Proc. of the European Safety and Reliability Conference (ESREL 2026)
Edited by José C. Matos, Paulo B. Lourenço, Michael Beer, Daniel Oliveira,
Edoardo Patelli, Hélder S. Sousa, Rui A. Silva and Monica Santamaria

doi: 10.3850/ESREL2026061419_esrel26-p27788-cd

Latent Risk Pattern Discovery in Hydrogen Incidents via Retrieval-Augmented Generation

Lucas Alencar S. da Silva

CEERMA, Universidade Federal de Pernambuco, Recife, Brasil, E-mail: lucas.alencarsantana@ufpe.br

July Bias Macedo

CEERMA, Universidade Federal de Pernambuco, Recife, Brasil, E-mail: july.bias@ufpe.br

Caio Souto Maior

CEERMA, Universidade Federal de Pernambuco, Recife, Brasil, E-mail: caio.maior@ufpe.br

Márcio das Chagas Moura

CEERMA, Universidade Federal de Pernambuco, Recife, Brasil, E-mail: marcio.cmoura@ufpe.br

Isis Lins

CEERMA, Universidade Federal de Pernambuco, Recife, Brasil, E-mail: isis.lins@ufpe.br

The transition to a global hydrogen economy introduces safety challenges driven by situations such as flammability and risk of leakage and by the conditions necessary for its storage. In this high-risk context, rigorous analysis of past incidents is fundamental for mitigating risks and having good public confidence. This study proposes a methodology based on Retrieval-Augmented Generation (RAG) with a domain-oriented risk analysis framework to automate the extraction of lessons learned from hydrogen-related accident reports. While the HIAD 2.1 database already includes a specific field for "lessons learned", this automated model enhances the process by uncovering non-explicit insights and ensuring consistency across large datasets. The approach enables contextual inference, allowing the model not only to retrieve information explicitly stated in the reports but also to derive latent risk patterns, provide cross-incident comparisons, and identify gaps in causal investigation that are not readily observable. Rather than answering predefined high-level questions, the RAG model analyses individual event descriptions to generate structured technical interpretations and prevention alerts, which can subsequently support cross-incident comparisons and broader risk insights. The methodology is applied to the HIAD 2.1 database, which contains structured and unstructured descriptions of accidents and incidents involving hydrogen in vehicles, transport systems, processing plants, and storage facilities. The study compares the RAG-generated insights with the existing lessons in the database to evaluate the model's effectiveness. The results show that the approach reduces interpretive bias and increases analytical consistency by converting incomplete descriptions into actionable safety information. Overall, the framework supports a shift from manual, retrospective analyses to a data-driven decision support process, contributing to more consistent hydrogen safety analysis and supporting the safe deployment of hydrogen technologies.

Keywords: Retrieval-Augmented Generation, Hydrogen Safety, Risk Analysis.

1. Introduction

Hydrogen is considered one of the fundamental pillars of the energy transition and today's global economy. Playing a key role in decarbonization strategies, proper handling becomes necessary for its implementation, given that hydrogen has a wide flammability range and low ignition energy, meaning any failure can become catastrophic. The Hydrogen Incident and Accident Database (HIAD 2.1) aggregates structured records of real-world accidents in the hydrogen industry to support learning from experience and risk mitigation. HIAD is managed by the European Commission's Joint Research Centre (JRC) as part of the Hydrogen Safety Network of Excellence. HIAD compiles reports on industrial accidents related to hydrogen, with each event being reviewed and validated by JRC experts prior to publication. In 2017, HIAD was updated to HIAD 2.1 and integrated into the activities of the European Hydrogen Safety Panel (Fernández-Arias et al., 2026), (Galassi et al., 2012), (Guo et al., 2024).

HIAD 2.1 documents hydrogen-related accidents using detailed narratives, each with a unique event ID and associated tables. Each event may include a free-text description, counts of injuries and fatalities, a cause label, and additional attributes such as facility context, leak/ignition descriptors, lessons learned, and references. The completeness and technical depth of these fields vary according to the primary source. In risk analysis, the manual verification of documents often proves to be slow and inefficient in most cases due to the dependence on detailed accident descriptions, which can be scarce at times (Cornelissen, Van Hoof, and De Jong, 2017), (Daniele, Meda, and Castello, 2019).

To address these challenges, this work employs Retrieval-Augmented Generation (RAG) to support contextual analysis across hydrogen-related incidents that are otherwise treated as isolated cases within the HIAD dataset. RAG assists a large language model in producing structured technical interpretations that would be difficult to obtain from a single report alone by retrieving information from technically similar past events. For instance, when an incident description lacks information on the ignition mechanism or failure progression, the model can use analogous events to infer plausible contributing factors, highlight missing investigative elements, or identify recurring vulnerabilities under comparable operating conditions (Jiang, Han, and Ahmed-Kristensen, 2025), (Shuang et al., 2024), (Wen, Zhang, and Liu, 2022).

RAG is a machine learning technique that combines the strengths of information retrieval and large language models (LLMs). Unlike approaches focused solely on assigning cause labels, which often oversimplify complex accident dynamics, the RAG-based framework enables cross-incident reasoning. This allows the analysis to move beyond categorization and toward the identification of latent risk patterns, such as recurring thermal exposure of transport systems, degradation mechanisms in pressurized components, or systemic investigation gaps that are not evident in isolated reports. For example, if a report does not specify whether an accident resulted in injuries or fatalities, RAG retrieves descriptions of past events with similar circumstances and presents them as context for an LLM tasked with inferring the missing label, ensuring that predictions are based on domain-specific knowledge (Campari, Al-Hajj, and Smith, 2023), (He et al., 2025).

This aligns with recent literature highlighting retrieval as fundamental for improving factuality and task specificity in LLM outputs. While past work (Macedo et al., 2025), focused on assigning cause labels, a process that often simplifies the complexity of the event, the RAG architecture here allows for cross-incident analysis. Thus, the study evolves from a mere categorization system into a decision support, capable of identifying latent risk patterns, such as the recurring degradation of high-pressure seals in refueling stations or the specific thermal vulnerability of transport trailers to external heat sources and causal gaps that remain invisible in isolated analyses. Additionally, this study evaluates the model's effectiveness by comparing the RAG-generated summaries with the "lessons learned" already provided in the database. The goal is to transform complex technical narratives into safety intelligence and actionable preventive alerts, providing a deeper understanding that goes beyond identifying a failure category, thereby supporting the safe implementation of hydrogen technologies.

2. Methodology

The proposed methodology aims to support the extraction of safety-relevant information and the identification of latent risk patterns in hydrogen-related incident reports. To achieve this, the process begins with filtering and preprocessing the HIAD 2.1 database to ensure sufficient data quality for analysis. Next, a RAG (Retrieval-Augmented Generation) workflow is implemented, using past events as contextual evidence to assist an LLM, DeepSeek-V3, in classifying data whose causes are classified as unknown. Subsequently, the data containing "Environment" as the cause and those lacking information on injuries or fatalities are cleaned again. Shortly thereafter, another RAG is added to generate structured technical interpretations, highlighting plausible contributors to failure and gaps in the causal investigation. Finally, the model's performance is evaluated through a systematic comparison between the analytical summaries generated by the RAG and the original "lessons learned" recorded in the database, assessing the consistency and technical value added of the generated outputs relative to traditional manual categorization. in the production of structured technical interpretations, highlighting plausible contributors to failure and gaps in the causal investigation.

2.1 *Data Overview*

We work with HIAD 2.1, which contains 755 events. The records are assigned a Quality Seal (QS) ranging from 2 (most descriptors missing) to 5 (root cause analysis and lessons learned with comprehensive technical details). The distribution by quality is shown in Table 1, where the number of unknown causes is 99 (13.1% of all events) [3].

Table 1. Hiad dataset by quality seal (n = 755).

Quality Seal (QS)	Count	Percentage (%)
2	385	51.00
3	210	27.80
4	76	10.10
5	84	11.10
Total	744	100

The raw data undergoes preprocessing to filter out records that lack sufficient descriptive textual depth. This filtering primarily targets reports that do not provide enough semantic context for the RAG system to establish meaningful correlations or make contextual inferences, ensuring that the model processes only narratives with sufficient technical substance to support the discovery of latent patterns. The filtered dataset (304 events) is divided in a 90/10 ratio to comprise, respectively, the retrieval knowledge base and the test set for model evaluation. Subsequently, RAG is used to classify all data with a QS >=3 whose cause is defined as unknown by the knowledge base.

2.2 *RAG Generation Testing*

Following this, RAG is applied, receiving the 273 events as a knowledge base for retrieval and thus receiving the descriptions of the events separated for testing, while being instructed by a prompt to generate analyses containing as content: technical clarity, which performs the logical reconstruction of the event to validate the understanding of the scenario; latent risk patterns, which search for evidence of recurring operational failures in the sector; investigation gaps, which provide explicit signalling of crucial information missing from the report; and prevention alerts, which translate the failure into an engineering or management countermeasure, using the DeepSeek-V3 generator.

2.3. *RAG Framework Implementation*

For the RAG framework implementation, prompt engineering is a fundamental part. This stage is necessary to define how the model should behave and ensure that it does not just summarize the text but performs a technical evaluation of the risks. The framework uses an instruction template that directs the model toward the tasks of identifying and mitigating risks in the hydrogen sector. The structure of this prompt is documented below:

- ROLE: Safety analysis specialist, risk engineer, and HIAD 2.1 database data analyst.
- INPUTS: {event_description} and the reasoning guides in {context_str}.
- ANALYSIS GUIDELINES:
  - Technical Clarity: Explain the logical sequence of the failure (leak → ignition).
  - Latent Patterns: Identify recurring risks in hydrogen systems by comparing the current incident with the evidence retrieved from the HIAD database.
  - Investigation Gaps: Point out what is missing from the report (e.g., unknown ignition source, missing maintenance history).
  - Prevention Alert: Proactive suggestion based on the identified failure.
- TASK: Analyze the accident report using past cases to explain clearly how the failure started and evolved. Identify key information missing from the original text and suggest practical safety solutions to prevent future occurrences, turning the raw data into a useful guide for prevention.

- OUTPUT: Return only the "Analytical Summary" text (3 to 5 sentences). Do not include binary lists, numbers, headers, or introductions such as “Here is the analysis”.
- CONSTRAINTS: Be purely analytical; go beyond surface-level description; identify what the original text does not say (gaps); use technical safety engineering terminology

3. Results and Discussion

This section evaluates the performance of the proposed RAG-based methodology by comparing the original event descriptions and lessons learned available in the HIAD database with the analytical summaries generated by the model. The analysis focuses on assessing consistency, technical depth, and the model’s ability to compensate for incomplete or low-quality incident records.

Applying the methodology to the 31 test events revealed robust consistency in the behavior of the DeepSeek-V3 model under the RAG architecture. Prior to individual analysis, it was observed that, in 100% of the inferences, the model adhered to the output constraints—analytical summaries of 3 to 5 sentences—demonstrating compliance with the established prompt design. Notably, the technical analyses generated by the model for events that were originally viewed as having low-quality technical descriptions achieved a level of depth and clarity comparable to the original QS=4 or QS=5 records. This result indicates that contextual retrieval can effectively fill the information gaps found in sparse incident descriptions, elevating them to a higher standard of technical utility.

To provide a comprehensive yet concise overview of the model’s performance, the following 10 examples summarize the 31 test cases using the Latent Risks, GAPs, and Recommendations derived from RAG. In this framework, Latent Risks address hidden hazards in hydrogen handling, GAPs highlight critical points that should be present in the original reports but are absent, and Recommendations provide the model’s suggestions for preventing similar incidents. These cases cover a variety of operational contexts, including transportation and chemical processing, illustrating different modes of mechanical and operational failure.

i. Incident: Fire on a hydrogen column of an ammonia manufacturing unit; extinguished with powder and N_2 injection.
   - Latent Risk: Recurring failure type associated with hydrogen handling and ammonia synthesis infrastructure in high-pressure process columns.
   - GAP: The description lacks critical information regarding the root cause of the fire ignition and the initial integrity status of the hydrogen column.
   - Recommendations: Implement enhanced, real-time integrity monitoring and leak detection systems specifically for hydrogen columns.

ii. Incident: Explosion during TIG welding on a brazing furnace cover due to residual hydrogen gas.
   - Latent Risk: Improper purging and lack of gas testing procedures during hot work.
   - GAP: The absence of witnesses and the speculative cause ("apparently") create a significant gap in verifying the purge procedure failure and the exact ignition mechanism.
   - Recommendations: Mandate a strict, multi-step purge-and-test protocol and use calibrated detectors before authorizing any welding.

iii. Incident: Collision where the right front side of a Fuel Cell Electric Vehicle (FCEV) was impacted by another car.
   - Latent Risk: "Vulnerability of fuel cell electric vehicles (FCEVs) to collision-induced damage.
   - GAP: Critical information missing includes the integrity of the hydrogen storage and delivery systems post-collision and the specific cause of the impact.
   - Recommendations: Enhance FCEV design with robust, impact-resistant shielding for hydrogen system components located in high-risk frontal zones.

iv. Incident: Hydrogen leak and ignition at an electrical connection during a sintering furnace restart phase after maintenance.
   - Latent Risk: This incident, involving a leak and ignition during the restart phase of a maintained furnace, is a recurring

failure type linked to post-maintenance procedures and ignition at electrical connections.
  - GAP: The description does not specify the exact source or cause of the hydrogen leak or the condition of the ignition protection at the electrical connection.
  - Recommendations: Implement mandatory, verified purge and leak testing protocols for all hydrogen systems following maintenance, prior to energizing electrical components during restart

v. Incident: Exothermic reaction between steam and zircaloy in a nuclear reactor core, leading to hydrogen accumulation and a small explosion.
  - Latent Risk: Severe accident progression due to loss of coolant and metal-water chemical reactions.
  - GAP: The specific ignition source, precise timing, and the exact volume or pressure of the hydrogen bubble before the explosion is not detailed in this summary.
  - Recommendations: Design and implement robust, redundant cooling systems and rapid hydrogen venting/detection mechanisms.

vi. Incident: Pipe rupture in a nuclear station caused by the detonation of hydrogen generated by water radiolysis.
  - Latent Risk: Hydrogen accumulation in stagnant or low-flow sections of nuclear systems due to radiolysis.
  - GAP: The description does not specify the exact ignition source or trigger for the hydrogen detonation, nor the precise location of hydrogen accumulation within the pipe system.
  - Recommendations: Implement routine monitoring and venting protocols for hydrogen concentration in all auxiliary piping systems.

vii. Incident: Fire on insulating plastic elements of an alkaline electrolyzer; controlled by CO_2 extinguishers and potash drainage.
  - Latent Risk: Use of combustible materials near high-current electrical interfaces in electrolysis units.
  - GAP: The specific ignition source or cause of the fire on the insulating plastics is not detailed in the description.
  - Recommendations: Implement routine thermal monitoring and inspection of insulating materials on electrolysers and use flame-retardant alternatives where possible.
  -

viii. Incident: Explosion in a condenser area two hours after a unit trip caused by a stator earth fault. The electrical fault damaged hydrogen cooler tubes, allowing gas to migrate into the condenser, where it ignited.
  - Latent Risk: Secondary system failure where electrical faults compromise the integrity of hydrogen cooling circuits, creating hidden leak paths into non-classified areas like the condensate system.
  - GAP: The description does not confirm the definitive ignition source, only suggesting possibilities without conclusive evidence.
  - Recommendations: Implement post-trip isolation and monitoring procedures for hydrogen cooling systems to detect and prevent migration into unintended areas like condensers.

ix. Incident: Frontal crash test of an FCEV leading to impact on hydrogen storage components.
  - Latent Risk: Potential for catastrophic tank failure during high-energy collisions.
  - GAP: The description lacks all critical information, including the cause of the crash, the integrity of the hydrogen system post-collision, and any resulting fuel release or fire.
  - Recommendations: Enhance frontal crash safety standards and integrate automatic hydrogen system isolation mechanisms in FCEVs.

x. Incident: Naphtha and hydrogen gas leak from a gasoline reformer during the operation of a relief valve shortly after it had been repaired.
  - Latent Risk: Maintenance-induced failure or improper calibration of safety-critical relief valves.

- GAP: The description lacks explicit confirmation of ignition or explosion, details on the cause of the relief valve operation, and the effectiveness of the repair.
- Recommendations Implement post-repair testing and verification procedures for relief valves and associated systems before returning to service.

The analysis of the results shows that the model goes beyond a simple review, transforming accident reports into practical safety guidelines. By identifying Latent Risks, the model highlighted hidden problems that recur in different situations. For instance, in Case iii, which contains information on risks observed in hydrogen vehicles similar to Case ix, the model was able to identify a specific pattern: structural vulnerability in frontal impact zones. Similarly, the model correlates Cases iv and x, where failures occurred during system restarts after maintenance or repairs; thus, the models recognizes that the latent risk lies not only in the component but in the process itself. In the nuclear incidents v and vi, the model associates different gas generation sources with a shared hazard of accumulation in low-flow areas, recommending ventilation and monitoring solutions that address the problem systemically rather than in isolation.

Beyond pattern identification, the model was effective in pointing out information GAPs in the original reports that hinder a thorough root-cause analysis. In Cases i, vii, and x, for example, the system noted that the descriptions failed to specify the exact ignition source or the initial integrity status of the equipment, information vital for preventing recurrence. In Case ii, the model identified subjectivity in the investigation due to the use of speculative terms and the lack of witnesses, indicating that the company does not follow a strict protocol and testing procedure before authorizing hot work.

Finally, the technical recommendations generated by the system translate these analyses into concrete and systemic preventive actions. Instead of generic suggestions, the model proposed specific solutions based on the nature of each failure, such as the implementation of real-time thermal monitoring for the electrolyzers in Case vii and post-trip inspection protocols for the cooling tubes in Case viii. Additionally, the system reinforced the importance of physical barriers and protective shielding for critical components in vehicles (Cases iii and ix), as well as redundant ventilation systems to mitigate gas accumulation in confined environments (Cases v and vi). This ability to transform sparse data into prevention alerts, such as the requirement for mandatory leak tests before re-energization seen in Case iv, reinforces the tool's utility as robust support for safety management within the hydrogen economy.

4. Conclusion

This study validates the use of RAG as a powerful tool for safety in the hydrogen economy. The study demonstrates that safety-relevant information can be systematically extracted from complex and often low-quality incident reports by applying the DeepSeek-V3 model to the HIAD 2.1 database. The proposed framework enables the correlation of events that would otherwise be treated in isolation, assists in identifying latent risks and gaps in investigations, and enriches originally low-quality records by filling in missing information, generating structured technical interpretations and preventive recommendations. Thus, the approach contributes to greater consistency and technical depth in hydrogen safety assessments, supporting a safer deployment of hydrogen technologies.

Future work will focus on extending the framework toward comparative and operational robustness. This includes benchmarking the RAG-based workflow against alternative large language models to assess model sensitivity and generalizability across hydrogen-related domains. In parallel, the development of an expert-oriented interface will be explored to facilitate interaction, validation, and integration of the tool into existing safety analysis workflows.

Acknowledgement

The authors acknowledge the financial support of the Human Resources Program of the National Agency of Petroleum, Natural Gas and Biofuels – PRH-ANP, specifically PRH 38.1 entitled "Risk Management, Reliability and Sustainability in the Exploration and Production of Oil, Natural Gas and Renewable Energies", funded by ANP and managed by Fapesp (São Paulo Research Foundation), process number 2025/03497-0.

References

Campari, A., A. Al-Hajj, and S. Smith (2023) Lessons learned from HIAD 2.0: Inspection and maintenance to avoid hydrogen-induced material failures. *Computers & Chemical Engineering* 173, 108199.

Cornelissen, P. A., J. J. Van Hoof, and M. D. T. De Jong (2017). Determinants of safety outcomes and performance: A systematic literature review of research in four high-risk industries. *Journal of Safety Research* 62, 127–141.

Daniele, P., A. Meda, and S. Castello (2019). The Hydrogen Incidents and Accidents Database (HIAD 2.0). *International Journal of Hydrogen Energy* 44, 2341–2350.

Fernández-Arias, P., A. del Bosque, G. Lampropoulos, and D. Vergara (2026). Green hydrogen and its contribution to environmental sustainability: A review. *Resources* 15, 15.

Galassi, M. C., E. Papanikolaou, D. Baraldi, E. Funnemark, E. Håland, A. Engebo, G. P. Haugom, T. Jordan, and A. V. Tchouvelev (2012). HIAD - hydrogen incident and accident database. *International Journal of Hydrogen Energy* 37, 17351–17357.

Guo, L. J., J. Z. Su, Z. Q. Wang, J. W. Shi, X. J. Guan, W. Cao, and Z. S. Ou (2024). Hydrogen safety: An obstacle that must be overcome on the road towards future hydrogen economy. *International Journal of Hydrogen Energy* 51, 1055–1078.

He, X., G. D. Yang, D. P. Kong, and D. K. Miao (2025, June). Deep learning-based method for predicting emergency resources demand at hydrogen refueling stations. *International Journal of Hydrogen Energy* 142, 400–412.

Jiang, P., J. Han, and S. Ahmed-Kristensen (2025, December) A two-stage retrieval-augmented generation framework for producing sustainable product design guidelines. *Sustainable Futures* 10, 101094.

Macedo, J., Silva, L. A. S. da, Valcamonico, D., Moura, M. J. C., Baraldi, P., & Zio, E. (2025). Multi-Task Learning and Retrieval-Augmented Generation for Cause and Consequence Prediction on the HIAD Hydrogen Safety Dataset. In *Proceedings of ICSRS* 2025.

Shuang, Q., X. S. Liu, Z. J. Wang, and X. X. Xu (2024, September). Automatically categorizing construction accident narratives using the deep-learning model with a class-imbalance treatment technique. *Journal of Construction Engineering and Management* 150, 04024107.

Wen, J., J. Zhang, and X. Liu (2022). Statistics, lessons learned and recommendations from analysis of HIAD 2.0 database. *International Journal of Hydrogen Energy* 47, 17082–17096.