Proceedings of the 35th European Safety and Reliability & the 33rd Society for Risk Analysis Europe Conference
Edited by Eirik Bjorheim Abrahamsen, Terje Aven, Frederic Bouder, Roger Flage, Marja Ylönen

doi: 10.3850/978-981-94-3281-3_ESREL-SRA-E2025-P4039-cd

Enhancement of a Hydrogen Incident and Accident Database Using Large Language Models

Gianluca Tabella

Dept. Gas Technology, SINTEF Energy Research, Norway. E-mail: gianluca.tabella@sintef.no

Ivan De Fazio

Dept. Civil, Chemical, Environmental, and Materials Engineering, University of Bologna, Italy.
Dept. Mechanical and Industrial Engineering, NTNU Norwegian University of Science and Technology, Norway.
E-mail: ivan.defazio@studio.unibo.it

Mohammed Ayalew Belay

Dept. Electronic Systems, NTNU Norwegian University of Science and Technology, Norway.
E-mail: mohammed.a.belay@ntnu.no

Elena Stefana

Dept. Mechanical and Aerospace Engineering, Sapienza University of Rome, Italy.
E-mail: elena.stefana@uniroma1.it

Valerio Cozzani

Dept. Civil, Chemical, Environmental, and Materials Engineering, University of Bologna, Italy.
E-mail: valerio.cozzani@unibo.it

Nicola Paltrinieri

Dept. Mechanical and Industrial Engineering, NTNU Norwegian University of Science and Technology, Norway.
E-mail: nicola.paltrinieri@ntnu.no

Marta Bucelli

Dept. Gas Technology, SINTEF Energy Research, Norway. E-mail: marta.bucelli@sintef.no

Hydrogen holds significant potential for decarbonizing various industries, including energy and mobility. However, the limited availability of accident data poses a significant challenge to effective safety risk analysis and assessment. This study leverages *large language models* to address the critical task of filling gaps in the *Hydrogen Incidents and Accidents Database* (HIAD) 2.1, a prominent repository of hydrogen-related unwanted events. A three-step *Artificial Intelligence*-driven algorithm is proposed: (i) a preprocessing phase to standardize and prepare an event description, (ii) a processing phase utilizing OpenAI's *sentence embedding* technology to extract semantic relationships, and (iii) an enhancement phase employing trained *multilayer perceptrons* to impute missing data. The algorithm demonstrates promising results in predicting categorical entries and is applied to enhance the entire database, with a specific focus on the 2019 fueling station fire in Sandvika (Norway). This case study highlights the proposed algorithm's potential to improve our understanding of hydrogen-related incidents and contribute to enhanced risk management strategies.

Keywords: Hydrogen, HIAD, safety, large language model, artificial intelligence, deep learning, machine learning.

1. Introduction

Hydrogen, with its high efficiency and energy density, is a promising candidate for the decarbonization of industries such as transportation, manufacturing, and power generation (U.S. Department of Energy, 2025; IEA, 2021). However, its classification as an *extremely flammable* substance, coupled with its wide flammability range and low minimum ignition energy, introduces significant safety challenges (Campari et al., 2023). Hydrogen's physicochemical properties, including low density and boiling point, necessitate its storage and transport as either high-pressure compressed gas or cryogenic liquid (Guo et al., 2024). Additionally, hydrogen's incompatibility with several industrial materials, such as metals, can lead to embrittlement and degradation of mechanical properties (Abohamzeh et al., 2021). Consequently, hydrogen production, handling, storage, transfer, and use require stringent measures to ensure safety, prevent ignition, and protect people, assets, and the environment. Comprehensive risk analysis and assessment are critical for supporting operators and decision-makers in designing, planning, and operating hydrogen facilities. The widespread deployment of hydrogen technologies across industries underscores the importance of systematically recording and analyzing unwanted events to identify trends and root causes. The *Hydrogen Incidents and Accidents Database* (HIAD) serves as a valuable repository of hydrogen-related unwanted events, enabling risk assessment and the development of safety protocols (European Commission JRC, 2023; Wen et al., 2022; Alfasfos et al., 2024). HIAD 2.1 stores incident details across various attributes, such as ignition source, location, and release type. However, extracting meaningful insights from this database is challenging due to its unstructured, multimodal, and sparse nature, which limits the effectiveness of traditional analytical methods. Therefore, the manual processing of such a dataset is time-consuming and potentially leads to errors in safety assessments.

Recent advancements in *artificial intelligence* (AI), particularly in *deep learning*, provide promising tools for addressing these challenges. *Large language models* (LLMs), a class of *foundation models*, excel in information retrieval, semantic analysis, and data generation by leveraging billions of parameters trained on diverse datasets (Brown et al., 2020). Built on *transformers*, LLMs model complex relationships in sequential data through *self-attention* mechanisms (Vaswani et al., 2017). Examples include OpenAI's GPT models and Google's BERT (OpenAI, 2025; Devlin et al., 2019). LLMs use *sentence embeddings* to convert textual data into dense, fixed-dimensional numerical vectors that encode semantic meaning. Embedding methods, such as *word2vec* and transformer-based techniques, enable efficient computation for downstream tasks like clustering, classification, and prediction (Mikolov et al., 2013; Vaswani et al., 2017). In safety-critical databases like HIAD 2.1, these embeddings are instrumental in extracting semantic relationships and contextualizing event descriptions, facilitating missing data imputation. Despite their demonstrated success in analyzing complex datasets, applying LLMs and deep learning architectures to unstructured hydrogen safety data, such as in HIAD 2.1, remains unexplored. Leveraging these methods can significantly improve the extraction of patterns and relationships within incident data, leading to enhanced data quality and more robust risk mitigation strategies.

This paper presents a novel algorithm to enhance the HIAD 2.1 database by integrating LLM embeddings with *multilayer perceptrons* (MLPs). The proposed algorithm aims to address data sparsity, improve information completeness, and uncover hidden patterns in the dataset.

The remainder of this paper is structured as follows: Section 2 details the proposed database enhancement algorithm, including the training and validation procedures. Section 3 presents and discusses the study's main results. Finally, Section 4 concludes the paper and outlines potential future directions.

2. Proposed Method

The proposed algorithm is designed to effectively process individual event descriptions from the HIAD 2.1 database and infer missing event information. Formally, the HIAD 2.1 database can be expressed as a set D defined as:

$$D \triangleq \{E_1, \ldots, E_N\} \ , \tag{1}$$

where E_n represents a unique event in the database, with $n = 1, \ldots, N$. The database attributes (i.e., the column headers) are denoted as:

$$A \triangleq \{a_1, \ldots, a_M\} \ . \tag{2}$$

Consequently, the generic event description E_n can be represented as a set of attribute-value pairs:

$$E_n \triangleq \{(a_1, v_{n1}), \ldots, (a_M, v_{nM})\} \ , \tag{3}$$

where v_{nm} denotes the value of attribute a_m for event E_n. If the value for a specific attribute a_m is missing for an event E_n, it is represented as $v_{nm} = \emptyset$. This mathematical representation provides a rigorous framework for describing the proposed algorithm, ensuring clarity and precision in handling event data and attributes. The focus of this work is to fill in missing entries in the database, particularly for attributes that are *categorical* in nature, allowing a *finite* amount of *known* possible values. This constraint enables the design of robust algorithms capable of accurately inferring missing values within a structured domain.

2.1. *Algorithm Description*

The proposed algorithm operates in three sequential steps, as illustrated in Fig. 1:

(i) *Event preprocessing:* This step standardizes and processes the event description as it appears in the database. It ensures that data are formatted and cleaned for the next steps, eliminating inconsistencies and preparing the input for numerical representation.
(ii) *Sentence embedding:* Using an LLM, the preprocessed event description is transformed into a dense numerical vector via sentence embedding. These embeddings capture the semantic relationships within the text, enabling efficient downstream analysis.
(iii) *Event enhancement:* The resulting numerical embeddings are passed to a set of MLPs. Each MLP is specifically designed to predict and infer missing categorical data for a particular attribute in the event description.

2.1.1. *Event Preprocessing*

The HIAD 2.1 database aggregates data from diverse sources, including news reports, inspection documents, other databases, and scientific literature. These heterogeneous sources vary in quality and detail, necessitating a preprocessing step to filter out errors, irrelevant content, and biases from the event descriptions. Out of 61 database columns, 45 are identified as relevant. Columns such as “event ID”, “quality”, and “event title” are excluded because they provided minimal additional information. However, several of the retained columns exhibit inconsistencies or ambiguities, such as cases where “deflagration” and “detonation” are both marked as “yes”, despite being mutually exclusive outcomes. The following interventions are implemented to address these issues:

- *Unit consistency:* Missing units for variables are identified and added to the column header.
- *Correction of inconsistent entries:* Erroneous data are rectified. For example, instances where non-hydrogen substances are incorrectly reported as 100% hydrogen are corrected.
- *Elimination of noisy entries:* Values such as “not yet specified”, “unknown”, “not specified”, and “NaN” are replaced with blank cells to indicate missing data. Typos and inconsistencies in capitalization are also standardized.
- *Numerical uniformity:* Columns expected to contain numerical values but exhibiting non-numeric data are reviewed, and inconsistent entries are excluded.

These adjustments can be automated by *hardcoding* them into the preprocessing pipeline. Given an event of interest with its initial description E_0, the preprocessing step transforms it into a standardized representation E'_0:

$$E'_0 \triangleq \left(\left(a'_1, v'_{01}\right), \ldots, \left(a'_{M'}, v'_{0M'}\right)\right) \ , \qquad (4)$$

where M' denotes the revised number of attributes. Also here, a missing value is indicated as $v'_{0m} = \emptyset$.

2.1.2. *Sentence Embedding*

After preprocessing, the event description must be transformed into a numerical vector for subsequent processing. This transformation is achieved through *sentence embedding* technology. Since embedding models typically require a text string rather than a structured table as input, the event data are first concatenated into a single coherent text representation. The concatenation follows these guidelines:

- *Consistent field formatting:* Delimiters are standardized (e.g., “:” separates attributes from their values, and “–” is used for lists). Column headers are converted to uppercase.

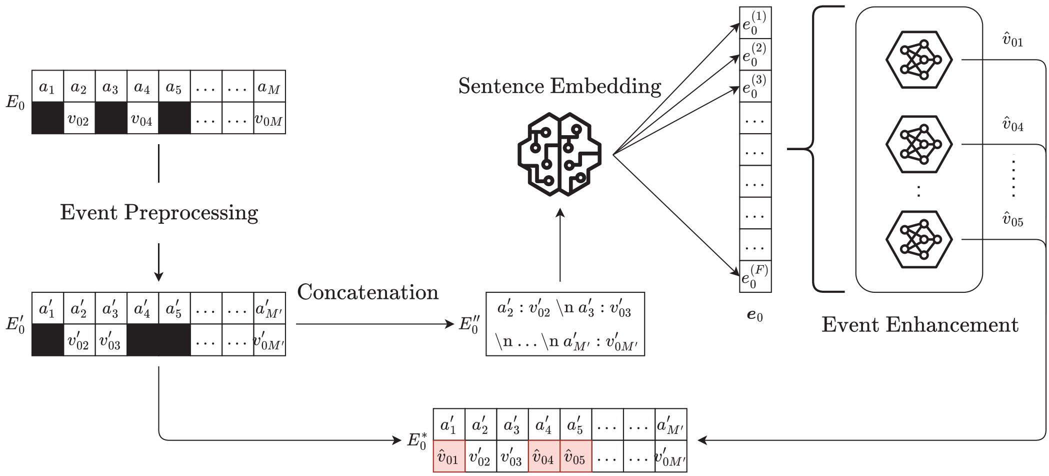

Fig. 1. Block diagram of the proposed algorithm.

- *Text standardization:* Unnecessary characters, such as multiple spaces, unwarranted line breaks, and superfluous symbols, are removed.
- *Conversion to bullet points:* Columns with complex or interrelated data (e.g., the *causes* column) are formatted as bullet points.
- *Separation of categories:* Different attributes are separated by line breaks, creating a structured text format.
- *Exclusion of empty fields:* Columns with missing values are omitted.

Following these steps, the event description undergoes the transformation $E_0' \rightarrow E_0''$, where E_0'' is a single text string representing the event's description in a structured format. This formatted text is then passed to a sentence embedding model. This study generates embeddings using OpenAI's *text-embedding-3-large* model (OpenAI, 2025). The embedding process converts the text string E_0'' into a *dense vector* $\boldsymbol{e}_0 \triangleq \left[e_0^{(1)} \cdots e_0^{(F)}\right]^{\mathrm{T}}$, where $F = 3072$ is the selected *embedding size*.

2.1.3. *Event Enhancement*

The embedding vector $\boldsymbol{e}_0$, generated as described in Section 2.1.2, is fed into multiple MLPs. The purpose of these MLPs is to infer information missing from the original database for the event of interest. Suppose that, after the preprocessing step described in Section 2.1.1, the event description is missing the value for the ith column, i.e., $v'_{0i} = \emptyset$. In this case, a specifically trained MLP predicts and fills the value of the ith column, denoted as $\hat{v}'_{0i}$. When applied across all relevant attributes, this step enhances the event description, transforming it into a more complete representation. This enhanced event can be formalized as:

$$E_0^* \triangleq \left((a'_1, v^*_{01}), \ldots, (a'_{M'}, v^*_{0M'}) \right) \,, \quad (5)$$

where the generic enhanced value v^*_{0i} is defined as:

$$v^*_{0i} \triangleq \begin{cases} \hat{v}'_{0i} \,, & \text{if } v'_{0i} = \emptyset \\ v'_{0i} \,, & \text{otherwise} \end{cases} . \quad (6)$$

Although each MLP operates independently to predict values for different attributes, they share a common structure:

- *Input layer:* Accepts the embedding vector $\boldsymbol{e}_0$ and has F (the embedding size) input nodes.
- *Hidden layers:* Consist of a constant number of nodes per layer, each employing the *ReLU activation function*.
- *Output layer:* Focuses on predicting categorical entries. Each node in this layer corresponds to a possible value of the attribute of interest. The *softmax activation function* assigns probabilities to each possible value. The predicted value corresponds to the node with the highest probability.

2.2. *Training and Validation*

The training and validation process for the MLP designed to predict values associated with the attribute a'_m involves the following steps:

(i) From the preprocessed database $D' \triangleq \{E'_1, \ldots, E'_N\}$, events missing the value of a'_m are removed. Specifically, if $v'_{nm} = \emptyset$, the event E'_n is excluded from D' for all n.
(ii) For the remaining events, the values of the attribute a'_m are hidden by setting $v'_{nm} = \emptyset$ for all n.
(iii) The remaining attributes of each censored event are concatenated, and sentence embedding is performed to generate numerical representations of the events.
(iv) The hidden variables are used as target variables and are encoded via *one-hot encoding*.
(v) 15% of the remaining events is set aside as the test set, while the other 85% is used for *cross-validation*.
(vi) Hyperparameters are selected, including the *number of hidden layers* ($\mathrm{N_{HL}}$), the *number of nodes per hidden layer* ($\mathrm{N_{NL}}$), and the *learning rate* (μ).
(vii) For the chosen $\mathrm{N_{HL}}$, $\mathrm{N_{NL}}$, and μ, a 10-fold cross-validation is performed. During each iteration, the network is optimized using *backpropagation* and the *Adam optimizer* to minimize the *cross-entropy loss function*, with *early stopping* applied based on the lowest validation loss achieved.
(viii) After completing the 10-fold cross-validation, the average of the lowest validation losses from each iteration is computed. This value is assigned to the corresponding configuration of $\mathrm{N_{HL}}$, $\mathrm{N_{NL}}$, and μ.

By iterating over steps (vi) to (viii) for various combinations of $\mathrm{N_{HL}}$, $\mathrm{N_{NL}}$, and μ, the optimal MLP configuration is identified as the one that minimizes the average validation loss.

3. Results and Discussion

This section describes the performance of the proposed algorithm on the test set. It provides an aggregated overview of the results obtained by applying the algorithm to enhance the entire HIAD 2.1 database and delves into insights from the Sandvika fueling station incident.

3.1. *Performance Evaluation*

The proposed model is evaluated using the test set. Performance metrics are *macro-averaged precision* (Prec), *macro-averaged recall* (Rec), and *accuracy* (Acc). This evaluation focuses on the following attributes: (1) *event initiating system*; (2) *classification of the physical effects*; (3) *nature of the consequences*; (4) *release type*; (5) *ignition source*; (6) *high pressure explosion*; (7) *high voltage explosion*; (8) *fire type*; (9) *application type*; (10) *specific application – supply chain stage*; (11) *storage/process medium*; (12) *location type*; and (13) *operational conditions*.

Table 1 presents the algorithm's performance after training and validating the MLPs using the procedure detailed in Section 2.2. The algorithm achieves performance above 0.5 across all three metrics for 10 out of the 13 selected columns. This indicates that the proposed algorithm effectively extracts information from event descriptions to infer missing table values accurately. The simultaneous achievement of high values of Prec, Rec, and Acc demonstrates the algorithm's effectiveness while mitigating the impact of class imbalance. However, the results also show that the proportion of missing values per attribute influences the performance. Attributes with higher absence rates tend to yield poorer results due to the limited number of events available for training and validation. Additionally, attributes with a larger number of possible values exhibit reduced performance, likely due to the higher complexity of the imputation task.

3.2. *Enhanced HIAD 2.1*

After the algorithm is presented and its performance evaluated, it is applied to the entire HIAD 2.1 database to fill in the missing data for the 13 attributes listed in Section 3.1. Due to the limited data availability, the MLPs are retrained following the procedure described in Section 2.2. However, in step (v), no portion of the remaining events is set aside for testing, allowing all available data to be used for training. Consequently, the algorithm imputes missing values within the

Table 1. Algorithm performances

Attribute	Prec	Rec	Acc	Absence (%)	No. values
1	**0.69**	**0.70**	**0.74**	0.79	2
2	**0.83**	**0.78**	**0.90**	1.07	3
3	**0.78**	**0.77**	**0.81**	0.00	5
4	**0.63**	0.40	**0.85**	54.97	4
5	0.12	0.15	0.27	87.15	11
6	**0.58**	**0.57**	**0.67**	81.99	2
7	**1.00**	**1.00**	**1.00**	85.83	2
8	**0.65**	**0.54**	**0.69**	86.49	4
9	**0.67**	**0.59**	**0.79**	2.52	12
10	**0.82**	**0.77**	**0.75**	3.71	12
11	**0.57**	**0.53**	**0.93**	11.52	4
12	0.36	0.39	**0.69**	43.31	4
13	**0.69**	**0.71**	**0.77**	50.86	2

Note: The metrics exceeding 0.5 are reported in bold.

investigated categorical attributes. This capability is particularly valuable for attributes with a high percentage of missing values, as they are essential for a comprehensive understanding of events and for proposing effective safety recommendations. For example, Fig. 2 and Fig. 3 illustrate the effect of the database enhancement on the *operational conditions* and *fire type* attributes, respectively, comparing the original data available in HIAD 2.1 with the algorithm's predicted values. In particular, results in Fig. 2 confirm that most events occur during normal operations (Wen et al., 2022). Understanding operational conditions during a broad set of undesired events serves as a critical basis for refining existing recommendations and identifying new ones related to personnel training, active and passive safety measures, equipment monitoring, and maintenance procedures (Wen et al., 2022). Additionally, uncovering patterns between operational conditions and other relevant attributes (e.g., *release type*) could provide valuable insights into safety barriers, benefiting a wide range of stakeholders, including policy-makers, hydrogen technology manufacturers, and plant operators.

3.3. *Sandvika Incident*

The developed algorithm is applied to the incident that occurred in 2019 at the hydrogen refueling station in Kjørbo, outside Oslo, Norway (Nel Hydrogen, 2019). In this incident, a hydrogen leak originated from the high-pressure storage unit due to the erroneous assembly of a specific plug in a hydrogen tank. The root cause was attributed to *human error*, as the inner bolts of the plug were not adequately torqued. The released hydrogen mixed with air, and ignition occurred when the concentration reached the flammability limits. Although the exact cause of ignition remains unclear, investigations suggest two possible triggers: auto-ignition or gravel movement underneath the storage unit. While no physical explosion occurred, the leaked hydrogen ignited in the open air, causing a pressure wave. It took approximately two hours to extinguish the fire fully. The incident caused significant disruptions, including traffic congestion and the closure of nearby roads. Three individuals were treated for minor injuries at the hospital, resulting from airbag deployments in vehicles near the site. No human fatalities or on-site injuries were reported.

For this analysis, the MLPs are retrained, excluding the Sandvika incident and treating it as a single-sample test set. Table 2 indicates that the incident description in HIAD 2.1 includes complete data for 9 out of 13 categorical columns. The table also presents the imputed values predicted by the model for the remaining four columns. Specifically, the *ignition source* is predicted as "other", aligning with reports that the cause of ignition remains unclear (Nel Hydrogen, 2019). Both *high pressure explosion* and *high voltage explosion* are correctly predicted as "no", as these scenarios do not apply to this incident. Finally, *operating conditions* is predicted as "normal", consistent with the conditions reported at the time of the incident.

To deepen the analysis, the imputation algorithm is reapplied to the event description after systematically censoring each of the nine known categorical values. The table shows that 5 out of 9 predicted values do not match the original database entries. However, these discrepancies highlight how multiple values can be appropriate descriptors for an attribute. In two instances, the predicted values are accurate alongside the original entries. For the category *specific application – supply*

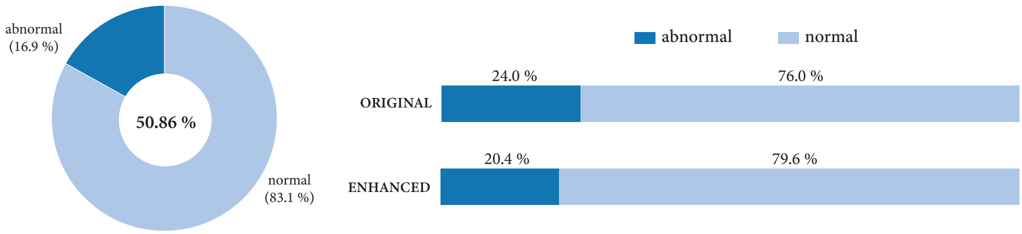

Fig. 2. *Operational conditions*. The circle shows inferred proportions and absence rate (% inside the circle).

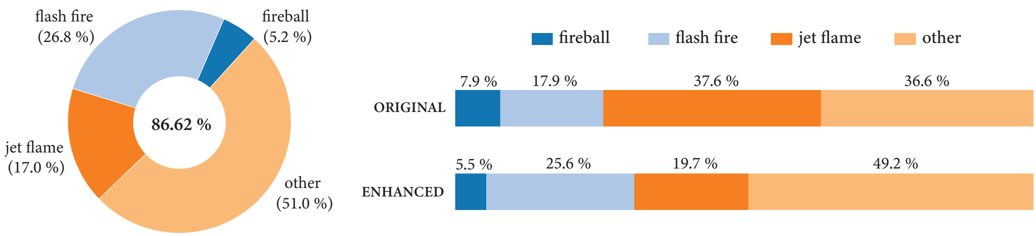

Fig. 3. *Fire type*. The circle shows inferred proportions and absence rate (% inside the circle).

Table 2. Proposed algorithm applied to the Sandvika incident.

Attribute	Original value and **imputed value**	Individual censoring	Individual censoring Attr. 3 always censored	Individual censoring Attr. 3 changed to “fire”
1	hydrogen system initiating event	✓	✓	✓
2	hydrogen release and ignition	✓	✓	✓
3	explosion	fire	fire (no new test needed)	fire (forced)
4	gas	✓	✓	✓
5	**other**			
6	**no**			
7	**no**			
8	jet flame	fireball	fireball	fireball
9	hydrogen refueling station	hydrogen production	✓	✓
10	hydrogen as a fuel	hydrogen storage	✓	✓
11	gas	✓	✓	✓
12	semiconfined	open	open	open
13	**normal**			

Note: A check mark (✓) indicates that the predicted value matches the value present in the original database.

chain stage, the model predicts “hydrogen storage” instead of “hydrogen as a fuel”. This prediction is contextually accurate, as the incident occurred while the hydrogen was contained in a storage tank despite its ultimate use as a fuel. Similarly, when the *nature of the consequence* category is censored, the model predicts “fire” instead of the original entry “explosion”. This is not an error, as the incident involved a continuous fire alongside the overpressure observed upon ignition.

As the incident is better described as a “fire” than an “explosion”, the analysis is extended under two settings: (i) permanently censoring the *nature of the consequence* category, and (ii) permanently changing it from “explosion” to “fire” before reanalysis. Interestingly, both experiments lead to an improvement in prediction outcomes. In both cases, categories that were previously predicted correctly retain their outcomes. At the same time, *application type* and *specific application – supply chain stage* are reclassified as “hydrogen refueling station” and “hydrogen as a fuel”, respectively, aligning with the original database entries. As a result, only *fire type* and *location type* remain misclassified.

This final analysis highlights the potential impact of inaccurate or imprecise information in an event description intended for enhancement. This is because while the MLPs are tuned using a combination of information directly from the training set and indirectly from the selected text embedding model, their outputs are heavily influenced by the quality of the input data. Since the algorithm treats this input as the ground truth, ensuring its accuracy is essential to achieve optimal performance when implementing the proposed AI solution.

4. Conclusions and Future Directions

We presented an algorithm integrating LLM embeddings with MLPs to enhance HIAD 2.1, addressing data sparsity and completeness challenges. Performance assessment and analysis of the Sandvika incident demonstrated promising results across most attributes, highlighting the algorithm’s flexibility in handling complex event descriptions. Key findings emphasize the importance of accurate input data and challenges like high absence rates and categorical complexity, underscoring the algorithm’s potential to improve hydrogen safety management by preventing similar incidents. Future directions include: (i) imputation of numerical values; (ii) tailored handling of data modalities; (iii) exploration of alternative embedding and enhancement models; (iv) deeper optimization of neural networks for better performance; and (v) application to other safety-related databases.

Acknowledgement

This publication was produced with support from the HYDROGENi Research Centre (hydrogeni.no), performed under the Norwegian Research Program FMETEKN (project no.: 333118). The authors acknowledge the industry partners and the Research Council of Norway for their contribution.

References

Abohamzeh, E., F. Salehi, M. Sheikholeslami, R. Abbassi, and F. Khan (2021). Review of hydrogen safety during storage, transmission, and applications processes. *J. Loss Prev. Process Ind. 72*, 104569.

Alfasfos, R., J. Sillman, and R. Soukka (2024). Lessons learned and recommendations from analysis of hydrogen incidents and accidents to support risk assessment for the hydrogen economy. *Int. J. Hydrogen Energy 60*, 1203–1214.

Brown, T. B., B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, and others (2020). Language Models are Few-Shot Learners. *Adv. Neural Inf. Process. Syst. 33*, 1877–1901.

Campari, A., E. Stefana, D. Ferrazzano, and N. Paltrinieri (2023). Analyzing Hydrogen-Related Undesired Events: A Systematic Database for Safety Assessment. In *Proc. 33rd ESREL*, Southampton (UK), pp. 420–427.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proc. NAACL HLT*, Volume 1, Minneapolis (MN), USA, pp. 4171–4186.

European Commission JRC (2023). HIAD 2.1.

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

IEA (2021). Net Zero by 2050 - A Roadmap for the Global Energy Sector. Technical report, IEA.

Mikolov, T., K. Chen, G. Corrado, and J. Dean (2013). Efficient Estimation of Word Representations in Vector Space. In *Proc. 1st ICLR*, Scottsdale (AZ), USA.

Nel Hydrogen (2019). Status and Q&A regarding the Kjørbo incident.

OpenAI (2025). Vector embeddings.

U.S. Department of Energy (2025). Hydrogen Basics.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin (2017). Attention Is All You Need. In *Adv. Neural Inf. Process. Syst.*, Volume 30, pp. 5998–6008.

Wen, J. X., M. Marono, P. Moretto, E.-A. Reinecke, P. Sathiah, E. Studer, E. Vyazmina, and D. Melideo (2022). Statistics, lessons learned and recommendations from analysis of HIAD 2.0 database. *Int. J. Hydrogen Energy 47*(38), 17082–17096.