Ocean Engineering 363 (2026) 126862

Contents lists available at ScienceDirect

Ocean Engineering

journal homepage: www.elsevier.com/locate/oceaneng

Research paper

Uncovering patterns in ferry near misses: A topic modelling approach with GPT-assisted labelling

Hollie Black [a,*], Louis de Wolff [b], Yasin Burak Kurt [a], Yaser Farag [a], Dwitya Waskito [a], Rafet Kurt [a], Osman Turan [a]

[a] University of Strathclyde, Department of Naval Architecture, Ocean and Marine Engineering, Henry Dyer Building, 100 Montrose Street, Glasgow, Scotland, G4 0LZ, UK
[b] CalMac Ferries Limited, The Ferry Terminal, Station Road, Gourock, Scotland, PA19 1QP, UK

ARTICLE INFO

Keywords:
Maritime incidents
Near-misses
Topic modelling
Natural language processing
Accident prevention

ABSTRACT

Near-miss incident reports provide valuable but underutilised insights for improving maritime safety. However, existing research emphasises failures, with limited focus on positive safety practices. This study addresses this gap by examining connections between incidents and preventive measures to identify what works and what doesn't. To overcome the limitations of traditional topic modelling approaches, this study applies BERTopic, a transformer-based method. Using this method, we analysed 4360 near-miss reports spanning five years from a ferry operator to identify recurring risk patterns and the preventative measures linked to them. GPT-assisted labelling improved interpretability, with outputs validated through structured multi-expert review. Over 75 distinct incident topics were identified across operational, environmental, and human factors domains, aligning with existing categories and uncovering previously unclassified risk themes. Time-series analysis highlighted seasonal variations, with passenger-related risks peaking in summer and infrastructure failures more common in winter. Prevention measures were modelled separately, producing 87 topics, with association rule mining identifying clear linkages between incident types and mitigations; for example, navigation-related incidents were frequently linked to improved bridge communication. Overall, combining unsupervised topic modelling with LLM-assisted labelling produces interpretable and operationally meaningful outputs. This approach offers a scalable framework for prioritising preventative measures using data-driven evidence.

1. Introduction

Near-miss reporting is widely recognised as a cornerstone of proactive safety management, providing early signals of systemic weaknesses before accidents occur (Heinrich, 1941; Simon et al., 1999). In the maritime domain, particularly with ferry operations that combine high passenger volumes with constrained harbour environments, learning from near misses is critical to reducing risk exposure.

Existing classification systems, such as the International Maritime Organisation (IMO)'s casualty reports or company-level taxonomies, rely heavily on predefined categories that may overlook nuanced patterns in human error, environmental conditions, and operational interactions. Moreover, lessons learned and preventative measures are usually left in textual format, making it difficult to extract meaningful trends without resource-intensive manual review. As a result, organisations often fail to capitalise on the predictive value of near-miss data, limiting their capacity to anticipate and prevent future incidents.

In recent years, advances in natural language processing (NLP) have opened new opportunities for systematic analysis of unstructured safety reports. Topic modelling methods have been applied across domains to uncover latent themes in large text corpora, allowing researchers to identify risk patterns without predefined labels. Traditional approaches such as Latent Dirichlet Allocation (LDA) have seen use in safety research but are constrained by their reliance on bag-of-words assumptions and parameter sensitivity (Lu and Cai, 2024). More recently, transformer-based models such as BERTopic have demonstrated superior performance by leveraging contextual embeddings and density-based clustering to produce semantically coherent topics (Egger and Yu, 2022). Unlike traditional topic modelling methods, BERTopic captures contextual relationships between words and phrases, making it particularly suitable for analysing complex safety narratives. These developments make advanced topic modelling a promising tool for maritime safety research.

At the same time, large language models (LLMs) such as GPT have

* Corresponding author. Department of Naval Architecture, Ocean and Marine Engineering, University of Strathclyde, Glasgow, Scotland, UK.
E-mail address: hollie.black.2017@uni.strath.ac.uk (H. Black).

https://doi.org/10.1016/j.oceaneng.2026.126862
Received 27 April 2026; Received in revised form 18 June 2026; Accepted 2 July 2026
0029-8018/© 2026 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

emerged as powerful tools for interpreting and labelling topics generated by unsupervised models (Costantino and Sabetta, 2025). While unsupervised clustering produced keywords that require expert interpretation, GPT can assist by generating candidate labels and grouping themes in a way that aligns with human reasoning. Recent studies have begun exploring the integration of LLMs with topic modelling pipelines, reporting improvements in topic interpretability and the discovery of cross-cutting themes. However, to date, no studies have combined BERTopic and GPT-assisted interpretation for analysing near-miss data.

This study addresses that gap by applying BERTopic and GPT-assisted labelling to a comprehensive dataset of near-miss reports from a ferry operator. The analysis aims to identify patterns in incident types, preventive measures, and their temporal dynamics, with validation provided by subject matter experts. Beyond identifying emergent risks, the study contributes a novel methodology framework for linking incident types to corresponding prevention actions, enabling data-driven recommendations for proactive safety planning. Since large-scale safety datasets typically require substantial expert input, extensive manual effort, and consistent classification frameworks, this can limit the speed and scalability of insights. By automating topic discovery and supporting expert-guided labelling, the proposed approach reduced these burdens and offers new opportunities for more timely, consistent, and comprehensive analysis.

The study makes three contributions. First, it demonstrates that transformer-based topic modelling, supported by GPT-assisted labelling, can extract operationally meaningful safety themes from large-scale ferry near-miss narratives, revealing contextual and activity-specific insights that are not fully represented within existing classification systems. Second, it introduces an integrated analytical framework that combines topic modelling and association rule mining to systematically link incident themes with corresponding prevention measures, providing empirical evidence of the adaptive actions, coordination mechanisms, and prevention controls through which safe ferry operations are maintained in practice. In doing so, it demonstrates how near-miss reporting systems can contribute not only to understanding why safety fails, but also to understanding how safety succeeds.

The structure of this manuscript is as follows. Section 2 reviews the current literature on the development of topic modelling and the utilisation of LLM in maritime safety analysis. Section 3 outlines the research framework, the explanation regarding data, BERTopic, and validation. Section 4 presents the main results of this study. Section 5 discusses the findings, implications, limitations and future research directions. Section 6 states the conclusion of this study.

2. Literature review

2.1. Maritime safety and near misses

Several studies have highlighted the multi-layered nature of causation in maritime accidents and near misses, where technical, human, organisational, and environmental factors interact in complex ways. Near-miss reporting is widely recognised as a critical element of proactive safety management across high-risk industries (Heinrich, 1941). In maritime operations, near misses serve as leading indicators for accident prevention, providing early warnings of systemic vulnerabilities before they escalate into major incidents (Hasanspahic et al., 2022; Madsen et al., 2025). The International Maritime Organisation (IMO) and International Safety Management (ISM) code encourage near-miss reporting as part of continuous improvement cycles (Guidance on Near, 2008).

Despite this, research consistently shows that the utility of near-miss data is undermined by chronic underreporting, inconsistent categorisation and limited analytical frameworks (Hassel et al., 2011; Oltedal and McArthur, 2011). Barriers include cultural reluctance to report, fear of blame, administrative burden, and doubts about feedback or learning value (Hasanspahi et al., 2020). As a result, many near-miss databases remain reactive records rather than proactive learning tools.

These challenges are particularly salient in ferry operations, which present a distinctive risk environment due to their dual role as passenger-vehicle carriers operating in constrained coastal and port settings. High-profile catastrophes, such as the Herald of Free Enterprise, have received detailed forensic investigation (Njoku, 2021). However, near-miss data in ferry contexts remain unexplored despite their predictive value for proactive risk management.

Comparative scoping studies have further highlighted the systemic nature of domestic ferry safety risks. Golden and Weisbrod (Golden and Weisbrod (2016), who compiled a dataset of 232 ferry accidents worldwide between 2000 and 2014, showing that 94% occurred in low- and middle-income countries. Their analysis found that human error contributed to between 62 and 86% of cases depending on coding assumptions, while hazardous weather played a role in 50% and overloading in nearly a third.

Regional case studies illustrate how multi-factorial causes manifest in domestic ferry operations. Njoku (2021) assessed ferry mishaps in Lagos State, Nigeria using the SHEL framework. Survey results from operators, passengers, and regulators showed that poor maintenance, engine failures, weak enforcement of safety rules, operator fatigue, negligence, and abandoned wrecks in navigation channels were the most significant causes of ferry mishaps.

In contrast, more recent work has attempted to address the limitations in near-miss data utilisation by developing a hybrid analytical framework. Bicen and Celik (2022) combined the Human Factors Analysis and Classification System (HFACS) with the Analytic Network Process (ANP) to systematically cluster causes and prioritise prevention actions. Their study analysed 2828 near-miss reports from 25 shipping companies and demonstrated how structured classification combined with multi-criteria decision methods can produce non-conventional and next-generation safety solutions, including integration of information and communication technologies (ICT) and safety engineering perspectives. Importantly, they argue that near-miss data, when systematically analysed, offers a proactive route to safety improvement rather than serving only as a reactive record.

Beyond causal analyses, numerous studies have evaluated human error and human performance as central contributors to maritime accidents. Human factors research in this field demonstrates that unsafe outcomes rarely stem from a single act or error, but from interactions among individuals, teams, organisations, and environmental conditions (Coraddu et al., 2020; Stroeve et al., 2023). Due to this, other lines of research have expanded to human-centred risk assessment, using quantitative and qualitative measures to evaluate how factors such as situational awareness, fatigue, stress and coordination influence operational reliability and risk perception (Ahn et al., 2022; Ahn and Kurt, 2020). Research has increasingly employed structured risk assessment methodologies to identify and prioritise hazards within critical shipboard systems. Fuzzy FMEA approaches have been applied to systems including hydraulic steering mechanisms and compressor systems, enabling the prioritisation of failure modes under uncertainty and supporting maintenance planning, reliability improvement and accident prevention (Velibor et al., 2024; BulutOzan, 2023). In parallel, system-based accident analysis methodologies to investigate accident causation within complex socio-technical environments have been increasingly adopted. Approaches such as Functional Resonance Analysis Method (FRAM) and Systems-Theoretic Accident Model and Processes (STAMP) have been applied to maritime incidents, including tanker fires and catastrophic engine failures, demonstrating how accidents emerge through interactions among human, technical, organisational, and environmental factors rather than from isolated failures (SalihHü et al., 2026; Ceylan et al., 2022). Similarly, studies examining complex maritime casualties have highlighted the importance of analysing multiple interacting root causes and system-level influences when investigating accident occurrence (Bulut et al., 2023). Together, these approaches reflect a shift away from traditional linear accident models

towards a more holistic understanding of safety performance in maritime operations.

Crucially, current research often fails to capture preventative factors and positive system performance, the conditions and actions that avert incidents even under high-risk situations. This aligns with the principles of Safety-II, which emphasise understanding why operations succeed and how everyday performance variability contributes to resilience, not only why they fail (Hollnagel, 2014).

2.2. Maritime safety and natural language processing

Maritime research has also increasingly adopted NLP-based approaches to extract structured safety knowledge from unstructured textual reports. Lichao et al. (Lichao et al., 2026) developed an automated framework integrating NLP and complex network theory to transform 564 maritime accident investigation reports into a quantitative risk network, enabling the identification of key causal factors and dynamic risk propagation pathways. Their findings demonstrated the value of large-scale text analytics for understanding systemic accident evolution and identifying intervention targets. However, the focus remained on accident causation and propagation mechanisms, with limited attention given to extracting recurring themes and prevention actions from near-miss narratives.

Similarly, Dongge et al. (Dongge et al., 2026) employed domain-adapted MBERT models to automatically extract accident characteristics, human factors, environmental conditions, and management deficiencies from 581 maritime investigation reports before linking these factors to casualty severity outcomes through ordered logistic regression. Their study further highlighted the growing role of NLP techniques in transforming unstructured maritime safety reports into actionable safety insights capable of supporting evidence-based decision making. Black et al. (Black et al., 2026) also employed a domain-adapted BERT trained on the SHIELD taxonomy to automatically classify fine-grained human and organisational factors from accident narratives. Their findings demonstrated the potential of domain-specific language models and association rule mining to support scalable organisational learning from maritime investigations while reducing the burden of manual coding.

Collectively, these studies demonstrate the growing adoption of NLP techniques for transforming unstructured maritime safety narratives into structured safety knowledge. However, existing approaches predominantly rely on supervised classification, predefined analytical objectives, or causal modelling frameworks. Comparatively less attention has been devoted to the unsupervised exploration of large collections of near-miss narratives to identify latent safety themes and understand how these themes relate to preventative actions. This limitation motivates the application of topic modelling techniques within maritime safety research.

2.3. Topic modelling in safety domains

Text mining and topic modelling have emerged as promising tools for extracting knowledge from unstructured safety data. Early applications in aviation and healthcare demonstrated that latent Dirichlet allocation (LDA) can surface themes beyond predefined taxonomies and support downstream analytics and visualisation (Kuhn, 2018; Rose et al., 2022; Mertes et al., 2024; Xing et al., 2024).

That said, traditional topic models such as LDA face well-documented limitations on technical, safety-critical text. Bag-of-words assumptions suppress semantic relations, which can lead to incoherent or overlapping topics and difficulty distinguishing contextually nuanced content. Numerous studies therefore explore improvements via semi-supervision (seed or anchor words), domain knowledge, or hybrids that stabilise topics and improve interpretability (Ahadh et al., 2021; Bao et al., 2023).

In maritime contexts, topic-modelling and related text-mining have been applied to accident or risk narratives to support investigation and risk coupling analysis, illustrating the approach's potential to augment safety management systems (Yan et al., 2023). However, the study employed Bidirectional Encoder Representations from Transformers (BERT), a transformer-based natural language processing model that captures contextual relationships between words, within a hybrid NLP framework rather than a topic modelling-based approach.

More recently, transformer-based topic modelling, such as BERTopic, have begun to appear in related maritime security studies, including analyses of piracy and armed robbery incident reports. These applications demonstrate the feasibility of transformer-based methods and their ability to produce clearer semantic groupings of contributory factors in piracy and armed robbery incidents (Fahreza and Hirata, 2024). BERTopic is one of the most widely used transformer-based approaches, utilising contextual embeddings, dimensionality reduction, and density-based clustering to produce topics that are more coherent and domain-interpretable compared to the LDA baseline. The method has been applied across many industries including, railway and wildfire incident analysis (Andrade and Walsh, 2023; Hong et al., 2024; Irshad and Walsh, 2024).

Despite this progress, applications of advanced NLP to maritime safety remain sparse. Much of the literature relies on structured accident databases or keyword-based analyses; few studies systematically analyse unstructured maritime near-miss narratives at scale using transformer-based topic models.

2.4. LLMs for topic interpretation

While topic modelling can reveal latent structures in text, interpreting clusters remains a challenge. Assigning meaningful labels to topics traditionally requires expert review, which is resource-intensive and prone to inconsistency. Large language models (LLMs) such as GPT-3 and GPT-4 have shown promise in bridging this gap by generating coherent, human-readable labels from unsupervised outputs.

Recent studies in social science and organisational research report that integrating LLMs into topic modelling pipelines improves interpretability, facilitates cross-cutting theme identification, and reduces expert burden (Emil et al., 2023). Moreover, experiments show that LLM-assisted labelling can enhance alignment between model outputs and domain expert reasoning, particularly in safety-critical applications where clarity of categories is paramount (Preiss et al., 2024).

Despite this growing literature, the combination of BERTopic with GPT-assisted labelling has not yet been applied to maritime near-miss data. Nor have previous studies attempted to link incident themes directly to reported prevention actions through data-driven association methods.

This study addresses these gaps by integrating advanced NLP and LLM techniques to uncover actionable insights from a large corpus of ferry near-miss reports, offering both methodological innovation and practical relevance for proactive maritime safety management.

2.5. Association rule mining (ARM) in safety domains

Association Rule Mining (ARM) is a data mining technique designed to uncover frequent co-occurrence patterns within large datasets. Originally developed for market basket analysis, ARM has been increasingly applied in safety research to identify hidden relationships between hazards, contributing factors, and outcomes.

In safety challenged industries ARM is usually paired with Bayesian Networks (BNs) to enhance causal inference. For example, Yuan et al. (Yuan et al., 2025) applied ARM and BNs to small-scale coal mining accidents, constructing a structured causal map transforming frequent co-occurrence into probabilistic dependencies. In the maritime domain, Bairami-Khankandi et al. (Bairami-Khankandi et al., 2025) employed ARM in combination with Causal Analysis based on System Theory (CAST) to identify systemic precursors. These studies show that ARM

can generate building blocks for Bayesian causal networks, linking descriptive patterns mining with predictive risk reasoning.

More limited applications have demonstrated the utility of ARM in combination with topic modelling. Sahoo et al. (Sahoo et al., 2024) combined ARM with LDA and fault tree analysis to investigate steel plant incidents, showing how textual themes could be linked through association rules into accident chains.

Most ARM applications in safety research rely on structured variables or coded datasets, while topic modelling is often deployed in isolation to map thematic distributions. The limited integration of these two approaches represents a methodological gap. Where topic modelling can provide semantically rich inputs to ARM, enhancing its ability to surface interpretable and domain-relevant rules for unstructured narratives.

3. Methodology

This study followed a multi-stage analytical pipeline combining transformer-based topic modelling, GPT-assisted interpretation, expert validation, temporal analysis, and association rule mining (ARM). The approach was designed to extract interpretable themes from unstructured near-miss narratives and link them systematically to recorded prevention measures.

3.1. Data

This section of the research utilises a comprehensive dataset of 4360 near-miss incidents sourced from a Ferry operator, covering a range of operational scenarios across marine and harbour environments. The data provides rich contextual information, which enables a multifaceted exploration of safety incidents, including the reason the near miss did not turn into an accident. The dataset contains the following information:

1. Incident metadata: including a unique near miss ID, incident date and time, and detailed vessel-specific information including asset name and region.
2. Open Text: "Near miss description" detailing what happened, "what in your opinion prevented it from developing into an accident?" detailing the prevention. These fields are central to topic modelling efforts and allow for the extraction of themes and contextual relationships through NLP.
3. Categorisations: "Near miss category" and "Near miss sub category" offer structured classification of the event type, which were used as a validation of the BERTopic classifications.
4. Other details: The database also contains information such as weather conditions, visibility, wind direction and strength, sea state, and impact score. Additionally, there are further details; however, the majority of these details are only completed for a very small number of incidents.

As BERTopic relies on transformer-based embeddings, traditional text preprocessing (e.g. stop word removal or lemmatisation) is unnecessary and may even reduce semantic quality. We therefore used only minimal cleaning prior to modelling. This included the removal of exact duplicate records, correction of character encoding issues, standardisation of whitespace and punctuation artefacts, and the exclusion of records with missing narrative contexts.

Abbreviations, spelling inconsistencies, and maritime-specific terminology were retained in their original form. This decision was made because transformer-based language models are generally robust to minor spelling variations and domain-specific vocabulary, and excessive normalisation may remove operational nuances present within safety narratives. Similarly, no manual synonym mapping or maritime terminology standardisation was applied. Very short reports were not removed; however, many of these records were subsequently identified as outliers by the HDBSCAN clustering stage due to insufficient semantic content for reliable clustering. This approach ensured that the original meaning and operational context of the reports were preserved while reducing obvious sources of textual noise.

3.2. Topic modelling

The central approach was BERTopic, selected for its ability to leverage transformer-based embeddings to extract semantically rich topics from unstructured text. The overall analytical workflow is illustrated in Fig. 1, which summarises the data, modelling approach and analysis methods used in this study. The pipeline involves four components:

1. Embedding model: Sentence-BERT (SBERT), an extension of BERT designed to generate meaningful sentence-level embeddings, was used to convert each incident description into high-dimensional embeddings that preserve semantic context. The specific model used was "all-MiniLM-L6-v2", which produced 384-dimensional sentence embeddings. This model is optimised for semantic similarity tasks, offering a balance between performance and computational efficiency. No fine-tuning or custom tokenisation was applied, and the text was passed directly to the transformer model after minimal non-linguistic cleaning.
2. Dimensionality reduction: Uniform manifold approximation and projection (UMAP) reduced the embedding dimensionality to enable effective clustering. This was applied with BERTopic's default parameters (n_neighbours = 15, n_components = 5, min_dist = 0.0, metric = 'cosine'), as these settings are widely adopted in the BERTopic literature and are recommended by the framework developers for preserving local semantic structure while producing compact representations suitable for density-based clustering. No additional hyperparameter optimisation was performed because the objective of the study was exploratory topic discovery rather than maximising predictive performance.
3. Clustering: Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) was employed to detect dense regions in the embedding space, automatically determining the number of clusters and labelling outliers. Again, the default parameters (min_cluster_size = 10, min_samples = None, cluster_selection_method = 'eom') were set. In our dataset, 1656 incident descriptions and 1287 prevention narratives were labelled as outliers. These texts were not then forced into clusters, as doing so would artificially inflate topic coherence and introduce noise into the interpretation. The relatively large number of outliers is partly attributable to the clustering strategy employed. HDBSCAN adopts a conservative approach to cluster assignment and intentionally labels low-density or weakly connected documents as noise rather than forcing them into potentially inappropriate topics. This behaviour is widely recognised in BERTopic applications and represents a trade-off between maximising document coverage and maintaining topic coherence. In the present dataset, the effect was amplified by the performance of short, highly specific and operationally diverse near-miss narratives, many of which lacked sufficient semantic similarity to form stable clusters.
4. Topic representation: Class-based TF-IDF (c-TF-IDF) extracted representative keywords from each cluster, providing interpretable descriptions of topic content:

$$W_{x,c} = \|tf_{x,c}\| \log\left(1 + \frac{N}{f_x}\right)$$

where $tf_{x,c}$ is the frequency of word x in class c, f_x is the frequency of word x across all classes, and N is the average number of words per class. This is a weighing scheme that quantifies how important a word is within a specific document relative to the entire corpus. The term fre-

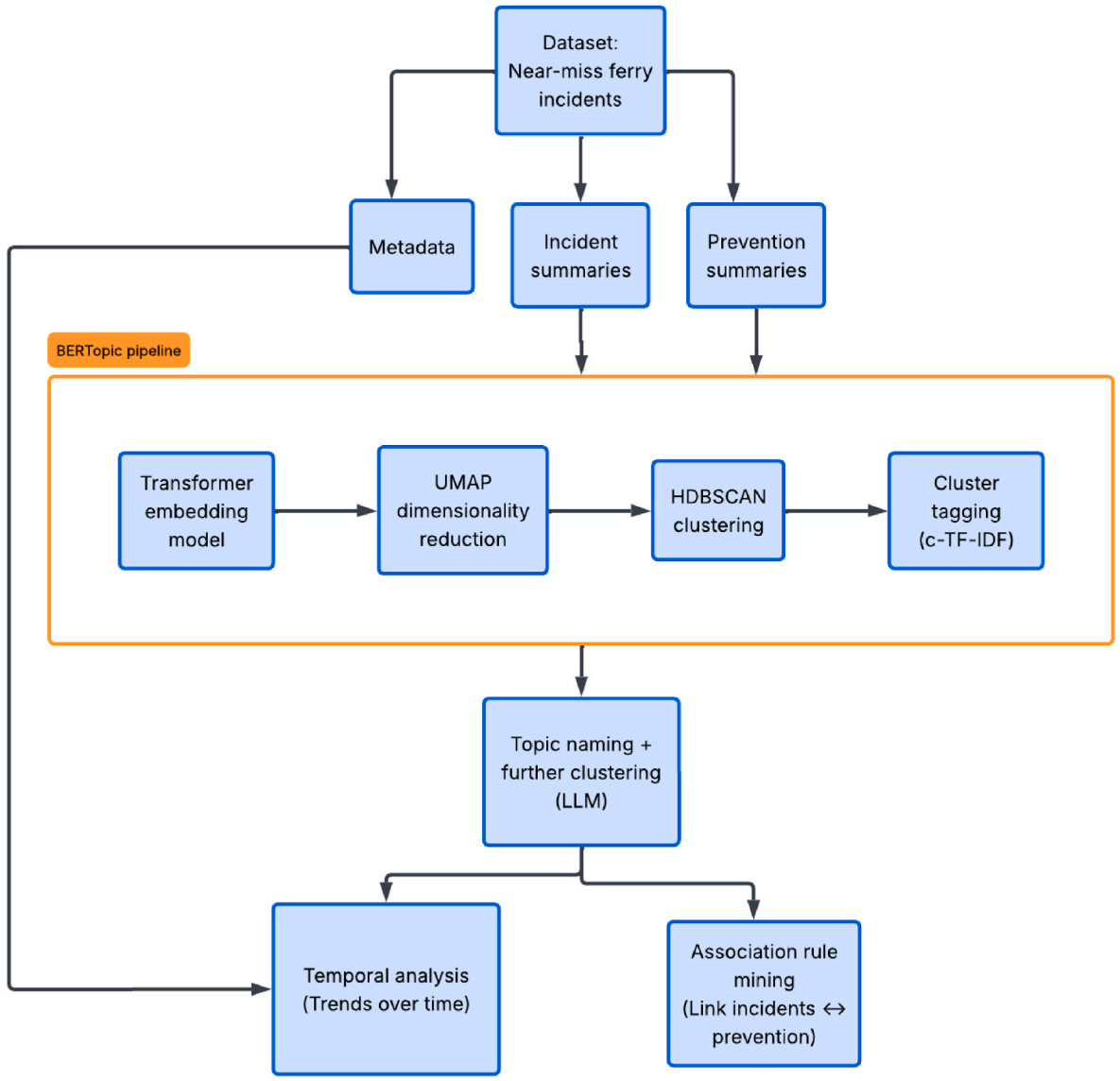

Fig. 1. Methodology flow diagram.

quency component assigns higher scores to words that appear frequently within a document.

The BERTopic model was run independently on the near-miss descriptions and the prevention summaries, enabling comparison between incident characteristics and associated prevention actions.

3.3. Topic labelling

While BERTopic generates representative keywords for each cluster, these outputs are often not sufficiently interpretable for direct use, particularly in safety-critical contexts where clear consistent terminology is required. To address this, GPT-4 was integrated into the workflow as a programmatic labelling component using the OpenAI API, enabling scalable and consistent topic interpretation.

The API was used to systematically process each cluster within the pipeline. For every cluster, the model was provided with the top-n keywords generated by the c-TF-IDF model and a set of representative incident descriptions (selected based on proximity to the cluster centroid in embedding space)

A one-shot prompt engineering approach was implemented within the codebase. The prompt structure was standardised and automatically applied to each cluster, consisting of: (1) a short explanation of the labelling task, (2) an illustrative example pairing keywords and sample texts with an ideal label, and (3) the cluster-specific inputs. The API calls were configured with deterministic parameters (temperature = 0) to minimise variability and ensure reproducibility of outputs across runs.

This setup allowed GPT-4 to act as a controlled natural language interpretation layer, transforming keyword clusters into concise, human-readable topic labels in a consistent and scalable manner. The outputs were parsed and stored programmatically, allowing integration with downstream analysis steps.

Importantly, GPT-4 was not used to replace expert judgement. All generated labels were subsequently reviewed and validated by domain experts, which will be explained in the concurrent sections.

3.4. Expert validation

To validate the topic modelling output, each GPT-labelled topic was first manually mapped to the nearest company-defined incident type, based on subject matter alignment and terminology. This step ensured that the derived topics could be meaningfully interpreted within an existing operational framework.

A formal expert validation process was then conducted to assess the quality and reliability of the topic model outputs. Four subject matter

experts in maritime safety participated in the evaluation. All experts held a minimum of master's-level qualifications and had substantial professional experience in the maritime domain.

The number of experts was selected based on recommendations from validation studies in topic modelling, qualitative content analysis, and expert judgement research, which indicate that panels of between three and ten domain experts are generally sufficient for evaluating thematic coherence, content validity, and interpretability when participants possess substantial subject-matter expertise. The objective of the validation process was not to estimate population-level opinion, but to assess whether the generated topics, labels, and representative documents were operationally meaningful within a maritime safety context. Using multiple experts with diverse professional experience allowed both consensus and disagreement to be quantified through formal inter-rater reliability measures while maintaining a manageable review workload.

Each topic was evaluated using three main criteria:

1. Keyword coherence: Whether the keywords formed a semantically meaningful grouping.
2. Label appropriateness: Whether GPT-generated names accurately described the theme.
3. Label-document fit: Whether representative documents were aligned with the assigned topic.

A hybrid validation strategy was adopted. A set of 26 core topics was reviewed by all experts to enable inter-rater reliability analysis, while an additional 10- 12 topics were distributed across individual reviewers to expand coverage without increasing cognitive burden.

To assess agreement and consistency in expert evaluations, a multi-method statistical approach was applied, following best practices in scale-based peer assessment (O'Neill, 2017). This included:

- Krippendorff's Alpha, used to measure overall agreement between raters while accounting for chance and disagreement:

$$\alpha = 1 - \frac{D_o}{D_e}$$

where D_o represents observed disagreement and D_e represents expected disagreement.

- Kendall's Coefficient of Concordance (W), used to assess the degree of consensus among raters:

$$W = \frac{12S}{m^2(n^3 - n)}$$

where S is the sum of squared deviations, n is the number of items, and m is the number of raters.

- Cronbach's Alpha, used to evaluate internal consistency of the ratings:

$$\alpha = \frac{N * \bar{c}}{\bar{v} + (N-1) * \bar{c}}$$

where N is the number of items, $\bar{c}$ is the mean covariance between items and $\bar{v}$ is the mean item variance.

These metrics are appropriate for Likert-scale data and help identify potential issues such as rater bias, inconsistent scale usage, and subjective variability in human judgement.

In addition, Intraclass Correlation Coefficient (ICC) was calculated using a two-way random effects model (ICC(2,k)) to assess the absolute agreement scores across multiple raters, treating Likert data as interval-level:

$$ICC = \frac{\sigma^2_{between}}{\sigma^2_{between} + \sigma^2_{within}}$$

This provides a complementary measure of agreement by quantifying the consistency of scoring patterns across raters.

Together, these metrics provide a robust assessment of both ordinal agreement and interval-level consistency in expert evaluations.

3.5. Time-series analysis

To explore how incident patterns evolve over time, a time-series analysis was conducted using both monthly (seasonal) and yearly (long-term) aggregations. The objective was to identify temporal trends in incident occurrence across the derived topic categories.

All analyses were implemented in Python using pandas for data processing and matplotlib for visualisation. Incident dates were parsed into a unified datetime field and converted into monthly and yearly variables. Records with invalid timestamps were removed, and the incident types mapped were taken from the clustered topics with outliers removed.

For yearly analysis, incidents were aggregated into a topic-by-year matrix, and years with partial data were excluded to avoid artificial suppression of frequencies. The matrix was then row-normalised:

$$p_{t,y} = \frac{n_{t,y}}{\sum_y n_{t,y}}$$

where $n_{t,y}$ is the number of incident topics t in year y and $p_{t,y}$ is the proportion of topic t's incidents occurring in year y. This transformation ensures that values representing the proportion of each topic's occurrences attributable to each year, allowing for comparison across topics with differing overall frequencies. Seasonal patterns of monthly analysis were also analysed using the same aggregation approach, replacing years with calendar months.

Heatmaps were produced to visualise both yearly and monthly distributions. To enhance visual contrast between low and moderate priorities, a power-law (gamma) scaling transformation was applied to normalise the values. This transformation improves visibility of lower-frequency patterns without distorting relative trends.

3.6. Association rule mining (ARM)

Then, to examine how reported incidents related to preventative measures, the categorised incident types and preventions were linked together using association rule mining. Each incident report was converted into a transactional record containing its incident-type labels and associated prevention labels, after first removing very rare or poorly defined categories (fewer than 10 occurrences, or labels such as 'unknown' and 'outliers') to improve stability and readability of the patterns. Association rules were mined using the Apriori algorithm (using the mlxtend library) with a minimum support threshold of 0.0025, allowing infrequent but operationally important incident prevention combinations to be retained.

Only rules with a confidence level above 0.07 and lift above 1.01 were retained, ensuring statistical strength and interpretability. These thresholds are lower than those typically applied in association rule mining. However, they were considered appropriate for the present study due to the exploratory objective and the characteristics of the dataset. The BERTopic stage produced a relatively large number of fine-grained incident and prevention categories, resulting in a sparse incident-prevention matrix in which many meaningful operational relationships occurred infrequently. Consequently, applying more restrictive thresholds would have limited the analysis to only the most common incident-prevention combinations and excluded lower-frequency associations associated with specialised operational contexts.

In safety applications, infrequent events can still offer valuable learning opportunities, particularly when they involve uncommon but potentially consequential scenarios. The selected thresholds, therefore, reflect a balance between retaining sufficient coverage of diverse

operational experiences and reducing the inclusion of spurious relationships. While alternative threshold configurations may produce different rule sets, the intention of the present analysis was exploratory knowledge discovery rather than predictive optimisation. Future studies could extend this work by conducting formal sensitivity analyses to assess the stability of the identified associations under different support, confidence, and lift criteria.

Each rule was also constrained to infer prevention strategies from incident topics, reflecting the operational objective of understanding how specific incident patterns trigger preventative behaviours. The retained rules were then visualised as a directed bipartite network, with incident types and prevention categories as separate node sets and edge colour and width scaled by lift and confidence, respectively, to highlight the strongest incident-prevention link.

4. Results

4.1. Near miss incident summary classification and topic naming

Application of BERTopic to the near-miss incident summaries generated 76 distinct topics. One large outlier cluster contained 1656 reports that were too generic or ambiguous for meaningful interpretation and was excluded from further analysis. The remaining 75 topics spanned a diverse set of operational, environmental, safety, and human-factors issues, shown in Table 1. Topic sizes ranged from 10 to 161 reports, with a median of 28. The largest topic accounted for approximately 6% of all classified reports, indicating that topic structure was not dominated by a small number of highly frequent incident types and instead captured a diverse range of near-miss themes across the dataset.

Comparison with the company's existing classification system showed that 66% of the BERTopic-derived clusters could be mapped directly onto predefined categories. The remaining 34% represented either (a) edge cases not explicitly covered by the taxonomy, (b) overlapping themes that cut across categories, or (c) novel patterns not previously recognised. This demonstrated that BERTopic can replicate established taxonomies while also surfacing under-specified or emergent issues, thereby extending the analytical value of incident reports beyond manual classification.

4.2. Time-series analysis

To explore how incident frequencies varied over time, a five-year dataset (2019-2023, plus early 2024) was analysed using monthly and annual decompositions (Figs. 2 and 3). The results reveal clear seasonal dynamics as well as longer-term shifts associated with operational stressors such as the COVID-19 pandemic.

Across months, incident patterns show clear seasonality. Passenger and traffic-related categories exhibit sustained peaks during the summer months (May-July), aligning with increased travel demand and peak tourist and commuting periods. The Traffic and Vehicle operations and Passenger and Public safety groups consistently intensify from late spring into summer, reflecting congestion pressures, higher passenger volumes, and increased pedestrian-vehicle interactions at terminals, car parks, and loading ramps. Similarly, cargo and logistics issues also show mid-year increases, potentially linked to heightened freight and baggage-handling activity during high-demand seasons.

Conversely, winter months demonstrate elevated rates of Mechanical and power system incidents, environmental and infrastructure incidents, and workplace and crew practices incidents. This pattern reflects environmental and operational conditions such as adverse weather, reduced daylight, tidal constraints, and maintenance-season activity. These groups show January and late-year concentrations consistent with maintenance cycles, asset wear, and environmental stress on infrastructure and equipment. Navigation-related events display more distributed, year-round patterns, indicating persistent operational exposure rather than seasonal concentration.

Table 1
Clustered topic groups of incident summary topic modelling.

Cluster	Topic numbers	Topic Names	N (% of Reports)
Passenger & Public Safety	2, 5, 9, 10, 11, 14, 15, 22, 35, 45, 55, 62, 66, 74	Walkway and Marshalling Lane Barriers, Slippery Surface and Growth, Maritime Equipment and Safety, Pier Safety and Public Access, Door and Gate Operation, Cyclist Safety and Bike Rental, Child Safety Around Railings, Maritime Accidents and Damages, Pet-related Incidents, Docking and Disembarkation Incidents, Passenger Vessel Toilet Issues, Seat and Chair Construction and Repair, Customer Service Desk Screen Incident, Police on Male Passenger Behaviour	642 (23.74%)
Mooring & Berthing Operations	0, 1, 16, 19, 34, 40, 41, 58, 61, 69	Maritime Mooring and Rope Maintenance, Waste Oil Disposal and Management, Maritime Navigation and Vessel Movements, Ferry Ramp Operations, Harbour Berthing Challenges and Vessel Safety, Maritime Engine Troubleshooting, Vehicle Discharge and Traffic Issues, Maritime Navigation Incidents, Marine Winch and Mooring Operations, Maritime Equipment Failures and Troubleshooting	508 (18.79%)
Traffic & Vehicle Operations	6, 7, 24, 26, 27, 29, 32, 33, 42, 44, 46, 65, 71, 73, 75	Road and Vehicle Incidents, Pedestrian and Vehicle Interaction, Car Handbrake Issues in Vehicles, HGV Trailer Traffic and Marshalling, Traffic Congestion at Roundabouts, Ferry Terminal Traffic Management, Public Transportation Queues and Disruptions, Vehicle Deck and Mezzanine Loading, Vehicle Parking and Traffic Issues, Coach Travel and Traffic Regulations, Vehicle Transportation and Safety, Passenger Vessel Incidents at Gourock, Workplace Car Park Incidents, Traffic Congestion at Kilchoan Pier, Speed Limit Violation at Pier	444 (16.42%)
Navigation & Bridge Management	3, 13, 21, 23, 31, 37, 38, 43, 54, 59	Marine Vessel Operations and Emergency Drills, Ship Thruster Failures, Mallaig Harbour Vessel Navigation, Marine Vessel and Yachting Operations, Yachting Incident at Rothesay Harbour, Boat/Ship Navigation Control, Marine Vessel Maintenance and Navigation, Autopilot System Issues, Schottel Steering System Issues and Solutions, Yacht Navigation and Safety Protocols	376 (13.91%)
Mechanical & Power Systems	4, 12, 20, 48, 53, 56, 64, 68	Generator and Power System Management, Engine	266 (9.84%)

(continued on next page)

Table 1 (*continued*)

Cluster	Topic numbers	Topic Names	N (% of Reports)
		Maintenance and Troubleshooting, Electrical Overheating and Safety Issues, Fire Safety Equipment Maintenance, Welding and Fire Safety, Fire Door and Fuse Safety, Fuel and Water Filter Maintenance, Fuel Pipe Leak and Repair in Ship Engines	
Fire & Emergency Systems	17, 18, 39, 47, 57, 67	Lifejacket Use and Safety, Smoking and Fire incidents, Passenger and Vehicle Disembarking Issues, Covid-19 and Ambulance Operations, Seaside Activities and Locations, Vehicle Activity at Lock Slipway	156 (5.77%)
Workplace & Crew Practices	25, 36, 49, 50, 63, 70, 72	Construction and Safety at Seaports, Night Watchman Duties and Challenges, Safety Equipment Regulations at Work Sites, Laundry and Dryer Equipment, Maritime Supervision and Vessel Management, UK Marine Infrastructure Maintenance, Intermittent Fault Investigation	136 (5.03%)
Cargo & Logistics	8, 30	Gas Shipment and Transport, Loading and Safety Procedures in Transport	92 (3.40%)
Environmental & Infrastructure	28, 51, 52, 60	Harbour and Maritime Activities, Fishing Activities at Loch Shira in Largs, Pier Lighting Issues, Seaman Booking and Organisation Issues	84 (3.11%)

Annual trends further emphasise the impact of broader systemic factors. The pandemic period (2020-2021) was marked by relative reductions in passenger and traffic-related categories, aligning with lower passenger and transport activity. These categories then demonstrate strong rebounds in 2022, consistent with post-restriction recovery and renewed pressures. By contrast, mechanical and power systems and workplace and crew practices show relative increases in 2023, potentially reflecting deferred maintenance, heightened inspection regimes, or a return to normal operational tempo, revealing latent system issues.

4.3. Classification preventions

Topic modelling of the free-text prevention measures yields 87 distinct topics after excluding 1287 outlier entries that were too generic or ambiguous for meaningful interpretation. These were grouped into eight higher-level clusters (Table 2), each capturing a coherent set of prevention practices. The distribution of reports across clusters was relatively balanced, with no single cluster dominating, reflecting the breadth of operational risk management strategies in place.

The preventative clusters (Table 2) show that safety measures were distributed across technical, operational, and human domains. The largest share related to vehicle and road safety, particularly traffic marshalling and pedestrian-vehicle interactions in terminal and ferry deck settings. Crew and team management was also highly represented, underscoring the centrality of vigilance, communication, and leadership in incident prevention. Other major clusters included passenger and public safety, maritime operations and vessel handing, and maintenance and technical safety, reflecting the breadth of frontline operational practices. Smaller but significant clusters captured port and shore operations, emergency preparedness, and risk/luck/uncertainty. Together, these categories illustrate a comprehensive landscape of preventative strategies, ranging from routine vigilance and technical reliability to readiness for rare but high-consequence events. The prominence of less systematic measures in the risk/uncertainty cluster further highlights the need for greater reporting consistency and formalisation of ad-hoc practices.

4.4. Linking incident types to prevention actions

The network diagram in Fig. 4 illustrated the relationships between ferry incident types (blue nodes) and their associated prevention measures (green nodes) identified through association-rule mining. The edge width represents the confidence of each rule, thicker lines denote stronger and more consistent relationships, while edge colour intensity indicates lift, showing the relative strength of the association beyond random chance.

Several clear clusters and associations were found linking technical incidents with asset-management controls. One notable association was the relationship between generator and power system management and generator capacity and functionality. This link reflects the strong dependence of power-related incidents on the adequacy and reliability

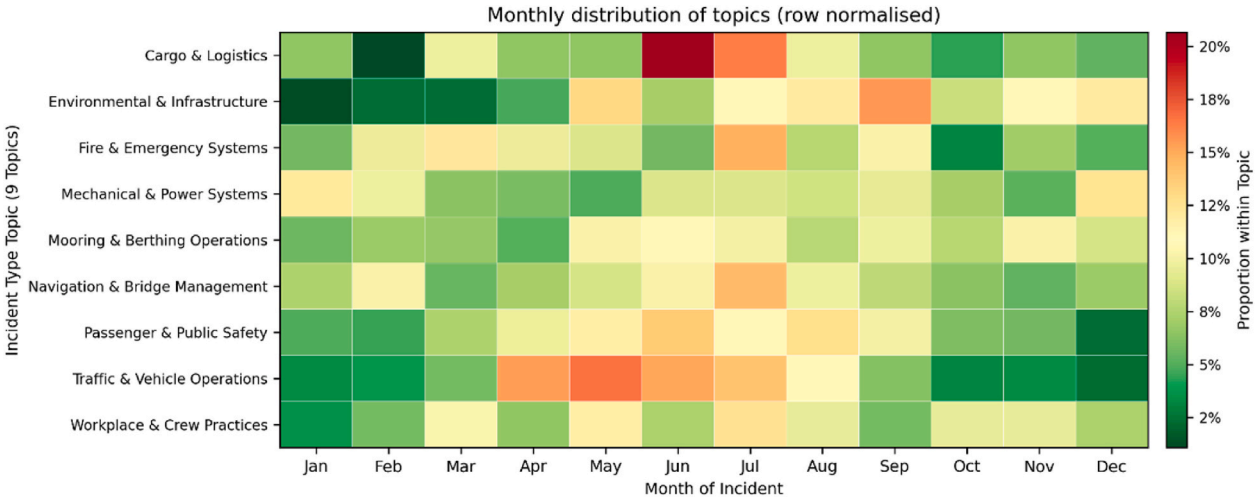

Fig. 2. Seasonal patterns of incidents by incident topic and month.

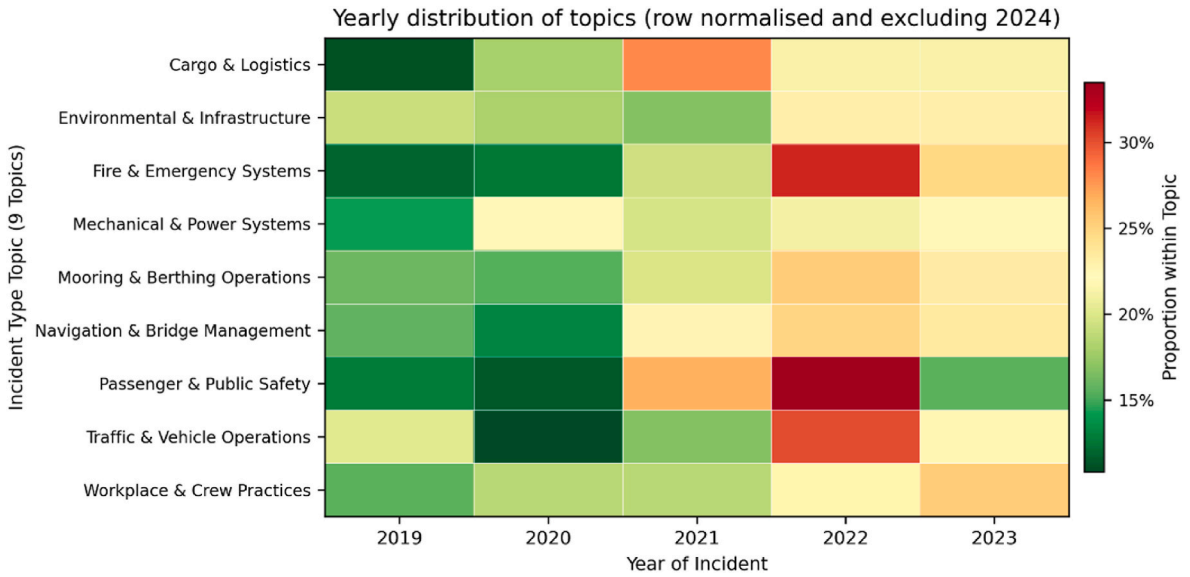

Fig. 3. Heatmap of yearly incidents by topic from 2019 to 2023.

of onboard generator systems. Examination of underlying reports revealed that these near misses were associated with a diverse range of technical failures, including generator trips, failed starts, cooling system deficiencies, fuel supply restrictions, electrical protection faults, and load management issues. Despite the diversity of failure mechanisms, the corresponding restoration actions consistently focused on maintaining adequate generating capacity, preventing redundancy through standby generators, and ensuring reserve power availability. This finding suggests that the primary safety mechanism was not the prevention of individual component failures, but maintenance of power system resilience. In many cases, vessel operations continued safely because reserve generating capacity was available to absorb load fluctuations or compensate for equipment failures before they affected operational performance.

Associations related to passenger and vehicle movements highlighted the importance of managing shared operational spaces. The relationship between passenger-vehicle interaction incidents and pedestrian and vehicle crossing controls revealed that these near misses frequently occurred during embarkation and disembarkation, when pedestrians crossed active vehicle lanes, moved between marshalling and terminal areas, or entered linkspan zones while vehicles remained in motion. The corresponding prevention measures consistently involved marshals stopping or slowing traffic, controlling vehicle movements, and facilitating designated pedestrian crossings. These findings suggest that the principal risk arose from the coexistence of pedestrian and vehicle flows rather than isolated unsafe acts. Consequently, traffic segregation, designated crossing points, and active marshalling emerged as the primary mechanisms through which passenger-vehicle interaction incidents were prevented. Related clusters involving walkway and marshalling lane barriers, customer safety, moving vehicles and traffic speed limitations reinforced this pattern. A further association linked marine vessel operations and emergency drills prevented by FRC (Fast Rescue Craft) Recovery and Launching Procedures. Many of these near misses occurred during training exercises, drills, inspections, or routine testing of FRC systems. The incidents exposed a range of conditions, including damaged wires, defective hooks, hydraulic system faults, limit switch failures, equipment wear, and operator handling errors. Prevention actions consistently involved controlled recovery procedures, equipment inspections, functional testing, servicing activities, and operational restrictions when defects were identified. These findings indicate that emergency drills served a dual safety function. Beyond preparing crews for emergency response, they also provided opportunities to identify equipment defects and procedural weaknesses before they contributed to operational failures.

Similarly, door and gate operations were strongly associated with door and gate security. The incidents frequently involved access-control failures, defective lock mechanisms, jammed doors, and situations where access to operational spaces was either unintentionally restructured or insufficiently secured. Prevention actions consistently involved prompt reporting by crew members, corrective maintenance, lock replacement and active intervention to restore safe access. Effective control, therefore, depended on both the integrity of locking systems and the vigilance of personnel who identified and responded to defects before they escalated.

Fire safety and environmental management linkages also emerge clearly. The relationship between smoking and fire incidents and fire prevention and detection demonstrated that many of the incidents originated from relatively minor ignition sources, including discarded smoking materials, hot-work activities, smouldering waste, overheating equipment, and vehicle-related defects. In most cases, the incidents were identified during the early stages of fire development when only smoke, unusual odours, or localised heating were present. The corresponding prevention actions consistently involved early detection, prompt reporting, rapid investigation, and immediate intervention by passengers, crew members, or other personnel. Several incidents were successfully interrupted because individuals recognised weak warning signs, such as smoke burning smells, or abnormal equipment behaviour, before the situation escalated into a more significant fire event. These findings suggest that effective fire prevention relied primarily on early detection, reporting, and intervention rather than the activation of firefighting systems after escalation has occurred. Likewise, the association between waste oil disposal and management and oil spill management and prevention reflects the environmental dimension of ferry safety. operations, illustrating how improper waste handling directly contributes to spill events.

Infrastructure-related controls were evident through the association between ferry ramp operations and marine ramp operations and maintenance. Many near misses occurred during ramp deployment, recovery, or alignment with shore infrastructure. The incidents frequently involved equipment degradation, ramp-linkspan interface issues, damaged components, and operational anomalies identified during routine use. Prevention actions consistently involved inspections, early defect identification, immediate maintenance interventions, and suspension of operations when abnormal conditions were observed. These

Table 2
Clustered topic groups of near miss prevention.

Cluster	Topic numbers	Topic Names	N (% of Report)
Vehicle & Road Safety	6, 8, 10, 15, 17, 21, 24, 36, 38, 48, 49, 56, 60, 61, 62, 83, 86	Staff directing traffic and park safety, Car marshalling and vehicle control, Traffic speed limitations, Stationary vehicles, Vehicle patrol and vigilance, Driver awareness and attention, Road traffic incidents, Driver safety and reaction time, Vehicle-pedestrian falls, Vehicle horn usage, Pedestrian and vehicle crossings, Vigilant vehicle drivers, Vehicle loading and unloading procedures, Customer safety and moving vehicles, Quiet morning road traffic, Traffic management in carparks, Traffic and pedestrian safety	565 (18.38%)
Crew & Team Management	0, 5, 12, 27, 31, 41, 66, 71, 72, 78, 82	Bridge Team Management and Communication, Maritime watchkeeping and machinery observations, Crew observation and awareness, Crew intervention and guidance, Crew vigilance at work, Staff reactions and responses, CFL staff communication and safety procedures, Master's decision support and development, Crew response and direction, Staff vigilance and resilience, Crew management and problem resolution	550 (17.91%)
Maritime Operations & Vessel Handling	4, 23, 26, 29, 33, 34, 53, 57, 58, 73, 77, 85, 87	Vessel service safety, Boat manoeuvring and thrust control, Boating and rope handling, Skipper experience and vessel operation issues, Maritime mooring techniques, Marine steering and control systems, Ship handling and seamanship knowledge, Ferry master's situational awareness, Maritime safety and vigilance, Tug operation and safety, Maritime berthing procedures, FRC recovery and launching procedures	384 (12.49%)
Passenger, Crew & Public Safety	9, 13, 16, 18, 25, 52, 59, 63, 64, 65, 81	Crew members' quick actions and reactions, Security and cleanliness in public places, Passenger and crew safety alerts, Describing favourable weather conditions, Work interruption and escalation, Desolate surroundings, Passenger and crew safety situations, Marine safety procedures, Marine staff operations, Marine alarm system and monitoring, Clothing storage and laundry management	368 (11.97%)
Maintenance & Technical Safety	3, 19, 30, 32, 37, 39, 47, 51, 54, 70, 74	Routine maintenance and inspections, Manual chain positioning on platforms, Deck equipment safety, Oil spill management and prevention, Generator Capacity and Functionality, Preventing injuries by maintenance balance, Preventing pedestrian falls, Door and gate security, Electrical power supply issues, Electrical circuit safety and maintenance, Valve operation and maintenance	354 (11.53%)
Risk, Luck & Uncertainty	2, 14, 20, 22, 67, 68, 69, 75, 79, 84	Luck and Chance, Good fortune and timing, Scottish isles and lochs, Undefined topic, Accidental falls and reactions, Managing customer balance and falls, Uncertainty and negation, Corrective actions and appreciation events, Discussions on luck, Knowledge and uncertainty	324 (10.55%)
Port, Pier & Shore Operations	1, 35, 40, 44, 45, 46, 50, 76, 80	Pier Workers, Port Assistant Observations and Actions, Port management and interventions, Port staff process management and intervention, Officers' intervention and observations, Port staff training and vigilance, Marine ramp operations and maintenance, Quiet pier and reduced sailings, Ferry traffic and disembarkation	320 (10.40%)
Emergency Preparedness & Response	7, 11, 28, 42, 43, 55	Fire prevention and detection, Emergency response and awareness, Potential accident risks, Emergency drill discoveries, Fast problem resolution & accident response, Emergency procedures and equipment operations	221 (7.18%)

findings suggest that ramp safety is influenced not only by the mechanical condition of the equipment but also by effective management of the vessel-shore interface, with incidents often prevented through the timely recognition of developing defects before operational failure occurred.

Navigational and bridge management relations form one of the largest and most coherent sub-clusters within the network. The connection between yacht navigation and safety protocols and ferry masters' situational awareness indicated that near-miss and collision risks involving smaller crafts were strongly influenced by the master's ability to maintain dynamic situational awareness and adapt to unpredictable movements of non-commercial vessels. Examination of the reports revealed that many of these events involved recreational vessels making unexpected manoeuvres, crossing the bows of large vessels, impeding traffic in constrained waterways, or failing to follow established navigation rules. In contrast, the corresponding prevention narratives consistently attributed successful avoidance of escalation to the situational awareness of the ferry master. The ability of bridge teams to continuously monitor vessel movement, anticipate developing conflicts, and intervene early therefore emerged as the primary mechanism

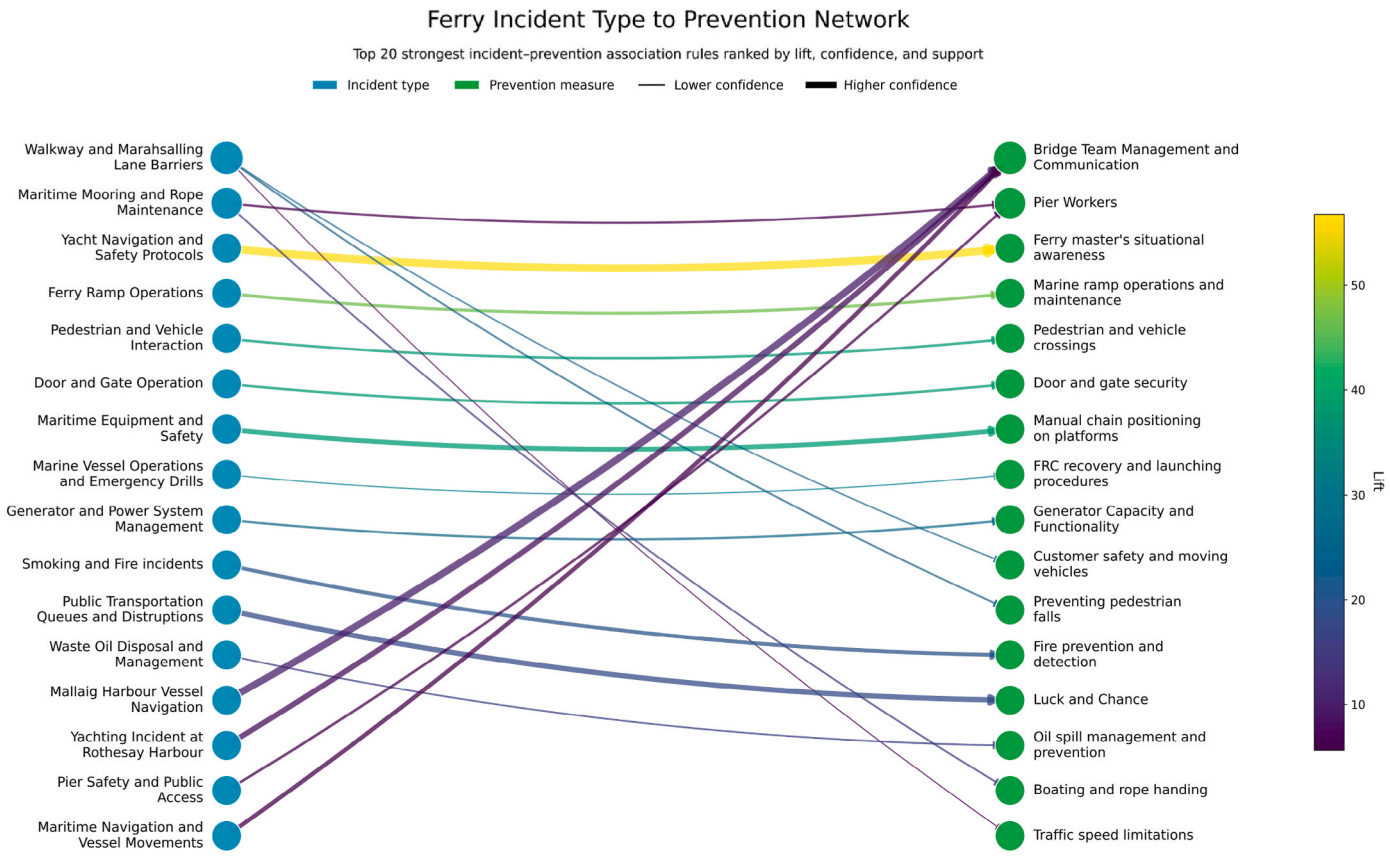

Fig. 4. Network of incident type and prevention association rule mining results.

preventing close-quarters situations involving recreational and non-commercial vessels.

A similar pattern was evident in the associations linking yachting incidents at Rothesay Harbour, Mallaig Harbour vessel navigation, and maritime navigation and vessel movements, with bridge team management and communication. The Rothesay incidents frequently occurred within congested harbour environments involving multiple interacting waterway users, including yachts, kayaks, recreational craft, harbour authorities, and commercial vessels. These situations involved unexpected manoeuvres, route conflicts, and interactions requiring rapid adaptation by ferry crews. The corresponding prevention actions consistently highlighted effective bridge-team communication, adherence to established navigational procedures, coordination with harbour authorities, and professional watchkeeping practices. Successful outcomes were associated with information sharing, collective decision-making, and proactive traffic management by bridge teams.

Other maritime operations also show similar relational clusters. For example, the linkage between maritime mooring and rope maintenance with both boating and rope handling and pier workers. Reports revealed that many near misses involved rope entanglement, snagging, unexpected tensioning, rope degradation, and interactions between mooring systems and harbour infrastructure. Several incidents involved conditions that could have resulted in snap-back events, rope failure, or personnel exposure to stored-energy hazards. The corresponding prevention actions consistently highlighted crew vigilance, continuous monitoring of rope behaviour, routine inspections, and the timely identification of developing hazards before operations continued. In many cases, incidents were prevented because abnormal rope movement, deterioration or tension changes were recognised early and corrective actions were taken before failure occurred. These findings suggest that safe mooring operations depend not only on the physical condition of ropes and equipment but also on the ability of personnel to recognise weak signals of deterioration and intervene before escalation. With the association to pier workers, prevention actions highlighted communication between vessel and shore personnel, rapid intervention, and the identification of unsafe conditions before operations continued. In several cases, incidents were prevented because either vessel crews or pier staff recognised developing hazards and communicated this information before personnel entered danger areas or line tensions increased further.

Taken together, these findings indicate that mooring safety depended on both equipment integrity and coordinated vessel-shore operations. Prevention was achieved not solely through procedural compliance, but through shared situational awareness, timely communication, and the collective ability of shipboard and shoreside personnel to recognise and respond to evolving hazards before they developed into injury events or equipment failures.

Finally, the association between public transportation queues and disruptions and luck and chance revealed a markedly different pattern from the other relationships identified within the network. Examination of the reports highlighted a recurring problem whereby passengers were required to queue on public roads, often without designated waiting areas, physical barrier, or effective separation from vehicle traffic. Several narratives explicitly noted that these situations occurred repeatedly during busy periods and had been raised on multiple occasions over several years. Unlike most other associations identified in this study, the corresponding prevention narratives did not describe active interventions, procedural safeguards, or established operational controls. Instead, incidents were frequently avoided simply because no vehicle happened to enter the occupied space at the time or because potentially hazardous circumstances did not materialise. In these cases, the successful prevention of harm appeared to depend largely on favourable circumstances rather than identifiable protective mechanisms. This finding suggests the absence of effective risk controls rather

than the presence of successful preventative measures. It further demonstrates how near-miss analysis can reveal persistent vulnerabilities that may remain hidden within traditional incident classification systems. Whereas many of the other associations identified adaptive responses that interrupted escalation pathways, this relationship exposed situations in which organisational resilience appeared weak, and safety performance relied heavily on change rather than robust defences.

These patterns provide evidence that certain types of incidents predictably trigger specific classes of prevention action. By quantifying the strength of these associations, the analysis highlights opportunities for evidence-based safety planning, such as prioritising bridge communication training for navigation hotspots.

4.5. Validation

4.5.1. Incident summary validation

To assess the interpretability of BERTopic clusters, four domain experts independently evaluated topic quality across three dimensions: keyword coherence, GPT-generated label fit, and document relevance. Agreement was assessed using multiple reliability metrics.

Krippendorff's α and Kendall's W indicated relatively low ordinal agreement (α = 0.24-0.39, W = 0.09-0.12), suggesting that experts did not consistently converge on identical interpretations of topic quality. Given that near-miss narratives often contain overlapping operational, technical, and contextual elements, some variation in semantic interpretation was expected. This observation also reflects a broader challenge in incident analysis, whereby even experienced practitioners may differ in how they categorise and interpret complex events.

Cronbach's α scores were consistently above 0.72 (range: 0.72-0.83), showing that individual raters applied their own standards consistently rather than randomly. Intraclass correlation coefficients (ICC) provided further evidence of reliability: keyword coherence achieved good agreement (ICC = 0.79, $p < 0.001$), while GPT label fit (ICC = 0.69) and document relevance (ICC = 0.70) demonstrated moderate agreement (Table 3). These results suggest that the identified topics were generally interpretable, although the precise boundaries and labels assigned to some topics remained subject to expert judgement.

Mean ratings across experts were positive for all three dimensions (Table 4). GPT-generated labels and document relevance were scored slightly higher than keyword coherence, suggesting that while topics were generally interpretable, the specific keyword groupings were less intuitive.

Qualitative reviewer comments provided further nuance to these findings. Some topics were judged to have overly narrow or overly specific labels (e.g. "waste oil disposal" applied to broader engine-related issues; "emergency drills" covering routine operational practices). Such observations reinforce the subjective nature of assigning semantic labels to emerging themes and illustrate the challenges inherent in achieving complete consensus when interpreting operational narratives. Despite these inconsistencies, most clusters were considered interpretable, relevant, and aligned with their underlying reports.

Taken together, the validation demonstrated that the majority of BERTopic-derived incident clusters were perceived as both meaningful and useable for safety analysis, although experts differed in their interpretation of some themes. The combination of low agreement coefficients and moderate reliability measures reflects the inherently subjective nature of interpreting complex near-miss narratives. These findings suggest that the extracted topics provide a plausible and operationally useful representation of the underlying reports, while acknowledging that alternative interpretations may also exist.

Table 4
Mean expert scores across topic validation dimensions.

Expert	Mean Keyword Coherence	Mean GPT Label fit	Mean label-document fit
Expert 1	4.029	3.943	3.743
Expert 2	3.632	3.868	4.022
Expert 3	3.575	3.725	3.875
Expert 4	3.707	3.697	3.973
Average	3.73	3.81	3.90

4.5.2. Prevention validation

As with incident classifications, prevention clusters were evaluated by four domain experts. Agreement was relatively low when assessed with Krippendorff's α (0.13-0.21), reflecting variability in how experts interpreted qualitative prevention themes. Given that many prevention narratives involved broad behavioural adaptations, overlapping operational practices, and context-dependent actions, some divergence in interpretation was anticipated.

Nevertheless, Intraclass correlation coefficients (ICC) indicated moderate agreement from keyword coherence (ICC = 0.62, $p < 0.001$), GPT label fit (ICC = 0.61, $p < 0.001$) and document relevance (ICC = 0.47, $p = 0.005$) (Table 5).

Internal consistency was high for three of the four raters (Cronbach's α = 0.80, 0.90, 0.92), with one expert showing lower reliability (α = 0.37). This suggests that while individual interpretations varied, most experts applied scoring criteria consistently. These findings suggest that, although experts did not always agree on the exact interpretation of individual prevention topics, most applied their evaluation criteria consistently. The resulting prevention clusters should therefore be viewed as exploratory yet operationally meaningful representations of the data rather than definitive classifications.

Mean ratings across dimensions (Table 6) were generally positive, with GPT-generated labels receiving slightly higher scores than keyword coherence or document relevance.

Taken together, these results mirror those for incident topics: although experts varied in their detailed interpretations, overall ratings were favourable, and agreement was sufficient to confirm that the prevention clusters were interpretable and practically meaningful.

5. Discussion

This study applied BERTopic, supported by GPT-generated labels, to a large corpus of near-miss reports from ferry operations. The results demonstrate that topic modelling can extract coherent and interpretable trends from unstructured safety data, replicate known taxonomies, and reveal emergent themes not captured in existing reporting frameworks. Validation with domain experts confirmed the practical interpretability of most clusters, while temporal and association analyses provided new insights.

Existing maritime safety studies have predominantly relied on accident investigations and predefined taxonomies to identify risk factors. In contrast, this study demonstrates how near-miss narratives can be systematically analysed to identify both emerging hazards and the operational mechanisms that prevent escalation. By linking incidents directly to the preventative actions documented by frontline personnel, the

Table 3
ICC scores per evaluation dimension.

Dimension	ICC(2,k)	Interpretation	p-value
Keyword Coherence	0.791	Good Agreement	<0.001
GPT Name Fit	0.689	Moderate Agreement	<0.001
Document Relevance	0.700	Moderate Agreement	<0.001

Table 5
ICC scores per evaluation dimension.

Dimension	ICC(2,k)	Interpretation	p-value
Keyword Coherence	0.620	Moderate agreement	<0.001
GPT Name Fit	0.612	Moderate Agreement	<0.001
Document Relevance	0.474	Moderate Agreement	0.005

Table 6
Mean expert scores across topic validation dimensions.

Expert	Mean Keyword Coherence	Mean GPT Label fit	Mean label-document fit
Expert 1	4.366	4.268	4.146
Expert 2	3.357	3.571	4.048
Expert 3	4.05	4.025	3.45
Expert 4	3.195	3.268	3.268
Average	3.742	3.783	3.728

findings provide insights into how safety is actively maintained within ferry operations.

5.1. Key findings and interpretation

The topic modelling results demonstrate that near-miss narratives contain a level of operational detail and contextual specificity that is not fully captured by the organisations existing incident taxonomy. While the existing classification system records incidents within broad categories such as traffic management, mooring operations, navigation incidents, and vehicle-related, the topics identified in this study revealed more granular subthemes relating to specific operational activities, locations, equipment, and preventive actions. This divergence suggests that the current classification system under-represents complex and cross-cutting interactions between people, vessels, equipment, and port environments. The results extracted several emergency topics combining human, technical and environmental elements, reflecting how incidents develop in practice. On the prevention side, staff routinely documented positive adaptations, crew vigilance, emergency preparedness, pier team interventions, and seamanship actions, yet these measures are seldom analysed in conventional Safety-I reporting. Together, the findings support a Safety-II perspective in which learning is not derived only from failures, but from understanding the everyday mechanisms that keep the system functioning safely. One of the most common incident clusters was passenger and public safety, accounting for near-misses involving walkways, boarding/disembarking, and pedestrian-vehicle proximity. Similarly, traffic and vehicle operations were found to be the third most common incident type cluster, including HGV marshalling, roundabout congestion issues, and deck loading. Seasonal analysis showed clear summer peaks in these categories. The high volume of tourists and seasonal passengers, many of whom are unfamiliar with port layouts, onboarding procedures, and driving within ferry-terminal environments, may contribute to the increase in these incidents. Kim et al. (Kim et al., 2022) highlighted how ferry-terminal walkability is influenced by the design and organisation of passenger movement routes, particularly during embarkation and disembarkation. The present findings complement this perspective by demonstrating that these environments are also associated with increased operational safety challenges, particularly during periods of elevated passenger demand. While Kim et al. focused on passenger perceptions and satisfaction, the near-miss data analysed here provide evidence of how pedestrian-vehicle conflicts and movement-related hazards emerge in practice and require active intervention from ferry personnel.

Mooring and berthing operations were also found to be a recurring incident type, including berthing, ramp operations, and line handling. Navigation and bridge management formed the fourth-largest cluster, covering harbour-approach navigation, bridge resource management, and control-system issues such as thruster failures, autopilot anomalies and Schottel steering faults. Findings from ARM show that these types of harbour-specific topics were strongly associated with pier-team actions and bridge communication. This supports previous maritime human-factors research highlighting communication, coordination, and shared situational awareness as a critical safeguard during port approaches and other high-workload operational phases (Hetherington et al., 2006). While previous studies have largely identified communication breakdowns and bridge-team deficiencies as contributors to accidents, the present findings extend this evidence by demonstrating that effective coordination between vessel and shore personnel contributes to the successful management of navigational and berthing hazards before they escalate.

In addition to these themes, we also see clear patterns related to mechanical and power-system reliability, which show noticeably higher incident and near-miss rates during the winter months. This aligns with the operational reality that winter is the primary period for major vessel overhauls, maintenance windows, and engineering interventions. While these activities are essential to ensure the fleet is fully available during the peak summer season, they can temporarily increase the likelihood of mechanical or power-system issues as vessels return to service after refit periods or operate on tighter engineering schedules. These findings suggest that increases in near-miss reporting do not necessarily indicate deteriorating safety performance. Instead, they may reflect periods of heightened operational sensitivity associated with maintenance transitions and engineering interventions. Ferry operators should therefore interpret near-miss trends alongside maintenance schedules and operational context when prioritising safety interventions and allocating engineering resources.

Annual patterns revealed a marked decline in near-misses during the COVID-19 period. The reduction in near misses during the pandemic reflects temporary simplification of operations, reduced system load, and lower coordination demands. This was followed by a rebound in 2022 which illustrates how risk re-emerges and complexity returns with the quick recovery of demand. Although previous studies have primarily examined accident trends and safety challenges in ferry systems, rather than near-miss reporting specifically, they similarly emphasise that changes in operational conditions, traffic levels, and system demands can alter the nature and frequency of safety events (Golden and Weisbrod, 2016; Hebbar et al., 2023). Consequently, safety-management systems should pay particular attention during periods of organisational or operational change, including demand recovery, route expansion, and service restructuring, as these phases may introduce new interactions, coordination requirements, and emerging risk pathways.

The prevention topics reveal a distributed set of effective control functions: traffic marshalling and pedestrian management, crew/bridge teamwork and vigilance, seamanship-based stabilisation, and routine technical reliability work, with additional contributions from port/shore operations and emergency preparedness. Collectively, these prevention themes provide empirical evidence of the adaptive capacities that enable ferry operations to remain resilient under varying conditions. Importantly, the findings demonstrate that such adaptations can be systematically identified from large-scale narrative datasets rather than solely through qualitative case studies.

The ARM results reinforce this perspective by connecting specific incident types directly to corresponding prevention actions. For example, the linkage between generator and power system management and generator capacity and functionality highlights the importance of proactive reliability assurance and load verification in maintaining power integrity. Other relationships emphasise the value of operational resilience achieved through situational awareness, procedural competence, and effective team communication and coordination. This finding complements previous accident-based research by Sandhåland et al. (Sandhå et al., 2015) who found that deficiencies in situational awareness frequently preceded collisions and navigational incidents. In contrast, the present study reveals the opposite perspective, demonstrating how effective bridge teamwork, communication, and situational awareness repeatedly prevented conflicts involving yachts, fishing vessels, harbour traffic, and technical anomalies from escalating into accidents. Whereas accident investigations primarily document the absence of effective controls, the present findings highlight the adaptive actions ferry crews use to manage operational disturbances and maintain safety in everyday operations.

The practical value of the ARM approach lies in its ability to identify where specific preventive mechanisms are most likely to interrupt escalation pathways. For ferry operators, this provides an evidence base for integrating near-miss information into maintenance planning, targeted training initiatives, and resource allocation. For port authorities, the findings highlight opportunities to strengthen passenger segregation measures, terminal layouts, and vessel-shore coordination practices. At the regulatory level, the systematic analysis of near misses provides a complementary source of leading safety indicators that can identify emerging vulnerabilities before they manifest as accidents. Rather than replacing existing safety-management frameworks, these approaches may enhance their predictive capability and support more proactive forms of risk governance. The expert validation process found that model outputs were coherent, applicable and useful for analysis. Average ratings across four domain experts were positive for keyword coherence, label fit, and document relevance (means ~3.7-3.9), and Intraclass correlation coefficients indicated moderate-good agreement. While ordinal agreement (Krippendorff's α, Kendalls' W) was modest, reflecting differences in rankings, internal consistency was high for most raters, and comments indicated only minimal issues (for example, overly narrow labels). Collectively, these findings suggest that the BERTopic-GPT pipeline provides sufficiently reliable outputs for exploratory trend analysis and ARM applications, while recognising that alternative interpretations of individual topics may remain plausible given the inherent subjectivity of semantic evaluation.

5.2. Practical implications

For ferry operators, these insights highlight the need for proactive resourcing that reflects seasonal patterns of demand. This includes scaling up traffic management and passenger safety measures during peak summer months, while using the quieter winter season for pre-emptive infrastructure inspections and maintenance. The associations identified through ARM further suggest that near-miss information can be integrated more directly into operational decision-making. For example, ramp-related near misses may act as early indicators of developing asset degradation at the vessel-shore interface, supporting more targeted inspections and preventative maintenance. Similarly, generator-related incidents emphasise the importance of maintaining reserve capacity and redundancy within critical systems. Training initiatives can also be aligned with recurrent operational risks, for example, improving bridge resource management in congested waterways involving recreational or non-compliant vessels, communication, reinforcing effective mooring and vessel-shore communication practices and improving the management of passenger-vehicle interfaces and fall prevention measures. Emergency drills should also be viewed not solely as compliance activities, but as opportunities to identify latent equipment and procedural weaknesses before failures occur in operational settings.

For port authorities, the findings highlight opportunities to strengthen the management of shared operational spaces. Passenger-vehicle interaction incidents frequently occurred where pedestrians and vehicle movements overlapped and were successfully mitigated through active marshalling, designated crossing, and traffic control measures. In contrast, the association between passenger queues and luck and chance identified situations in which safety appeared to depend on favourable circumstances rather than robust preventative controls. These findings suggest that terminal layouts, designated waiting areas, physical segregation measures, and traffic-flow arrangements should be reviewed to minimise reliance on frontline intervention and improve resilience during periods of peak demand.

At an industry level, our findings show that NLP-based approaches can add significant value to existing reporting frameworks. By surfacing emergent risks, linking incidents to preventative strategies, and enhancing the predictive use of near-miss data, operators can move from reactive safety management to a more proactive, system-wide approach. More specifically, narrative-based analyses of near misses can complement traditional lagging indicators by identifying both emerging vulnerabilities and the adaptive mechanisms through which escalation is prevented. This provides regulators and operators with an evidence base to support targeted interventions, facilitate cross-organisational learning, and strengthen the predictive capabilities of existing safety management systems. Ultimately, this strengthens resilience, reduces operational disruptions, and builds greater passenger confidence in ferry services.

5.3. Limitations

Several limitations should be considered when interpreting the findings of this study. First, the dataset was obtained from a single ferry operator, which may limit the generalisability of the results. Near-miss reporting practices, operational procedures, safety cultures, and risk exposures can vary considerably between operators, vessel types, geographic regions, and regulatory environments. Consequently, the topics and prevention measures identified in this study should not be assumed to represent the wider ferry industry.

Additionally, a substantial proportion of the records were classified as outliers by BERTopic. This outcome reflects the density-based nature of the HDBSCAN clustering algorithm, which intentionally avoids forcing semantically weak or ambiguous records into clusters. While this improved the coherence and interpretability of the resulting topics, it also means that potentially relevant information contained within these records was not incorporated into the final topic structure. The relatively large number of outliers is likely attributable to a combination of factors, including the presence of short and contextually sparse narratives, variability in reporting styles and terminology, and the conservative clustering behaviour of the BERTopic pipeline. This behaviour is recognised in BERTopic applications and represents a trade-off between maximising document coverage and maintaining topic quality. Consequently, the findings presented in this study may be more representative of incident types characterised by richer and more consistently documented narratives, while rare, highly specific, or poorly described events may be underrepresented within the extracted topic structure. This potential bias should be considered when interpreting the prevalence and relative importance of identified topics and associations.

The quality and level of detail contained within the near-miss narratives varied considerably. Many reports consisted of brief descriptions, incomplete contextual information, or highly operator-specific terminology. Such variability can affect the quality of sentence embeddings and reduce the ability of topic-modelling algorithms to identify consistent semantic patterns. Short or ambiguous reports were particularly likely to be assigned to the outlier category, while sparse narratives may have limited the granularity of some identified topics. These challenges are common in operational safety reporting systems and highlight the importance of consistent reporting practices and sufficient narrative detail for effective text analytics.

As with all voluntary reporting systems, the findings are also dependent upon reporting behaviour and organisational reporting culture. Not all near misses are necessarily reported, and the likelihood of reporting may vary according to perceived severity, operational workload, individual attitudes, and organisational priorities. Consequently, the dataset may under-represent routine, low-consequence, or operationally sensitive events while over-representing incidents that were considered noteworthy by personnel. This potential reporting bias should be considered when interpreting the prevalence and relative importance of identified topics and prevention measures.

Finally, the expert validation exercise revealed that semantic interpretation of near-miss narratives remained inherently subjective, even among experienced maritime practitioners. The relatively low levels of absolute agreement observed across some validation metrics suggest that alternative topic interpretations may be equally plausible. This limitation reflects a broader challenge associated with manual incident

classification and reinforces the importance of combining automated approaches with expert review when analysing complex safety narratives.

5.4. Recommendations for future research

Future research could build on these findings in several important ways. First, extending the analysis beyond a single operator's dataset would provide a stronger basis for comparative and generalisable insights across regions and the wider ferry sector. This would also act as an external validation of the identified topics, prevention measures, and association-rule relationships, helping determine which findings are operator-specific and which represent broader industry-wide safety patterns. Incorporating broader or multi-operator datasets could reveal systemic and cross-regional risk patterns, enabling a deeper understanding of how vessel types, routes, weather conditions, and sea states interact with specific incident types and preventive actions. Such integration would enhance both the explanatory and practical value of topic-based safety analysis.

Second, future studies would benefit from structured expert triangulation to refine and validate topic interpretations. The observed variability in expert feedback highlights an opportunity to formalise multi-reviewer validation frameworks that capture diverse operational perspectives and improve reliability across contexts.

Third, the use of GPT-assisted topic labelling shows clear promise for improving interpretability, but further exploration of hybrid labelling methods, combining machine-generated descriptors with iterative expert refinement, could balance efficiency with accuracy and generalisability.

Finally, methodological refinement in text pre-processing and outlier handling represents another key direction. Enhancing data quality and harmonising reporting practices across incident summaries would improve model stability and reduce bias from inconsistent terminology or phrasing. Advancing these areas would strengthen the robustness of NLP-based safety analytics and support more proactive, evidence-based risk management in the maritime domain.

6. Conclusion

This study applied a combined methodology of BERTopic, GPT topic labelling, association rule mining, and temporal analysis to ferry near-miss narratives, demonstrating the value of advanced NLP in maritime safety research. By uncovering 75 interpretable topics, we showed that data-driven approaches can both replicate existing taxonomies and reveal emergent or cross-cutting issues overlooked by predefined categories. The 87 prevention topics highlighted the critical role of crew vigilance, communication, and technical maintenance, underscoring the need for Safety-II perspectives that recognise not only failures but also the everyday actions that keep operations safe. Temporal patterns further illustrated how systemic shocks, such as COVID-19, reshape incident dynamics, while association rules revealed strong, actionable links between incident types and prevention strategies.

Together, these findings illustrate that unstructured near-miss reports are a rich but underutilised resource for proactive safety management. Leveraging transformer-based topic modelling and hybrid analytical pipelines can transform these narratives into actionable insights, supporting evidence-based interventions and advancing resilience in ferry operations.

Data availability

The data utilised in this study is held by an external organisation and cannot be shared publicly due to confidentiality agreements and data protection regulations.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Hollie Black: Conceptualization, Formal analysis, Methodology, Software, Validation, Visualization, Writing – original draft. **Louis de Wolff:** Conceptualization, Data curation, Validation. **Yasin Burak Kurt:** Validation, Writing – review & editing. **Yaser Farag:** Validation, Writing – review & editing. **Dwitya Waskito:** Validation, Writing – review & editing. **Rafet Kurt:** Conceptualization, Supervision, Writing – review & editing. **Osman Turan:** Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Ahadh, A., Binish, G.V., Srinivasan, R., 2021. Text mining of accident reports using semi-supervised keyword extraction and topic modeling. Process Saf. Environ. Prot. 155, 455–465. https://doi.org/10.1016/j.psep.2021.09.022. https://www.sciencedirect.com/science/article/pii/S0957582021004882.

Ahn, S.I., Kurt, R.E., Turan, O., 2022. The hybrid method combined STPA and SLIM to assess the reliability of the human interaction system to the emergency shutdown system of LNG ship-to-ship bunkering. Ocean Eng. 265. https://doi.org/10.1016/j.oceaneng.2022.112643, 112643-112643. https://www.sciencedirect.com/science/article/pii/S0029801822019266.

Ahn, S.I., Kurt, R.E., 2020. Application of a CREAM based framework to assess human reliability in emergency response to engine room fires on ships. Ocean Eng. 216, 108078 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0029801820310209.

Andrade, S.R., Walsh, H.S., 2023. Machine learning framework for hazard extraction and analysis of trends (HEAT) in wildfire response. Saf. Sci. 167, 106252 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925753523001947.

Bairami-Khankandi, S., Bolbot, V., BahooToroody, A., Goerlandt, F., 2025. A systems-theoretic approach using association rule mining and predictive Bayesian trend analysis to identify patterns in maritime accident causes. Reliab. Eng. Syst. Saf. 258, 110911 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0951832025001140.

Bao, J., Chen, Y., Yin, J., Chen, X., Zhu, D., 2023. Exploring topics and trends in Chinese ATC incident reports using a domain-knowledge driven topic model. J. Air Transport. Manag. 108, 102374 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0969699723000170.

Bicen, S., Celik, M., 2022. A hybrid approach to near-miss report investigation towards next-generation safety solutions on-board ships. Ocean Eng. https://doi.org/10.1016/j.oceaneng.2022.112768, 112768-112768. https://www.sciencedirect.com/science/article/pii/S0029801822020510.

Black, H., Rafet, K., Osman, T., 2026. A domain-adapted NLP pipeline for human factors classification of maritime accidents. Ocean Eng. 357, 125625. https://doi.org/10.1016/j.oceaneng.2026.125625.

Bulut Ozan, C., 2023. Shipboard compressor system risk analysis by using rule-based fuzzy FMEA for preventing major marine accidents. Ocean Eng. 272, 113888. https://doi.org/10.1016/j.oceaneng.2023.113888.

Bulut Ozan, C., Gizem, E., Yasin, A., 2023. A novel approach for the analysis of engineering complex system accidents: the casualty of the M/V Vitaspirit. Ocean Eng. 276, 114217. https://doi.org/10.1016/j.oceaneng.2023.114217.

Ceylan, B.O., Akyuz, E., Arslanoğlu, Y., 2022. Modified quantitative systems theoretic accident model and processes (STAMP) analysis: a catastrophic ship engine failure case. Ocean Eng. 253, 111187 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0029801822005911.

Coraddu, A., Oneto, L., Navas de Maya, B., Kurt, R., 2020. Determining the most influential human factors in maritime accidents: a data-driven approach. Ocean Eng. 211, 107588. https://doi.org/10.1016/j.oceaneng.2020.107588.

Costantino, F., Sabetta, N., 2025. Large Language Models for topic analysis: insights from robotic accident narratives. Procedia Comput. Sci. 253, 1462–1472 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050925002169.

Dongge, L., Liang, C., Bo, A., Rui, X., Xidong, W., Hengshuai, S., 2026. Determinants of casualty severity in maritime accidents: evidence from text mining of 581 investigation reports, 2014–2023. Ocean Eng. 351, 124456. https://doi.org/10.1016/j.oceaneng.2026.124456.

Egger, R., Yu, J., 2022. A topic modeling comparison between LDA, NMF, Top2Vec, and BERTopic to demystify Twitter posts. Front. Sociol. 7. https://doi.org/10.3389/fsoc.2022.886498, 886498-886498.

Emil Rijcken, F.S., Zervanou, Kalliopi, Spruit, Marco, Mosteiro, Pablo, Kaymak, Uzay, 2023. Towards interpreting topic models with ChatGPT. In: Presented at the the 20th World Congress of the International Fuzzy Systems Association, Korea, Republic of.

Fahreza, M.I., Hirata, E., 2024. Maritime piracy and armed robbery analysis in the Straits of Malacca and Singapore through the utilization of natural language processing. Marit. Pol. Manag. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85209904719&doi=10.1080%2f03088839.2024.2405157&partnerID=40&md5=5d7909b399e360e146b0074edf29004b.

Golden, A.S., Weisbrod, R.E., 2016. Trends, causal analysis, and recommendations from 14 years of ferry accidents. J. Public Transport. 19 (1), 17–27 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1077291X2200087X.

I. M. O. (IMO) MSC-MEPC.7/Circ.7 Guidance on Near-Miss Reporting, 2008.

Hasanspahić, N., Frančić, V., Vujičić, S., Maglić, L., 2020. Reporting as a key element of an effective near-miss management system in shipping. Saf. Now 6 (4), 53 [Online]. Available: https://www.mdpi.com/2313-576X/6/4/53.

Hasanspahic, N., Vujicic, S., Kristic, M., Mandusic, M., 2022. Improving safety management through analysis of near-miss reports—A tanker ship case study. Sustainability-Basel 14 (3), 1094. https://doi.org/10.3390/su14031094.

Hassel, M., Asbjørnslett, B.E., Hole, L.P., 2011. Underreporting of maritime accidents to vessel accident databases. Acc. Anal. Prev. 43 (6), 2053–2063. https://doi.org/10.1016/j.aap.2011.05.027.

Hebbar, A., A, M., Jr, M.Q., Hanzu-Pazara, R., 2023. Safety of Domestic Ferries: a Scoping Study of Seven high-risk Countries.

Heinrich, H.W., 1941. Industrial Accident Prevention. Mcgraw-hill Book Company Inc., New York and London.

Hetherington, C., Flin, R., Mearns, K., 2006. Safety in shipping: the human element. J. Saf. Res. 37 (4), 401–411. https://doi.org/10.1016/j.jsr.2006.04.007. https://www.sciencedirect.com/science/article/pii/S0022437506000818.

Hollnagel, E., 2014. Safety-I and Safety-II: the past and Future of Safety Management, 1St Ed. CRC Press.

Hong, W.-T., Clifton, G., Nelson, J.D., 2024. A data-driven conceptual framework for understanding the nature of hazards in railway accidents. Transp. Policy 152, 102–117 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0967070X24001240.

Irshad, L., Walsh, H., 2024. Identifying human errors and error mechanisms from accident reports using large language models.

Kim, H., Choi, J., Nam, Y., Youn, J.-H., 2022. Characteristic analysis of the built environment of ferry terminals: a case study of mokpo, South Korea. Sustainability 14 (4), 2115 [Online]. Available: https://www.mdpi.com/2071-1050/14/4/2115.

Kuhn, K.D., 2018. "Using structural topic modeling to identify latent topics and trends in aviation incident reports,". Transport. Res. C Emerg. Technol. 7, 105–122. https://doi.org/10.1016/j.trc.2017.12.018. https://www.sciencedirect.com/science/article/pii/S0968090X17303881.

Lichao, Y., et al., 2026. From text to network: a framework for identifying causal factors and risk propagation paths in maritime accidents. Reliab. Eng. Syst. Saf. 271, 112282. https://doi.org/10.1016/j.ress.2026.112282.

Lu, H., Cai, J., 2024. "Experimental comparison of three topic modeling methods with LDA. Top2Vec and BERTopic 1998, 376–391. https://doi.org/10.1007/978-981-99-9109-9_37.

Madsen, P.M., Dillon, R.L., Morris, E.T., 2025. Using near misses, artificial intelligence, and machine learning to predict maritime incidents: a U.S. Coast Guard case study. Risk Anal. 45 (4), 830–845. https://doi.org/10.1111/risa.15075.

Mertes, P.M., et al., 2024. Validation of a natural language processing algorithm using national reporting data to improve identification of anesthesia-related ADVerse evENTs: the "ADVENTURE" study. Anaesth. Crit. Care Pain Med. 43 (4), 101390 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352556824000481.

Njoku, I., 2021. Assessment of causal factors and effects of ferry mishaps in inland waterways in Lagos state, Nigeria. Int. J. Scient. Res. Mulidisc. Stud. 7 (9), 45–52.

Oltedal, H.A., McArthur, D.P., 2011. Reporting practices in merchant shipping, and the identification of influencing factors. Saf. Sci. 49 (2), 331–338. https://doi.org/10.1016/j.ssci.2010.09.011.

O'Neill, T.A., 2017. An overview of interrater agreement on likert scales for researchers and practitioners. Front. Psychol. 8, 777. https://doi.org/10.3389/fpsyg.2017.00777.

Preiss, C.A. A. Alexander J., Berghammer, Anthony, Bollenbacher, John, McCarthy, John V., Brom, Madeline G., Enger, Mike, Villacorta, Nicholas Rios, Straughn, Shaquavia, 2024. Evaluation of text cluster naming with generative large language models. J. Data Sci. 22 (3), 376–392. https://doi.org/10.6339/24-JDS1149.

Rose, R.L., Puranik, T.G., Mavris, D.N., Rao, A.H., 2022. Application of structural topic modeling to aviation safety data. Reliab. Eng. Syst. Saf. 224. https://doi.org/10.1016/j.ress.2022.108522, 108522-108522. https://www.sciencedirect.com/science/article/pii/S0951832022001776.

Sahoo, S., Maiti, J., Tewari, V.K., 2024. A framework to model contractors' hazard and risk exposure at process plants using unsupervised text mining. Process Saf. Environ. Prot. 183, 24–45 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957582023011643.

Salih Hüsrev, Ü., Bulut Ozan, C., Ali, E., 2026. Functional resonance analysis (FRAM) of a chemical tanker fire: a systematic approach to enhancing maritime accident analysis. Ocean Eng. 355, 125077. https://doi.org/10.1016/j.oceaneng.2026.125077.

Sandhåland, H., Oltedal, H., Eid, J., 2015. Situation awareness in bridge operations – a study of collisions between attendant vessels and offshore facilities in the North Sea. Saf. Sci. 79, 277–285 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925753515001630.

Simon, J., Christian, K., Willy, B., 1999. The importance of near miss reporting to further improve safety performance. J. Loss Prev. Process. Ind. 12 (1), 59–67. https://doi.org/10.1016/S0950-4230(98)00038-2.

Stroeve, S., et al., 2023. SHIELD human factors taxonomy and database for learning from aviation and maritime safety occurrences. Saf. Now 9 (1), 14. https://doi.org/10.3390/safety9010014.

Velibor, K., Bulut Ozan, C., Mitar, J., 2024. Reliable ships: a fuzzy FMEA based risk analysis on four-ram type hydraulic steering system. Ocean Eng. 314, 119758. https://doi.org/10.1016/j.oceaneng.2024.119758.

Xing, Y., et al., 2024. Discovering latent themes in aviation safety reports using text mining and network analytics. Int. J. Transp. Sci. Technol. 16, 292–316 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2046043024000297.

Yan, K., Wang, Y., Jia, L., Wang, W., Liu, S., Geng, Y., 2023. A content-aware corpus-based model for analysis of marine accidents. Accid. Anal. Prev. 184. https://doi.org/10.1016/j.aap.2023.106991, 106991-106991.

Yuan, C., et al., 2025. "Causal relationship analysis and risk assessment of contributing factors in coal mine ordinary accidents: a case study (2002–2023),". Process Saf. Environ. Prot. 197, 107065. https://www.sciencedirect.com/science/article/pii/S0957582025003325.