG, which reflects its causative influence. On the contrary, PC_i quantifies the degree to which entity i can be triggered by other entities, which represents its susceptibility to upstream causative factors.

$$AC_i = 1/(\sum_{n \in Nodes} dist_min_{in} / \sum_{n \in Nodes} dist_true_{in}) \quad (7)$$

$$PC_i = 1/(\sum_{n \in Nodes} dist_\min_{ni} / \sum_{n \in Nodes} dist_true_{ni}) \quad (8)$$

where n is any entity node in the aviation accident graph except i; *dist_min* represents the shortest path from its subscript node 1 to node 2; *dist_true* signifies whether a causal path exists between its subscript node 1 and node 2; if the path exists, then *dist_true* = 1; otherwise, *dist_true* = 0.

The calculative results of all types of entities are shown in Figure 18. All KI and KD entities have only PC_i value, which is reasonable given that they represent the outcomes of accidents and therefore reside at the end of the causation chain. Most causations in groups CE and CO have only AC_i value, which means that most CE and CO causations predominantly exhibit outgoing relations, with only a small number of causations possessing incoming relations. By contrast, CA and CH causation groups demonstrate a more balanced distribution of AC_i and PC_i values, which implies that most of their causations possess incoming and outgoing relations. This observation suggests that environmental and organizational factors often serve as originating causes in the propagation of aviation accidents, whereas human and aircraft factors tend to act as intermediary transmission nodes.

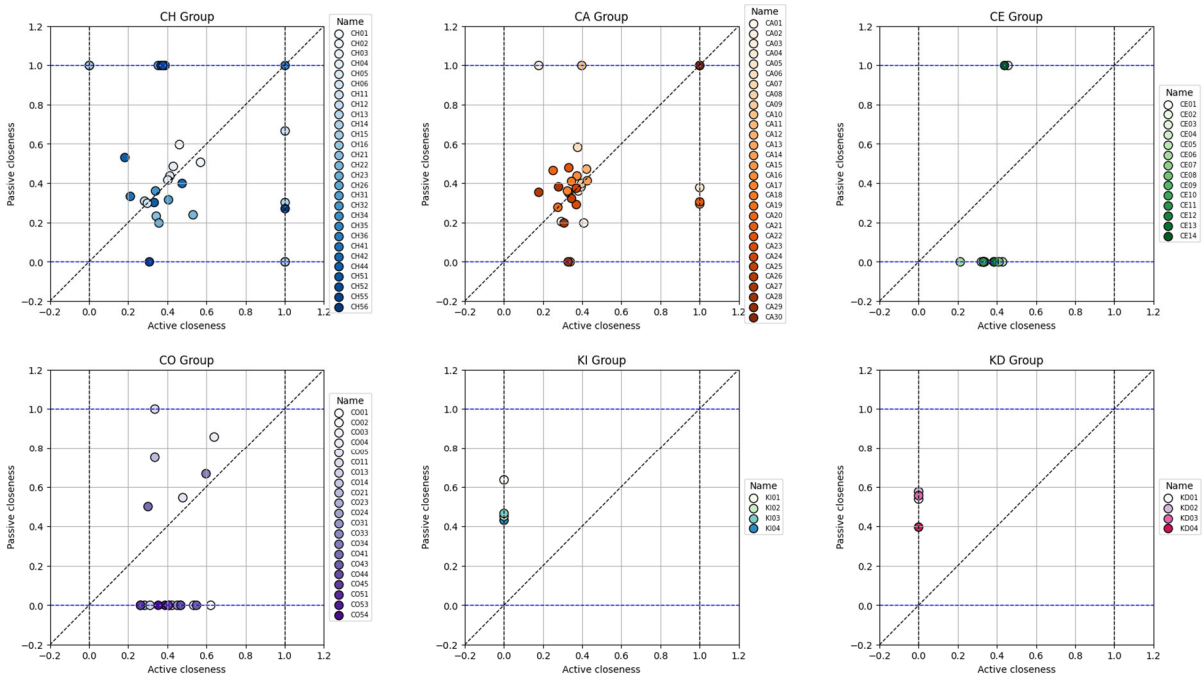


Figure 18. Active/passive closeness of all entities.

Overall, these findings reveal a typical causation propagation path in aviation accidents that can be characterized as

(Environment/Organization) → (Human/Aircraft) → Consequence

In other words, aviation accidents often originate from environmental or organizational deficiencies, which subsequently trigger equipment failures or human errors. These phenomena ultimately lead to undesired outcomes

4.4. Result of Special Case Analysis

Notably, some causations have AC_i value for CE and CO groups, which may cause certain confusion. To verify the plausibility of incoming relations for certain Environment and Organization causation nodes, we conducted a detailed causation chain analysis of two representative aviation accidents.

In the case of CE01 (Bird Strike), a query on the constructed KG revealed the following causal relationship:

"Airport Operator—Policy or Procedural Deficiency"→ "Bird Strike"

This relation originates from an incident that occurred on 18 February 2008, in Austin, Texas. This incident was documented under NTSB report number DFW08IA073. The corresponding textual recording for *"Airport Operator—Policy or Procedural Deficiency"* is shown as follow:

> *"According to a representative of the airport, they do not have an FAA approved Wildlife Management Plan since there had been no previous incidents that fit the criteria set aside by FAR 139.337(a) or Advisory Circular 150/5200-33B, titled Hazardous Wildlife Attractants on or Near Airports that required such a plan."*

The narrative indicates that the lack of an FAA-approved Wildlife Management Plan—a direct result of procedural deficiencies on the part of the airport operator—contributed to the bird strike, which justifies the presence of an incoming edge to the environmental factor in the graph.

Similarly, for CO21 (*Air Traffic Control Unit—Process Deviation*), the graph shows the following incoming relationship:

> *"Regulatory Authority—Policy or Procedural Deficiency"→ "Air Traffic Control Unit—Process Deviation"*

This relation is exemplified by an accident that occurred on October 21, 2009, in Minneapolis, Minnesota. This accident was documented under NTSB report number DCA10IA001. The corresponding causation texts are as follows:

> *"No national standardized procedures exist when automated information transfers are used instead of the paper flight progress strips to nonverbally document and confirm air traffic control information among controllers."—"Regulatory Authority—Policy or Procedural Deficiency".*

> *"ATC management did not complete the required notifications for a NORDO airplane in a timely manner as required by Federal Aviation Administration directives."—"Air Traffic Control Unit—Process Deviation"*

The absence of standardized procedures indirectly contributed to the failure of the air traffic control unit to perform timely notifications, which demonstrates how upstream regulatory deficiencies can trigger downstream procedural deviations within operational units.

These examples reinforce the validity of representing Environment and Organization nodes with incoming causal links in the constructed aviation accident causation graph.

5. Discussion on High-Risk Causations

To further identify high-risk causation combinations, this section focuses on the three high-risk causation factors identified in Section 4.2, which are CH01, CA12, and CH42, as the primary research targets. By designing specific evaluation metrics, the most probable and impactful causation combinations are located within the previously constructed KG.

5.1. Evaluation Metrics

Considering the bidirectional nature of relations, this section introduces two evaluation metrics to quantitatively assess the association between a given node and its connected nodes: *Support* ($S(X \mid Y)$) and Development ($D(Y \mid X)$). As shown in Equation (9), Support measures the probability that the occurrence of node Y occurs is caused by node X. Conversely, as defined in Equation (10), Development reflects the likelihood that node Y occurs following the occurrence of node X.

$$S(X|Y) = C_r(X- > Y)/C_{r_{\text{in}}}(Y) \tag{9}$$

$$D(Y|X) = C_r(X- > Y)/C_{r_{out}}(X) \tag{10}$$

where $C_{r_{in}}(X)$ represents the incoming relation number of node X, $C_{r_{out}}(X)$ denotes the outgoing relation number of node X, and $C_r(X- > Y)$ refers to the count of edges from X to Y.

5.2. CH01: Pilot—Decision Error

By extracting all incoming and outgoing relations that are directly connected to CH01, the results are visualized in Figure 19. Preliminary statistics reveal that CH01 has a total of 41 upstream nodes and 17 downstream nodes, which indicates that pilot decision errors are influenced by different factors and, in turn, have substantial downstream impacts on various safety-related entities. These decisions can directly contribute to accident outcomes of varying severity. Considering that this section focuses solely on inter-causation relationships, all nodes related to consequences are excluded from the analysis. Using Equations (8) and (9), we identify the top three upstream causations with the highest support values (S(X | CH01)) and the top three downstream causations with the highest development values (D(Y | CH01)), as shown in Figure 20. The most strongly associated upstream nodes are:

> *X = {CO04: Aircraft Operator—Policy or Procedural Deficiency, CH02: Pilot—Procedural Violation, CE07: Runway Contamination},*

while the most likely downstream nodes are:

> *Y = {CH02: Pilot—Procedural Violation, CA12: 32 Landing Gear, CA22: 57 Wings}.*

These findings suggest a high-risk causation propagation chain centered around *CH01: Pilot—Decision Error*, represented as follows:

> *Aircraft operator—policy or procedural deficiency/pilot—procedural violation/Runway contamination*
>
> *→ pilot—decision error*
>
> *→ pilot procedural violation/32 landing gear/57 wings*

This chain illustrates how systemic, human, and environmental factors converge to influence pilot decision-making. It highlights how these decisions propagate through technical and procedural failures, which ultimately leads to serious incident outcomes. To reduce the occurrence of similar accident chains, airlines should establish additional detailed and reasonable regulations, pilots must strictly adhere to operational procedures, and airport authorities should enhance the efficiency and reliability of runway availability management.

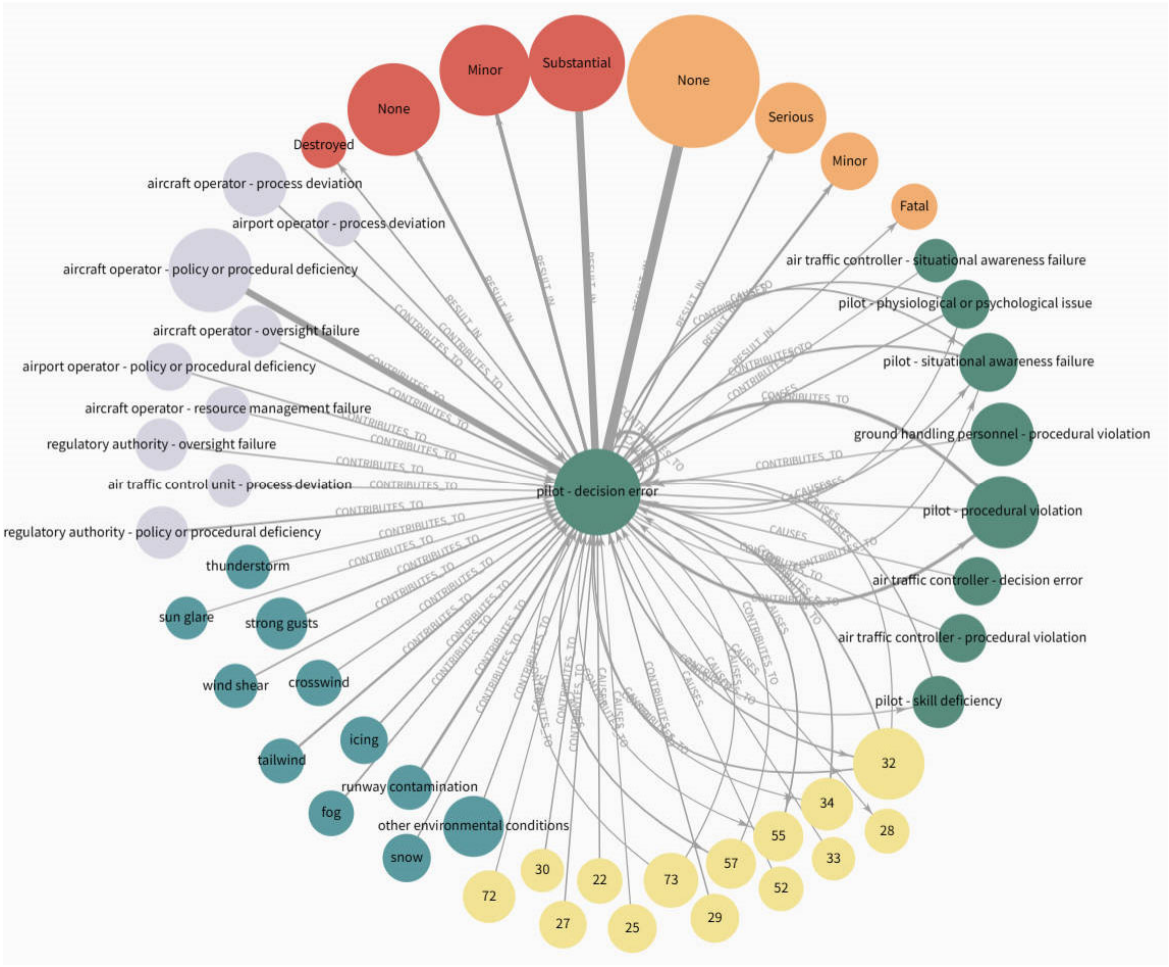


Figure 19. All in/out connections of pilot-decision error (CH01) in causation-relation knowledge graph.

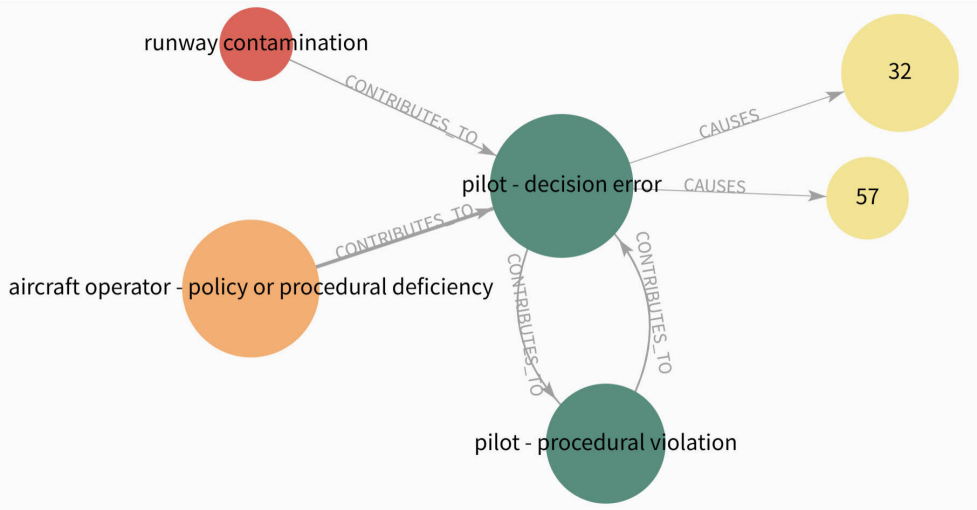


Figure 20. Top 3 in/out connections of pilot-decision error (CH01) in causation-relation knowledge graph.

5.3. CA12: 32 Landing Gear

Following the same analytical approach as in Section 5.2, all nodes directly connected to CA12 are shown in Figure 21. As depicted in the figure, CA12 is associated with 13 upstream nodes and 11 downstream nodes. After excluding consequence nodes, the top

three upstream and downstream causations with the highest $S(X \mid CA12)$ and $D(Y \mid CA12)$ score are identified and visualized in Figure 22:

> *X = {CO04: aircraft operator—policy or procedural deficiency, CH22: maintenance personnel—procedural violation, CO34: manufacturer (aircraft/component)—policy or procedural deficiency}*
>
> *Y = {CH01: pilot—decision error, CH02: pilot—procedural violation, CH03: pilot—skill deficiency}*

These findings reveal a high-risk causation propagation chain centered on *CA12: landing gear*, which can be represented as

> *aircraft operator—policy or procedural deficiency/maintenance personnel—procedural violation/manufacturer (aircraft/component)—policy or procedural deficiency*
>
> *→ landing gear malfunction*
>
> *→ pilot—decision error/procedural violation/skill deficiency*

This chain suggests that failures in the landing gear system are highly correlated with policy and procedural deficiencies, particularly those originating from aircraft operators, maintenance personnel, and manufacturers. Given that the landing gear is a critical subsystem for ensuring safety during the landing phase, its malfunction significantly challenges the ability of the pilot to make timely decisions, execute appropriate responses, and manage emergency situations that may exceed routine training. To prevent such chain from happening, aircraft manufacturers and operators should develop additional reasonable and comprehensive landing gear maintenance procedures. They should also ensure that maintenance personnel strictly follow the instructions outlined in maintenance manuals and other relevant documents. Moreover, pilot training should be enhanced to include additional scenarios involving landing gear failure, which improves the ability of the pilot to respond effectively in such emergencies.

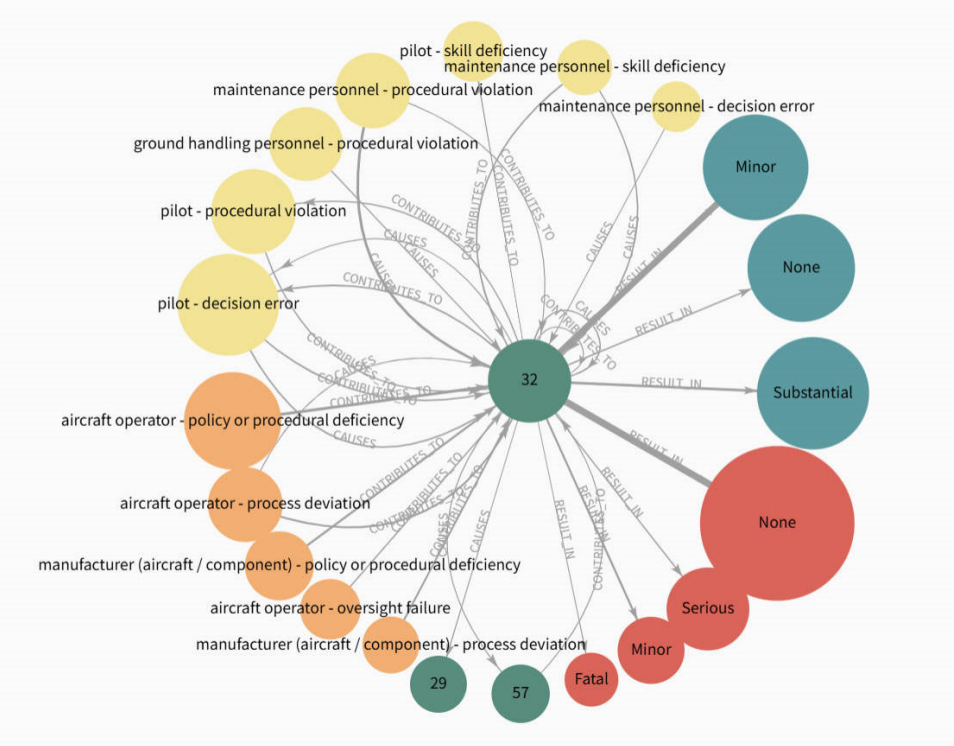


Figure 21. All in/out connections of landing gear (CA12) in a causation-relation knowledge graph.

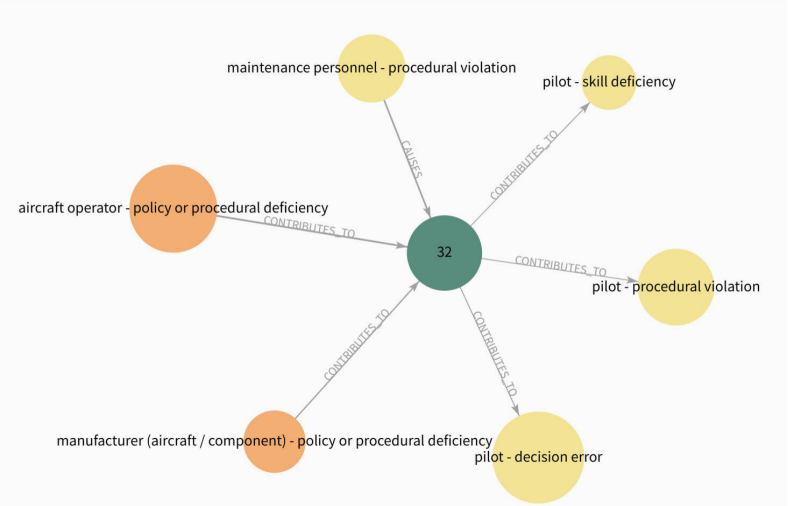


Figure 22. Top 3 in/out connections of landing gear (CA12) in a causation-relation knowledge graph.

5.4. CH42: Ground Handling Personnel—Procedural Violation

All nodes that are directly connected to CH42 are illustrated in Figure 23, which comprise 13 upstream nodes and 17 downstream nodes. After filtering out consequence-related nodes, the top-scoring nodes based on $S(X \mid CH42)$ and $D(Y \mid CH42)$ are as follows (Figure 24):

X = {CO04: aircraft operator—policy or procedural deficiency, CO01: aircraft operator—process deviation, CO03: aircraft operator—oversight failure}

Y = {CA20: 55 stabilizers, CA12: 32 landing gear, CH41: ground handling personnel—decision error}

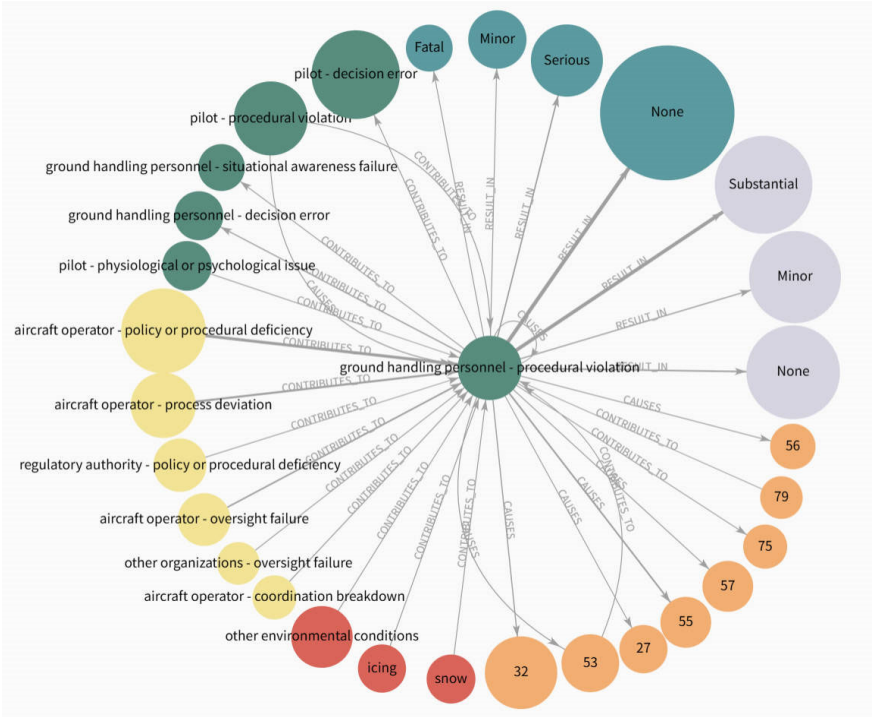


Figure 23. All in/out connections of ground handling personnel—procedural violation (CH42) in a causation-relation knowledge graph.

These results reveal a high-risk causation propagation chain centered on *CH42: ground handling personnel—procedural violation*, which can be represented as

aircraft operator—policy or procedural deficiency/process deviation/oversight failure

→ *ground handling personnel—procedural violation*

→ *55 stabilizers/ 32 landing gear/ground handling personnel—decision error*

This causation chain highlights that organizational deficiencies, including inadequate policies, procedural lapses, and oversight failures within the aircraft operator, are major contributors to procedural violations by ground handling personnel. In turn, such violations can lead to damage to critical aircraft systems such as stabilizers and landing gear, as well as negatively impact pilot decision-making. To prevent such accident chains, ground handling personnel must strictly comply with all relevant regulations and procedures. At the same time, aircraft operators should improve the integrity of their safety management systems by enhancing the completeness of regulations, the rationality of procedures, and the effectiveness of supervision.

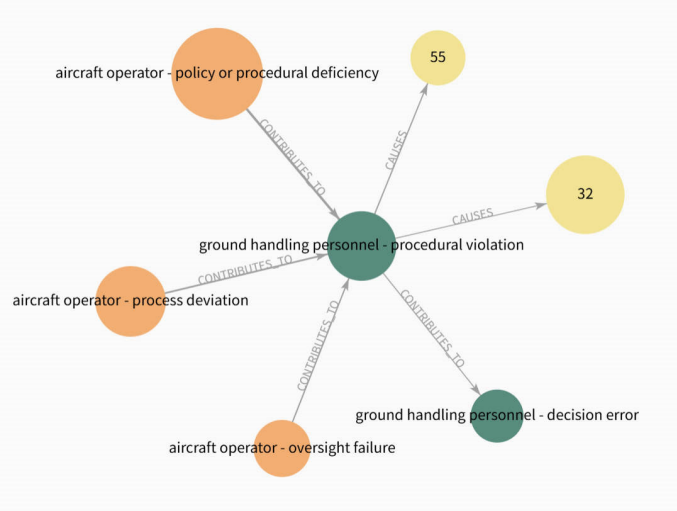


Figure 24. Top 3 in/out connections of ground handling personnel—procedural violation (CH42) in a causation-relation knowledge graph.

6. Conclusions

Given the diverse causative factors and complex propagation mechanisms of aviation accidents, this study leverages historical accident investigation reports as primary research data and integrates domain expertise with practical requirements to design a customized causation clustering schema. Utilizing PE and RAG—two advanced techniques derived from LLMs—the study extracts causation entities and their interrelationships and conducts in-depth causation analysis via KG modeling. The main contributions of this study are summarized as follows:

Development of an LLM-powered causation mining and clustering framework: A novel method integrating PE and RAG is proposed for extracting and clustering aviation accident causations and their relationships. From 343 accident investigation reports, the method identifies 102 unique causation entities and 560 types of inter-causation relationships. The clustering accuracy for the four major categories—Human, Aircraft, Environment, and Organization—reaches 95.8%, 86.5%, 97.9%, and 90.3%, respectively, with an overall accuracy of 92.6%.

Construction of a KG for causation network analysis: Based on the extracted causation entities and relationships, a causation-centric aviation accident KG is constructed.

A suite of risk evaluation metrics is proposed to quantitatively assess the centrality and propagation roles of various causations in the accident chain.

Design of a high-risk causation combination identification method: A specialized analysis framework is introduced to identify high-probability propagation chains originating from high-risk causations. This framework offers actionable insights and references for aviation accident risk management and mitigation. Among the four major causation categories, pilot decision error, landing gear system failure, bird strike, and aircraft operator policy or procedural deficiency are identified as the most frequent causations, respectively. Furthermore, key causation chains involving multiple categories—such as Aircraft operator—policy or procedural deficiency → pilot—decision error → 32 landing gear—have been discovered. These insights offer valuable reference points for understanding and interrupting potential accident pathways.

Overall, this study presents an efficient and automated aviation accident causation analysis methodology, which is underpinned by LLM technologies. By integrating advanced AI techniques with domain-specific knowledge, the proposed approach enables scalable and in-depth analysis of aviation accident data. Thus, this study offers valuable support for risk monitoring and control in aviation safety practices.

7. Future Work

Although the current work has made some progress, the method proposed in this paper still can be optimized and further investigated from the following aspects:

(1) **Extension of the causation clustering schema:** Current schema was designed from a high-level perspective, which means minor variance between different causations overlooked—a potential issue in practical applications. For example, the current aircraft causation schema categorizes causations only by system, which becomes inadequate when analyzing accident data tied to specific subsystems. To address this, the schema should be refined to include subsystem-level or even component-level granularity.

(2) **Extension of the knowledge graph:** The extension here encompasses two key aspects: accident report volume and information richness of nodes and edges. Applying the proposed method to analyze and incorporate more accident reports into the existing aviation accident knowledge graph can broaden the graph's scope and enhance its value for aviation accident analysis. Enriching node and edge information involves integrating pre-clustered causation descriptions and their source accident reports into nodes or edges.

While the current work demonstrates promising results, future research should prioritize refining the causation clustering schema (e.g., subsystem/component-level granularity) and expanding the knowledge graph (e.g., broader report coverage and richer node-edge metadata). These improvements would not only address existing limitations but also unlock deeper analytical capabilities, such as fine-grained causation tracing and cross-report pattern discovery. Ultimately, such advancements could transform the knowledge graph into a more robust tool for both retrospective accident analysis and proactive aviation safety planning.

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

Funding: This research received no external funding.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

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

Appendix A

Table A1. Prompt templates.

Task	Prompt Template
Preliminary causations	You are an expert in aircraft accident analysis. Your task is to analyze the causes of an aircraft accident from four aspects (human, aircraft, environment, and organizational) based on the provided factual narrative, analysis narrative and probable cause. Make sure that all returned results are returned in JSON format and correctly followed the given return requirements below: ===return requirements=== 1. The output must be a properly formatted JSON object using the following structure: {{ “human”: [“analysis of human factors”], “aircraft”: [“analysis of aircraft factors”], “environment”: [“analysis of environmental factors”], “organizational”: [“analysis of organizational factors”], }} 2. The “human” field should find out WHO performed WHAT faulty operations. If there is no “human” factors, return “unknown” with no additional explanations. 3. The “aircraft” field should specify WHICH component or system had WHAT failure or fault. If there is no “aircraft” factors, return “unknown” with no additional explanations. 4. The “environment” field must specify which extreme environmental conditions influenced specific aspects of the flight. These conditions are limited to [“Fog”, “Thunderstorm”, “Lightning”, “Heavy Rain”, “Icing”, “Snow”, “Wind Shear”, “Tailwind”, “Crosswind”, “Gusts”, “wet runway”, “bird strike”, “Sunglare”, “Freezing Rain”, “others”]. If there is no “environment” factors, return “unknown” with no additional explanations. 5. “organizational” field should state WHICH authority had WHAT shortcomings or failures. If there is no “organizational” factors, return “unknown” with no additional explanations. 6. All causes must—and only—be derived from the conclusions explicitly stated in the provided factual narrative, analysis narrative, and probable cause. No external assumptions or interpretations are permitted. 7. Do NOT include any additional explanation, commentary, or text outside of the JSON structure. 8. The output must include only the JSON object and no additional text before or after it. ===end of return requirements=== ===factual narrative=== {factual_narrative} ===analysis narrative=== {analysis_narrative} ===probable cause=== {probable_cause}

Table A1. *Cont.*

Task	Prompt Template
Root causation	You are an expert in aircraft accident analysis. Below is a list of known causes of an aircraft accident, along with the conclusion of the probable cause. Your task is to identify the **most directly responsible and conclusive cause** of the accident based on the given probable cause conclusion. Match it with the most relevant item in the cause list, and return its **ID** and **description** in the specified JSON format. Ensure that your output is: - Based solely on the probable cause conclusion. - Chosen from the provided cause list. - Returned strictly in the expected JSON format. ===Probable Cause Conclusion=== {probable_cause} === Cause List === {cause_list} === Expected Output Format === {{ “id”: “id of the final cause”, “description”: “description of the final cause” }}

Relations	You are given **accident information** and a list of **known causes**. Your task is to analyze and extract valid relationships between these causes according to the **Instructions** below.
	=== accident information ===
	“factualNarrative”:{factualNarrative}
	“analysisNarrative”:{analysis_narrative}
	===known cause===
	{cause_list}
	=== Cause ID Definitions ===
	Each cause ID is prefixed by a digit that represents its type:
	- “1x”: Human cause
	- “2x”: Aircraft cause
	- “3x”: Environmental cause
	- “4x”: Organizational cause
	=== Instructions ===
	1. Format the result as a list of dictionaries like this:
	[
	{{
	“source_cause_id1”: {{
	“relation_type1”: [target_cause_id1, target_cause_id2],
	“relation_type2”: [target_cause_id3]
	}}
	}},
	{{
	“source_cause_id2”: {{
	“relation_type1”: [target_cause_id4]
	}}
	}},
	...
	]
	2. Use **only the cause IDs** (e.g., “10”, “40”) for all keys and values.
	3. **Valid relation types** include:
	- ‘”causes”‘ — for strong direct causal links.
	- ‘”contributes to”‘ — for indirect or supporting contributions.
	- You **must not invent** custom relation types unless clearly supported by the evidence.

Table A1. *Cont.*

Task	Prompt Template
Relations	4. **STRICT ID RULES** — YOU MUST FOLLOW: - '"3x"' (environmental) and '"4x"' (organizational) causes can only appear as **source_cause**, not as targets if the source is a '"1x"' or '"2x"' cause. - Invalid: '"10": {{"causes": ["40"]}}' - Valid: '"40": {{"contributes to": ["10"]}}' - '"3x"' and '"4x"' causes can be **targets only** if the source is also a '"4x"' cause. - Valid: '"40": {{"contributes to": ["41"]}}' - '"1x"' and '"2x"' causes can be either sources or targets. - Never place '"3x"' or '"4x"' causes in the target list of '"1x"' or '"2x"' sources. 5. Do **not duplicate** any relationship in both directions. - If '"40"' contributes to '"10"', you must **not** also say '"10"' causes '"40"'. 6. Exclude any causes that have **no valid outgoing relationships**. - Do not include empty dictionaries such as '"41": {{}}'. 7. Your final output must be **only** the valid JSON result. Do not include any explanation.
Classification -CA	===Role Capability=== You are an aircraft fault diagnosis expert. Based on the input [Aircraft Fault Symptoms], analyze and identify the corresponding 4-digit system/component code and description from the [JOINT AIRCRAFT SYSTEM/COMPONENT CODE TABLE AND DEFINITIONS]. ===JOINT AIRCRAFT SYSTEM/COMPONENT CODE TABLE AND DEFINITIONS=== [{context}] ===Aircraft Fault Symptoms=== [{content}] ===Output Requirements=== 1. The returned result must be selected exclusively from the provided "JOINT AIRCRAFT SYSTEM/COMPONENT CODE TABLE AND DEFINITIONS". 2. Only one CODE AND TITLE should be returned without any additional explanation, formatting, or comments. Return the 4-digit code and corresponding title of the system or component involved in the aircraft fault.

Classification -CH	You are tasked with analyzing the human cause of an aircraft accident and categorizing it as a combination of one "personnel type" and one "factor type" from the provided [Human Cause Category List]. You may refer to the provided [personnel and factor definitions] to ensure precision in your classification.
	Please be aware that the return should be a combination of "personnel type" and "factor type" in [Human Cause Category List] instead of the detail definition in [personnel and factor definitions], for example, if you find out that the "first officer" had a "decision error", return "pilot—decision error" instead of "first officer—decision error".
	===Human Cause===
	{cause}
	===Human Cause Category List===
	{{
	"**personnel type**": ["pilot", "cabin crew", "maintenance personnel", "air traffic controller", "Ground handling personnel", "other personnel"],
	"**factor type**": ["decision error", " procedural violation", "skill deficiency", "physiological or psychological issue", "situational awareness failure", "other unsafe behavior"]
	}}
	===personnel and factor definitions===
	"personnel_type_definitions": {{
	"**pilot**": "The individual(s) responsible for operating and controlling the aircraft during flight, including the captain and first officer.",
	"**cabin crew**": "Personnel on board responsible for passenger safety and comfort, including flight attendants and in-flight service staff.",

Table A1. *Cont.*

Task	Prompt Template
Classification -CH	"**maintenance personnel**": "Technicians and engineers tasked with inspecting, repairing, and maintaining aircraft systems, components, and airworthiness.", "**air traffic controller**": "Ground-based personnel who provide instructions, clearances, and separation to aircraft for safe and efficient navigation and traffic management.", "**ground handling personnel**": "Staff responsible for servicing the aircraft on the ground, such as baggage handling, refueling, towing, catering, and ground equipment operations.", "**other personnel**": "Any other individuals involved in aviation operations not classified above, such as dispatchers, load planners, third-party vendors, or ramp security." }}, "factor_type_definitions": {{ "**decision error**": "A mistake arising from a poor judgment or incorrect choice made in planning, execution, or response to a situation.", "**procedural violation**": "A deliberate or unintended failure to adhere to established procedures, regulations, checklists, or standard operating protocols.", "**skill deficiency**": "Inadequate technical or motor skills required to correctly perform a task or operate equipment, often due to lack of training or experience.", "**physiological or psychological issue**": "Impairment due to fatigue, illness, stress, distraction, or emotional disturbance that affects performance or judgment.", "**situational awareness failure**": "A loss or degradation of understanding regarding the aircraft's status, environment, or intended flight path, leading to inappropriate actions.", "**other unsafe behavior**": "Any other improper or unsafe action not covered by the above categories, including complacency, distractions, or negligence." }} ===Return Requirements=== 1. The output must be exactly one combination: one "personnel type" + one "factor type" from the [Human Cause Category list]. 2. Return only the category combination as a string in the format: "<personnel type>—<factor type>". 3. Do not include any additional text, formatting, or comments.
Classification -CE	Your task is to classify the environmental cause of an aircraft accident using one environmental factor from the given [Environmental Cause Category List]. ===Environmental Cause=== {cause} ===Environmental Cause Category List=== ["Fog", "Thunderstorm", "Lightning", "Heavy Rain", "Icing", "Snow", "Wind Shear", "Tailwind", "Crosswind", "Strong Gusts", "Runway Contamination", "Bird Strike", "Sun Glare", "Other Environmental Conditions"] === Return Requirements === 1. The output must be exactly one category in the given Environmental Cause Category List. 2. Return only the category as a string without any additional text, formatting, or comments.
Classification -CO	You are tasked with analyzing the organizational cause of an aircraft accident and categorizing it as a combination of one "organization type" and one "factor type" from the provided list. You may refer to the provided [personnel and factor definitions] to ensure precision in your classification. ===Organizational Cause=== {cause}

Table A1. *Cont.*

Task	Prompt Template

Classification -CO	===Organizational Cause Category List===
	{{
	"**organization type**": ["Aircraft Operator", "Airport Operator", "Air Traffic Control Unit", "Manufacturer (Aircraft/Component)", " Regulatory Authority", "Other Organizations"]
	"**factor type**": ["Process Deviation", "Coordination Breakdown", "Oversight Failure", "Policy or Procedural Deficiency", "Resource Management Failure", "Other Organizational Cause"]
	}}
	===organization and factor definitions===
	"organization_type_definitions": {{
	"**Aircraft Operator**": "The organization responsible for operating and managing aircraft, including airlines and their management departments, accountable for flight safety and daily operations.",
	"**Airport Operator**": "The entity responsible for overall airport operations and management, covering runways, terminals, and ground services.",
	"**Air Traffic Control Unit**": "The organization that provides air traffic control services, ensuring safe separation and routing of flights.",
	"**Manufacturer (Aircraft/Component)**": "Companies that design and manufacture aircraft and their components, responsible for product design quality and airworthiness assurance.",
	"**Regulatory Authority**": "Government or administrative bodies that establish, oversee, and enforce aviation regulations and standards.",
	"**Other Organizations**": "Other organizations involved in aviation safety management not classified above, such as third-party service providers or training institutions."
	}},
	"factor_type_definitions": {{
	"**Process Deviation**": "Organizational or cross-organizational processes that are not performed according to established procedures or expectations, resulting in safety risks or errors.",
	"**Coordination Breakdown**": "Failures in effective communication, collaboration, or information sharing between departments or organizations leading to task execution failure.",
	"**Oversight Failure**": "Inadequate supervision, auditing, or inspection that fails to identify or correct potential safety issues in a timely manner.",
	"**Policy or Procedural Deficiency**": "Deficiencies, inadequacies, or absence of organizational policies, rules, or operating procedures that impact safety assurance.",
	"**Resource Management Failure**": "Insufficient or mismanaged allocation of human resources, materials, funding, or time affecting organizational operations and safety.",
	"**Other Organizational Cause**": "Other organizational-level factors not covered above that negatively impact safety through behaviors or conditions."
	}}
	===Return Requirements===
	1. The output must be exactly one combination: one "organization type" + one "factor type" from the [Organizational Cause Category List].
	2. Return only the category combination as a string in the format: "<organization type>—<factor type>".
	3. Do not include any additional text, formatting, or comments.

References

1. ICAO. The Annual ICAO Safety Report 2024. International Civil Aviation Organization (ICAO). 2024. Available online: https://www.icao.int/sites/default/files/sp-files/safety/Documents/ICAO_SR_2024.pdf (accessed on 10 June 2025).
2. Cirium. Airline Safety and Losses Annual Review 2024. Cirium. 2024. Available online: https://assets.fta.cirium.com/wp-content/uploads/2025/03/11112122/Airline-Safety-Review-2024-prmc.pdf (accessed on 10 June 2025).

3. Tanguy, L.; Tulechki, N.; Urieli, A.; Hermann, E.; Raynal, C. Natural language processing for aviation safety reports: From classification to interactive analysis. *Comput. Ind.* **2016**, *78*, 80–95. [CrossRef]
4. Subramanian, S.V.; Rao, A.H. Deep-learning based Time Series Forecasting of Go-around Incidents in the National Airspace System. In Proceedings of the 2018 AIAA Modeling and Simulation Technologies Conference (AIAA 2018-0424), Kissimmee, FL, USA, 8–12 January 2018.
5. Zhou, D.; Zhuang, X.; Zuo, H.; Cai, J.; Zhao, X.; Xiang, J. A model fusion strategy for identifying aircraft risk using CNN and Att-BiLSTM. *Reliab. Eng. Syst. Saf.* **2022**, *228*, 1. [CrossRef]
6. Bleu-Laine, M.-H.; Puranik, T.G.; Mavris, D.N.; Matthews, B. Predicting adverse events and their precursors in aviation using multi-class multiple-instance learning. In Proceedings of the AIAA Scitech 2021 Forum, Online, 11–15 and 19–21 January 2021; p. 0776.
7. Kojima, T.; Gu, S.S.; Reid, M.; Matsuo, Y.; Iwasawa, Y. Large language models are zero-shot reasoners. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 22199–22213.
8. Lewis, P.; Perez, E.; Piktus, A.; Petroni, F.; Karpukhin, V.; Goyal, N.; Küttler, H.; Lewis, M.; Yih, W.-t.; Rocktäschel, T.; et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In *Advances in Neural Information Processing Systems*; Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M.F., Lin, H., Eds.; Curran Associates, Inc.: New York, NY, USA, 2020; Volume 33, pp. 9459–9474.
9. Liu, C.; Yang, S. Using text mining to establish knowledge graph from accident/incident reports in risk assessment. *Expert Syst. Appl.* **2022**, *207*, 117991. [CrossRef]
10. Xiong, M.; Wang, H.; Wong, Y.D.; Hou, Z. Enhancing aviation safety and mitigating accidents: A study on aviation safety hazard identification. *Adv. Eng. Inform.* **2024**, *62*, 102732. [CrossRef]
11. Brooker, P. Experts, Bayesian Belief Networks, rare events and aviation risk estimates. *Saf. Sci.* **2011**, *49*, 1142–1155. [CrossRef]
12. Zhang, H.; Wang, Q. Risk identification model of aviation system based on text mining and risk propagation. *Eksploat. I Niezawodn. Maint. Reliab.* **2025**, *27*, 192767. [CrossRef]
13. Zhang, X.; Srinivasan, P.; Mahadevan, S. Sequential deep learning from NTSB reports for aviation safety prognosis. *Saf. Sci.* **2021**, *142*, 105390. [CrossRef]
14. De Vries, V. Classification of aviation safety reports using machine learning. In Proceedings of the 2020 International Conference on Artificial Intelligence and Data Analytics for Air Transportation (AIDA-AT), Singapore, 3–4 February 2020; pp. 1–6.
15. Wang, X.; Gan, Z.R.; Xu, Y.X.; Liu, B.N.; Zheng, T. Extracting Domain-Specific Chinese Named Entities for Aviation Safety Reports: A Case Study. *Appl. Sci.-Basel* **2023**, *13*, 11003. [CrossRef]
16. Olive, X.; Basora, L. Detection and identification of significant events in historical aircraft trajectory data. *Transp. Res. Part C: Emerg. Technol.* **2020**, *119*, 102737. [CrossRef]
17. Baigang, M.; Yi, F. A review: Development of named entity recognition (NER) technology for aeronautical information intelligence. *Artif. Intell. Rev.* **2022**, *56*, 1515–1542. [CrossRef]
18. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 261–272.
19. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, USA, 2–7 June 2019; pp. 4171–4186.
20. Liu, Z.; Lin, W.; Shi, Y.; Zhao, J. A robustly optimized BERT pre-training approach with post-training. In Proceedings of the China National Conference on Chinese Computational Linguistics, Hohhot, China, 13–15 August 2021; pp. 471–484.
21. Reimers, N.; Gurevych, I. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv* **2019**, arXiv:1908.10084. [CrossRef]
22. Chandra, C.; Ojima, Y.; Bendarkar, M.V.; Mavris, D.N. Aviation-BERT-NER: Named Entity Recognition for Aviation Safety Reports. *Aerospace* **2024**, *11*, 890. [CrossRef]
23. Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A. Language models are few-shot learners. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 1877–1901.
24. Hu, E.J.; Shen, Y.; Wallis, P.; Allen-Zhu, Z.; Li, Y.; Wang, S.; Wang, L.; Chen, W. Lora: Low-rank adaptation of large language models. *ICLR* **2022**, *1*, 3.
25. Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; Xia, F.; Chi, E.; Le, Q.V.; Zhou, D. Chain-of-thought prompting elicits reasoning in large language models. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 24824–24837.
26. Kıcıman, E.; Ness, R.; Sharma, A.; Tan, C. Causal Reasoning and Large Language Models: Opening a New Frontier for Causality. *arXiv* **2024**. [CrossRef]
27. Liu, Q.L.; Li, F.; Ng, K.K.H.; Han, J.S.; Feng, S.S. Accident investigation via LLMs reasoning: HFACS-guided Chain-of-Thoughts enhance general aviation safety. *Expert Syst. Appl.* **2025**, *269*, 126422. [CrossRef]

28. Chen, L.; Xu, J.H.; Wu, T.Y.; Liu, J. Information Extraction of Aviation Accident Causation Knowledge Graph: An LLM-Based Approach. *Electronics* **2024**, *13*, 3936. [CrossRef]
29. Ren, T.F.; Zhang, Z.P.; Jia, B.; Zhang, S.W. Retrieval-Augmented Generation-aided causal identification of aviation accidents: A large language model methodology. *Expert Syst. Appl.* **2025**, *278*, 127306. [CrossRef]
30. Qu, J.Y.; Wang, J.T.; Zhao, Z.Y.; Chen, X.G. MBJELEL: An End-to-End Knowledge Graph Entity Linking Method Applied to Civil Aviation Emergencies. *Int. J. Comput. Intell. Syst.* **2024**, *17*, 237. [CrossRef]
31. Xu, J.H.; Chen, L.; Xing, H.X.; Tian, W.J. Causation Correlation Analysis of Aviation Accidents: A Knowledge Graph-Based Approach. *Appl. Sci.* **2024**, *14*, 6887. [CrossRef]
32. Han, H.; Wang, Y.; Shomer, H.; Guo, K.; Ding, J.; Lei, Y.; Halappanavar, M.; Rossi, R.A.; Mukherjee, S.; Tang, X.; et al. Retrieval-Augmented Generation with Graphs (GraphRAG). *arXiv* **2025**. [CrossRef]
33. Chen, N.; Sun, Y.; Wang, Z.; Peng, C. Identification of flight accidents causative factors base on SHELLO and improved entropy gray correlation method. *Heliyon* **2023**, *9*, e13534. [CrossRef]
34. NTSB. Aviation Coding Manual. NTSB. 1999. Available online: https://www.ntsb.gov/GILS/Documents/codman.pdf/ (accessed on 10 June 2025).

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.