

Review

Not peer-reviewed version

New Trends in the Use of Artificial Intelligence and Natural Language Processing to Occupational Risks Prevention

[Natalia Orviz-Martínez](#) , [Efrén Pérez-Santín](#) , [José Ignacio López-Sánchez](#) *

Posted Date: 11 October 2025

doi: 10.20944/preprints202510.0868.v1

Keywords: safety management; predictive safety strategies; real-time risk mapping; large language models; accident prevention



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

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

Review

New Trends in the Use of Artificial Intelligence and Natural Language Processing to Occupational Risks Prevention

Natalia Orviz-Martínez, Efrén Pérez-Santín and José Ignacio López Sánchez *

Escuela Superior de Ingeniería y Tecnología (ESIT), Universidad Internacional de la Rioja - UNIR, Avenida de la Paz, 137, 26006 Logroño, Spain.

* Correspondence: joseignacio.lopez@unir.net; Tel.: +34941210211

Abstract

Workplace safety remains a critical issue worldwide, with approximately three million workers dying annually each year from work-related accidents and diseases, highlighting the urgent need for more effective prevention measures. Technological advances, particularly in Artificial Intelligence (AI) and Natural Language Processing (NLP), offer innovative approaches to occupational risk prevention in various sectors. This literature review systematically examines the application of AI models, in particular, Large Language Models (LLMs) and NLP, to occupational risk prevention. The review includes studies published between 2013 and 2024, sourced from Web of Science and Google Scholar, using a combination of key terms related to occupational safety and AI technologies. The results show the increasing integration of AI and NLP across multiple industries, including aviation, construction, chemical, and transport sectors, to enhance safety management. Notable applications include real-time risk mapping, automated classification of safety incidents, and predictive modelling of occupational hazards. Thus, this review highlights the potential of AI and NLP technologies to transform occupational risk prevention by providing more accurate, efficient, and predictive safety strategies. However, challenges such as data quality, model transparency, and multilingual support remain. Future research should focus on addressing these limitations and further exploring AI and NLP applications to effectively mitigate workplace hazards.

Keywords: safety management; predictive safety strategies; real-time risk mapping; large language models; accident prevention

1. Introduction

Occupational Safety and Health (OSH) is a major concern for all countries worldwide (Zhang et al., 2019), and its management is an ongoing challenge to protect the health and safety of workers and ensure a safe and healthy working environment. The latest International Labor Organization (ILO) estimates that nearly three million workers die each year from work-related accidents and diseases, an increase of more than 5 per cent compared to 2015 (ILO, 2023), denoting an urgent need for more action to prevent work-related accidents and diseases. Although the statistics remain alarming, working conditions have improved tremendously over the years (Badri et al., 2018). Advances in science and technology, such as engineering controls, safer machinery and processes, collective and individual protective equipment, and the implementation of regulations and labor inspections, have significantly reduced the incidence of occupational accidents and diseases associated with industrialization (Kim et al., 2016). In addition, the development of a preventive culture in organizational settings have been crucial in optimizing health and safety management. Industrial development began worldwide in the eighteenth century (Sharma & Singh, 2020) and has been characterized by a series of events that have triggered different revolutions over the years, as shown in **Figure 1**. These revolutions have been driven by technological transformations that have

led to changes in the way industry operate and important social changes (Vinitha et al., 2020; Xu et al., 2018).

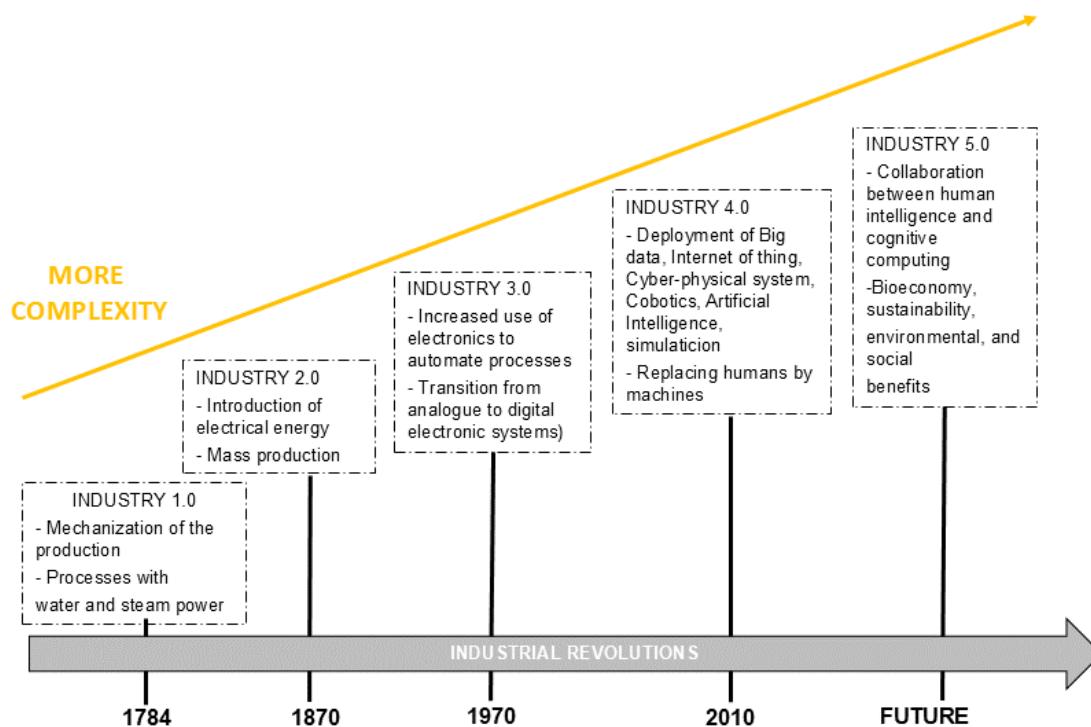


Figure 1. Evolution of industrial revolutions driven by technological transformations and increasing complexity.

The First Industrial Revolution began with the introduction of the power loom in 1784 and was characterized by shifts towards the intensification of work activities. In this period, water power and the steam engine played a decisive role, both in their contributions to industry and transport (Sharma & Singh, 2020). In the 19th century, the Second Industrial Revolution was born with the invention of electricity production, innovations in development, the use of new materials (alloys, synthetic plastics), mass production and assembly lines (Zhang & Yang, 2020). With the appearance of the first programmable logic controller (PLC) in 1969, the transition from the invention and manufacture of analogue to digital electronic devices, automation, and the incorporation of information technologies (ICT) into industrial processes, the Third Industrial Revolution was born (Fonseca, 2018), which encouraged the globalization of production and the relocation of jobs (Roberts, 2015).

The Fourth Industrial Revolution, or Industry 4.0 (early 21st century), is marked by technological developments with a certain autonomy and self-behavior, mainly focused on industrial automation robotic production through the integration of digital technology, information and communication technologies within an intelligent environment (Leesakul et al., 2022; Milea & Cioca, 2024; Gomez Miranda and Gonçalves, 2024). Specifically, this digital transformation is driven by technologies such as Blockchain, the Internet of Things, Big Data, Cyber-physical systems, Cobotics, Artificial Intelligence (AI), Natural Language Processing (NLP), Cloud Computing, Augmented Reality (AR) and Virtual Reality (VR), which aim to optimize production processes, increasing productivity and efficiency (Milea & Cioca, 2024). Finally, from 2021 onwards, the futuristic Fifth Industrial Revolution emerged, based on humans-robot collaboration to increase creativity and innovation by allowing robots to perform monotonous activities (Miraz et al., 2022).

On the other hand, the evolution in the field of OSH has always followed the different revolutionary advances in the industry (Badri et al., 2018) which has made it possible to react and propose effective solutions to be able to control occupational risks that may manifest themselves in

the face of technological advances, innovations, changes in working methods, organization, work teams, processes, products, and workplace itself. Indeed, the nature of a changing work environment brings with it a series of OSH challenges and opportunities, which is why its management is essential to ensure workers' health, business sustainability and social stability (Wang et al., 2020)

Over the years, in most industrialized and developed countries, reactivity has given way to proactivity (Badri et al., 2018) which promotes a fully preventive approach to OSH that allows action from the source to eliminate risks. This allows the appropriate decisions to be made sufficiently in advance to anticipate possible undesirable events that could harm workers. To this end, the incorporation of digitalization is now beginning to offer new opportunities to innovate, improve and address new and emerging risks in the field of occupational risk prevention, through the incorporation of neurocognitive computing technologies, AI and NLP. Recent studies highlight the role of AI in the risk of occupational disease, analyzing workplace hazards, and enhancing safety measures (Garvin & Kimbleton, 2021; Mollaei et al., 2023; Pishgar et al., 2021; Westhoven, 2022; Yimyam & Ketcham, 2022).

Thus, the evolution of occupational risk prevention has been characterized by a continuous expansion of its focus and methods. Initially focused on occupational health, it has transformed into a comprehensive approach that includes safety and health management and the early use of emerging technologies for the benefit of workers. This shift reflects a deeper understanding of workplace hazards, including chemical risks and psychosocial factors. The development of a preventive culture within organizational settings has been crucial in optimizing safety and health management systems. Recent strategies include dynamic risk assessment which helps organizations become better able to adapt to rapidly changing business or technological dynamics, putting them in a better position to respond to changes in business processes and their associated OSH risks (van Gulijk, 2021).

Today, this evolution is experiencing breakthrough, especially with the integration of AI and NLP. Recent studies highlight the role of AI in assessing the risk of occupational disease risk, analyzing workplace hazards, and enhancing safety measures. (Mollaei et al., 2023; Yimyam & Ketcham, 2022); Pishgar et al., 2021; Westhoven, 2022; Garvin & Kimbleton, 2021) Specifically, NLP is demonstrating great potential in processing and interpreting large data sets for risk analysis. These technologies are central to the development of more efficient, accurate, and predictive occupational risk prevention strategies.

The objective of this study is to conduct a systematic review of the application of Artificial Intelligence (AI) models, particularly Large Language Models (LLMs) and Natural Language Processing (NLP), within the domain of occupational risk prevention. This review aims to elucidate how these advanced technologies are being employed across various industrial sectors, including aviation, construction, the chemical industry, and transportation, to improve workplace safety and risk management. Specifically, the study will identify key technological applications such as real-time risk mapping, automated safety incident classification, and predictive modeling of occupational hazards. Furthermore, it seeks to address current challenges related to data quality, model transparency, and multilingual support while providing insights for future research to overcome these limitations and advance the efficacy of AI-driven occupational risk prevention strategies.

2. Methodology

This literature review explored the application of AI models, specifically LLM and NLP, in different areas of occupational risk prevention. The review focused on relevant studies published between 2013 and 2024, using a systematic search strategy across multiple databases and considering specific inclusion criteria.

2.1. Data Sources and Search Strategy

The primary sources for this review were the Web of Science and Google Scholar databases. The search was carried out using specific combinations of key terms: "occupational risk prevention" AND

“artificial intelligence”, “occupational risk prevention” AND “Large Language Models”, AND “occupational risk prevention” AND “Natural Language Processing”. In addition to direct retrieval of articles, a snowballing technique was employed to review the citations within these works and the articles that cited them. This approach ensured comprehensive coverage of the existing literature relevant to the research topic.

2.2. Inclusion Criteria

To be included in this review, studies had to meet several criteria:

1. **Publication venue:** Only articles published in peer-reviewed scientific journals indexed in the Journal Citation Reports (JCR) were considered.
2. **Publication period:** Studies had to be published between 2013 and 2024.
3. **Content relevance:** The studies needed to contain the specified key terms either in the title, abstract, or keywords or deal with the application of specific methods based on AI, LLM, and NLP in the context of occupational risk prevention.
4. **Methodological focus:** Each study was manually reviewed to ensure that it specifically applied LLM and NLP methodologies to the field of occupational risk prevention.

2.3. Screening and Selection Process

The initial search yielded a large number of potential articles. Each article was subjected to a multi-step screening process. Firstly, the titles and abstracts were reviewed to identify studies that appeared to meet the inclusion criteria. Articles passing this stage underwent a full-text review to confirm the relevance and ensure that the research applied specific LLM and NLP methodologies in the context of occupational risk prevention. Studies were excluded if they did not directly address the application of these technologies in this specific area or if they were not empirical research articles or reviews published in peer-reviewed journals.

2.4. Data Analysis and Synthesis

The included studies were categorized according to of application in areas across different industrial sectors, reflecting the diverse contexts in which LLM and NLP technologies have been used for occupational risk prevention in recent years. Data extracted from the studies included the type of AI technology used, the specific methods and models employed, the industrial context, the objectives of the application, and the reported outcomes or findings. This structured approach facilitated the identification of trends, gaps, and emerging themes in the use of AI for occupational risk prevention.

This methodology ensured a comprehensive and systematic review of the literature, focusing on the application of LLM and NLP models in occupational risk prevention. By adhering to strict inclusion criteria and employing a rigorous screening process, the review provides a robust foundation for understanding how these advanced AI methodologies are being applied across various industrial sectors to enhance workplace safety and risk management. The selected works were those that met the inclusion criteria, as shown in Figure 2.

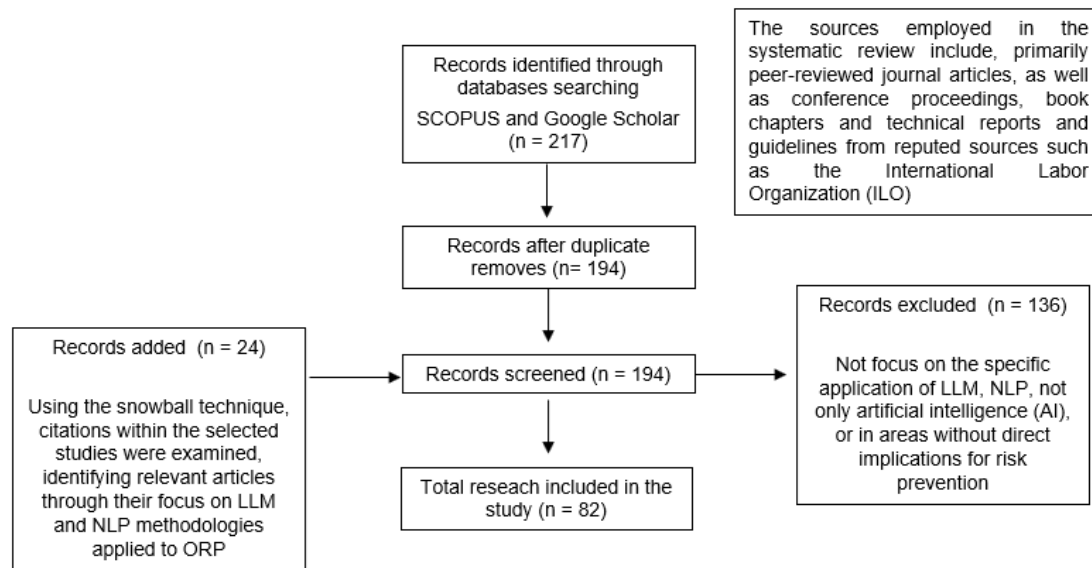


Figure 2. Selection process for articles included in the review.

3. Results

As pointed out by Zhao *et al* (2018), the benefits of using NLP methods in occupational risks prevention are multiple; for example, these approaches allow valuable information to be extracted and processed from large amounts of data. Future research directions include pattern recognition, in-situ identification of actual events, and fully automated methods (Zhao *et al.*, 2018). More specifically, previous reviews have highlighted the high potential of AI, LLMs and NLP methods in different areas of occupational risk prevention, as summarized in **Table 1**, such as exploring the impact of NLP applications in the field of aviation safety (Yang & Huang, 2023; Kierszbaum & Lapasset, 2020), and in other safety-critical industries such as transport, medical and construction (Ricketts *et al.*, 2023), as well as for occupational injury analysis (Khairuddin *et al.*, 2022), unveiling the influential aspects of this field through descriptive and scient metric analyses (Sarkar & Maiti, 2020).

Table 1. Application of artificial intelligence (AI) and natural language processing (NLP) in occupational risk prevention.

| Appli cation | Mode l | Domain/Dataset | Advantages | Limitations | Years of review | Ref |
|------------------|--------|---|---|---|-----------------|----------------------|
| Aviati on Safety | NLP | Analysis of aviation incident/accident reports and air traffic control communications | 1. Enhance situational awareness 2. Reduce workload 3. Improve decision-making capabilities | 1. Ambiguity in language interpretation 2. Scarcity of adequate training data 3. Lack of multilingual support | 2010-2022 | (Yang & Huang, 2023) |

| | | | | | | | |
|---|----------|---------------------------------------|------------------|--|---|-----------|-------------------------------|
| Aviation on safety | BER T | Aviation Reporting dataset | Safety System | 1. About 70% accuracy correctly answering the posed question 2. Uncovers information does not present in the dataset | 1. More questions are necessary to improve the model 2. Transparency of the model | 2011-2019 | (Kierszbaum & Lapasset, 2020) |
| Safety - critical indus tries | NPL | Safety reports | occurrence | 1. Automatically classifies occurrence reports 2. Extract critical information 3. Allows semantic searches | 1. Limited availability of occurrence reporting databases 2. Data privacy restrictions | 2012-2022 | (Ricketts et al., 2023) |
| Occupatio nal injury | NPL | Narratives occupational reports | from injury | 1. Classify accident types 2. Identify causal factors 3. Predict occupational injuries | 1. Low quality and quantity of data 2. Unbalanced data distribution 3. Inconsistent terminologies | 2016-2021 | (Khairuddin et al., 2022) |
| Occupatio nal injury | ML | Occupational accident analysis | | 1. Prediction of incident outcomes 2. Extraction of rule-based patterns 3. Prediction of injury risk 4. Prediction of injury severity | 1. Review focused on citation network analysis, with no critical comments on limitations | 1995-2019 | (Sarkar & Maiti, 2020) |

Natural Language Processing (NLP); Bidirectional Encoder Representations from Transformers (BERT); Citation Network Analysis (CNA); Machine Learning (ML).

LLM and AI are demonstrating great potential for development in different areas of occupational risk prevention, in sectors such as aviation and construction, even in specific risks such as Fall from Height (FFH), and in the chemical industry, as well as in the transport system, including railway, in the nuclear power generation sector, and for the protection of mine workers and to avoid medical errors.

Among the advantages that we can find in the use of methodologies based on AI, LLM and NLP in the field of occupational risk prevention, some are particularly interesting due to their high general applicability to multiple sectors, such as the generation of risk maps in real-time. For example, the application of dynamic real-time analysis using multimodal data fusion to enhance occupational risks prevention through the development of risk maps for workplaces, using machine/deep learning techniques by analysing data from diverse sources such as images, videos, documents, mobile applications and sensors/IoT. Thus, the combination of computer vision, NLP techniques, and sensor data analysis enables automated root cause identification, damage prevention, and disaster recovery, dynamically updating risk assessments in real-time. (Dalal & Bassu, 2020).

It is also worth mentioning that an important part of the success of the application of LLM and NLP-based methods lies in their ability to extract and analyze information in an automated way from large datasets contained in reports (e.g., accident reports), where the information can be structured to address a variety of problems, such as the limitations of generic and static checklists, which often do not apply to specific workplace contexts (Westhoven & Jadid, 2023), or more interestingly, the information may not have been previously structured.

Thus, in relation to the use of unstructured information, recent research highlights innovative integrations of AI, specifically through NLP and Machine Learning (ML), to refine safety and risk assessments. For example, Kamil, Taleb-Berrouane et al. (2023) combine a variety of NLP and text mining techniques with fuzzy set theory to transform unstructured accident reports into useful data, a methodology that contrasts with others used by Hou et al. (2022), who rationalize incident classification using NLP techniques for text vectorization. On the other hand, Paraskevopoulos et al. (2022) extend the functionality of AI in safety management by introducing a multimodal architecture that synergizes textual and visual data, distinguishing it from other studies primarily focused on text. In addition, Zhao et al. (2020) and Macedo et al. (2023) extend text analysis in different ways, as Zhao focuses on summarizing accident reports, while Macedo aims to correct inaccuracies in report. Furthermore, Baker et al. (2020a) and G. Liu et al. (2021) both refine data extraction and prediction methods, but differ in their approaches, as Baker emphasizes predictive modelling for safety outcomes, while Liu explores causal relationships using clustering techniques. Similarly, Dorsey et al. (2020) and Ekramipooya et al. (2023) aim to improve data quality and analysis efficiency through the use of NLP and AI methods.

Next, a perspective is presented on the impact of methodologies based on AI, LLM and NLP on the advancement of occupational risk prevention in different industrial sectors is presented.

3.1. Aviation

In the field of aviation safety, Miyamoto et al. (2022) and Dong et al. (2021) both used NLP techniques to analyze aviation safety reports. Miyamoto et al. focused on categorizing the causes of flight delay using clustering techniques, revealing maintenance issues as the primary cause. In contrast, Dong et al. combined NLP with deep learning models to automate the identification of primary factors in incident reports, demonstrating superior performance over traditional methods but limiting their scope to the most frequent incident categories. Moreover, Jiao et al. (2022) introduced a novel classification scheme using the XGBoost classifier and OC-POS vectorization to identify risk factors from Chinese aviation reports, indicating great potential for broader applications. Similarly, Kierszbaum et al. (2022) developed a compact, domain-specific language model,

demonstrating that specialized pre-training can effectively address the scarcity of domain-specific data in aviation safety NLU tasks, highlighting a trend towards creating more specialized NLP and AI tools tailored to specific data challenges in aviation safety. In addition, Madeira et al. (2021) investigated human factors in aviation incidents, using a hybrid approach of semi-supervised and supervised learning to tackle the challenge of limited labelled data sets, a common issue in AI applications in safety analysis. This study aligns with the work of Rose et al. (2020), who also used NLP and clustering to categorize and visualize safety narratives, but with a focus on integrating numerical and text-based data to enhance accident investigation processes. These studies highlight a significant trend towards using advanced AI and NLP methods to dissect and understand large volumes of aviation safety data, as it is shown in Table 2.

Table 2. Applications of AI and NLP methodologies in the analysis of aviation safety data.

| Objective | Methodology | Results | Reference |
|---|---|---|-------------------------|
| Categorize and visualize the textual narratives from safety incident reports from the Aviation Safety Reporting System (ASRS) | NLP and clustering techniques, K Means clustering and t-distributed Stochastic Neighbor Embedding (t-SNE) | Seven major categories and 23 sub-clusters of flight delay causes were identified, revealing that maintenance issues, rather than weather conditions, are the main contributors to delays. | (Miyamoto et al., 2022) |
| Analysis of voluminous aviation incident reports to prevent occupational hazards | NLP techniques: Universal Language Model Fine-Tuning (ULMFiT) and Averaged Stochastic Gradient Descent Weight-Dropped LSTM (AWD-LSTM) for unsupervised language modelling and text classification. Deep recurrent neural networks and attention-based Long Short-Term Memory (LSTM) models. | High accuracy in predicting multiple primary factors, providing a better understanding of incident factors, but limited to the six most common incident categories, with rarer categories not addressed due to insufficient data. | (Dong et al., 2021) |
| Classify and extract risk factors from Chinese civil aviation incident reports, which are | Machine learning: Extreme Gradient Boosting (XGBoost) classifier, combined with Occurrence | Identification of incident causes from 25 empirically determined factors covering equipment, | (Jiao et al., 2022) |

| | | | |
|--|--|---|---------------------------|
| traditionally underutilized due to their incoherence, large volume, and poor structure. | Position (OC-POS) vectorization strategy. | human, environmental, and organizational domains. | |
| Comparison of two language models in aviation safety: pre-trained ASRS-CMFS and RoBERTa model, without domain-specific pre-training. | Natural Language Understanding (NLU) and fine-tuning. | The RoBERTa model's size advantage does not outperform the ASRS-CMFS, which demonstrates greater computational efficiency. This highlights the advantage of pre-training compact models in scenarios where domain-specific data is limited. | (Kierszbaum et al., 2022) |
| Prediction of human factors in aviation safety incidents, identification and classification of human factor categories in aviation incident reports. | NLP for feature extraction, coupled with semi-supervised Label Spreading (LS) and supervised Support Vector Machine (SVM) techniques for data modelling. Use of TF-IDF models as an alternative to Doc2Vec (D2V), and Bayesian optimization to find near-optimal hyperparameter combinations | The semi-supervised LS algorithm is particularly suitable for classification with fewer labels, while the supervised SVM is more reliable for larger and more uniformly labelled datasets. | (Madeira et al., 2021) |
| To enhance flight safety by analyzing aviation safety reports | NLP with preprocessing routines, in particular TF-IDF text representation model for document classification. Categorization and | Robust and repeatable framework for identifying class categories in aviation safety event narratives, capable of identifying 31 class | (Rose et al., 2020) |

| | | | |
|--|--|---|------------------------|
| | visualization of narratives through k-means clustering and t-distributed Stochastic Neighbor Embedding (t-SNE) and post-processing through metadata-based statistical analysis | categories for ASRS event narratives | |
| Management and analysis aviation incident reports | Advanced NLP and text mining techniques, including algorithm design for active learning approaches, document content similarity methods, and topic modelling using TreeTagger and Gensim library | A range of developed tools to improve access to and analysis of aviation safety data | (Tanguy et al., 2016) |
| Overcome the difficulties of manually reviewing over 45,000 aviation reports. | Automatic text classification. Random forest algorithm for ICAO Occurrence Category | Text classification with an accuracy range of 80-93% | (de Vries, 2020) |
| Prevention of occupational hazards in aviation safety by efficiently extracting critical information from complex narratives | Common pattern specification language and normalized template expression matching in context | Overcome previous issues in these narratives, handle variants of multi-word expressions and improve accuracy. | (Posse et al., 2005) |
| Automated identification of human factors in aviation accidents | NLP techniques, Semantic Text Similarity approaches, Distributional Semantic theory, Vector Space Model (VSM), and document embeddings, integrated with the | Precision rate exceeding 86% and 30% reduction in time and cost compared to conventional methods | (Perboli et al., 2021) |

| | | | |
|---|--|---|----------------------|
| | Software Hardware Environment Liveware (SHEL) accident causality model | | |
| Improve the analysis of accident reports, by overcoming the limitations of effective analysis of unstructured information | Automated, semi-supervised, domain-independent approach | User-defined classification topics and domain-specific literature, such as handbooks and glossaries, to autonomously identify and categorize domain-specific keywords with an average classification accuracy of 80%, rivalling traditional supervised learning methods | (Ahadh et al., 2021) |
| The critical issue in the analysis of aviation safety reports is the reliance on manually labelled datasets for traditional classification modelling, which has proven to be inadequate. | Latent Dirichlet Allocation (LDA) topic modelling to cluster aviation safety reports into meaningful sets for subsequent analysis. | Considerable reduction in dependence on aviation experts and improves in flexibility and efficiency | (Luo & Shi, 2019) |
| Delve into the vast repository of over a million confidential aviation safety incident reports within the Aviation Safety Reporting System (ASRS) to uncover latent structures and hidden trends. | NLP and structural topic modelling, demonstrating flexibility and reduced dependence on subject matter experts | Uncover previously unreported issues, such as fuel pump, tank, and landing gear problems, while underscoring the relative insignificance of smoke and fire incidents in private aircraft safety | (Kuhn, 2018) |

| | | | |
|---|---|--|--------------------------|
| Visualization of safety narratives to prevent occupational risks through the integration of NLP techniques | Latent semantic analysis (LSA) to uncover latent relationships and interpret meaning within safety narratives, followed by isometric mapping to project this information. | Primary safety problems at the different phases of flight were revealed | (Robinson, 2016) |
| Classification of aviation safety reports to avoid the time-consuming and resource-intensive process of manual categorization and classification narratives | NLP models with ULM-FiT procedures | Outperforming alternative models, increasing the F1 score from 0.484 to 0.663. | (Marev & Georgiev, 2019) |

3.2. Construction

Additionally, occupational risk prevention (ORP) in the construction industry has made significant advances through the integration of AI and NLP, as resumed in **Table 3** and commented on from now on. For example, Shen et al. (2022) proposed a novel integration of Building Information Modeling (BIM) with an ontology-based safety rule library and NLP, creating a dynamic safety rule-checking system for construction sites and automating the identification of safety risks. In addition, Baker et al. (2020b) used advanced deep learning techniques, including Convolutional Neural Networks and Hierarchical Attention Networks, to analyze construction accident reports, enabling visual interpretation of model predictions to identify injury precursors. Also, Thompson et al. (2020) introduced a Named Entity Recognition scheme tailored to construction safety documents, which aims to structure free-text data into safety strategies. On the contrary, Xu et al. (2021) applied a text mining approach using an information entropy weighted term frequency metric to efficiently extract safety risk factors from metro construction reports, providing a novel quantitative tool for safety risk assessment through the effective analysis of extensive text data.

Furthermore, Fang et al. (2020) and Gadekar & Bugalia (2023) improved text classification in construction safety reports, focusing on the use of Bidirectional Transformers (BERT) for deep learning-based text classification and innovating with a semi-supervised model, respectively, achieving high accuracy with reduced dependence on pre-labelled data. In addition, F. Zhang et al. (2019) and Wang et al. (2021) addressed the classification and categorization of construction accident causes and safety risks, respectively, using different ensemble and text mining techniques to improve the precision of safety data analysis, while Tian et al. (2023) and Fan & Li (2013) allowed the effective information retrieval and analysis from construction documents, employing different AI techniques to optimize the extraction and utilization of safety-related data. Finally, Cheng et al. (2020) combined advanced NLP preprocessing with novel AI methods to enhance the accuracy and effectiveness of construction safety analyses from accident reports. Taken together, these studies highlight a significant shift towards automated, precise, and effective methods of risk identification and safety management in construction through the advanced integration of AI and NLP tools. In relation to the

Fall From Height (FFH) risk, Ben Abbes et al. (2022) using NLP to analyze and extract crucial information from the DBkWik database, sourced from 40,000 wikis, by means of knowledge graphs (KGs), highlighting their advantages over deep learning methods, particularly in addressing the limitations of the latter.

Table 3. Advances in occupational risk prevention in the construction industry through AI and NLP implementation.

| Objective | Methods | Results | Reference |
|--|---|--|-------------------------|
| To establish an automatic inspection mechanism | Use of NLP to integrate Building Information Modeling (BIM) with a safety rule library. | Development of a safety rule-checking system for the construction process | (Shen et al., 2022) |
| Identify injury precursors from construction accident reports to predict and prevent workplace injuries. | Convolutional Neural Networks (CNN) and Hierarchical Attention Networks (HAN), combined with Term Frequency-Inverse Document Frequency (TF-IDF) and Support Vector Machines (SVM) | Improve the understanding, prediction, and prevention of in the workplace injuries and provide tools that allow users to visualize and understand the predictions. | (Baker et al., 2020) |
| Effective management of occupational risks in the field of construction safety | NLP with a Named Entity Recognition (NER) scheme specifically designed for the construction safety domain | Effective and reliable annotator scheme with an agreement rate of 0.79 F-Score, overcoming previous limitations such as scope issues within hazard classification and the lack of coverage for specific construction activities, body parts injured, harmful consequences, and protective measures | (Thompson et al., 2020) |
| Identification of the critical causes of metro construction accidents in China | Development of a text mining strategy incorporating metric - information entropy weighted term | Successful extraction of 37 safety risk factors from 221 metro construction accident reports, | (Xu et al., 2021) |

| | | | |
|---|--|---|---------------------------|
| | frequency ($TF - H$) - metric to evaluate the importance of terms | demonstrating effective distillation of important factors from accident reports regardless of their length | |
| Analysis of near-miss reports to prevent potential accidents in the construction industry | Bidirectional Transformers for Language Understanding (BERT) for automatic classification of near-miss data | Outperforms the performance of other current state-of-the-art automatic text classification methods | (Fang et al., 2020) |
| Occupational risk prevention in the construction industry using NLP and semi-supervised machine learning techniques | Yet Another Keyword Extractor (YAKE) with Guided Latent Dirichlet Allocation (GLDA). | Effectiveness of the YAKE-GLDA approach, achieving an F1 score of 0.66 for OSHA injury narratives and an F1 score of 0.86 for specific categories, significantly reducing the need for manual intervention. | (Gadekar & Bugalia, 2023) |
| Text mining and NLP techniques are used to classify accident causes and identify common hazardous objects from construction accident reports. | Five baseline models (Support Vector Machine, Linear Regression, K-Nearest Neighbor, Decision Tree, Naive Bayes) and an ensemble model, with the Sequential Quadratic Programming (SQP) algorithm to optimize the weights of classifiers within the ensemble | Optimized models in terms of average weighted F1 score, even with low support, enabling automatic extraction of common objects responsible for accidents. | (Zhang et al., 2019) |
| Extract and categorize safety risks from records, focusing on high-frequency but | Text mining Word2Vec models integrated with NLP. | 7 unsafe-act-related and nine unsafe-condition-related risks were uncovered, | (Wang et al., 2021) |

| | | | |
|---|---|--|--------------------------|
| low-severity risks that are often missed by traditional methods. | | revealing predominant inappropriate human behaviors and the primary sources of safety hazards on-site | |
| Mining of safety hazard information in construction documents presented in unstructured or semi-structured formats. | Term recognition models using semantic similarity and information correlation and term frequency-inverse document frequency methods (TF-IDF). | Automatic extraction and visualization of safety hazard information. | (Tian et al., 2023) |
| Effective retrieval of relevant historical cases to prevent occupational risks in the construction industry. | Euclidean distance measure, cosine similarity measure, and the co-occurrence, and structured term vector model to represent unstructured textual cases. | Demonstration of the superior information retrieval of NLP-based models over traditional methods in a construction management information system | (Fan & Li, 2013) |
| More effective precautionary strategies and, consequently, improved safety assessments for construction projects. | Symbiotic Gated Recurrent Unit (SGRU) using NLP for text data preprocessing. | Improved classification accuracy and removal of human error in accident analysis and root cause identification. | (Cheng et al., 2020) |
| Prevention of Fall From Height (FFH) accidents in the context of occupational safety. | NLP combined with knowledge graphs (KGs). | A robust approach to enhance occupational safety, using NLP and knowledge graphs, to mitigate FFH risks and improve prevention strategies. | (Ben Abbes et al., 2022) |

3.3. Chemical Industry

The integration of AI and NLP in chemical industry safety has the potential to enhance occupational and environmental safety (see **Table 4**). Thus, Kamil, Khan et al. (2023) used NLP, Interpretive Structural Model (ISM), and probabilistic techniques to predict and analyze fire and explosion risks, leveraging accident databases for predictive accuracy in safety management

practices. On the other hand, Kabir et al. (2023) improved the accuracy of flare system failure analyses in the oil and gas industry by integrating traditional Fault Tree Analysis (FTA) with Dynamic Bayesian Networks (DBNs). In contrast, Kumari et al. (2022) advanced incident prediction by means of Artificial Neural Networks (ANNs) for cause and sub-cause analysis, surpassing traditional models to offer causation clarity. Moreover, B. Wang & Zhao (2022) introduced a novel deep learning framework combining BERT, BiLSTM-CRF, and CNN models to automate the extraction and classification of risk factors from accident reports in confined spaces, addressing the manual labor-intensive and subjective traditional analysis. Additionally, Xu et al. (2022) and Jing et al. (2022) utilized deep learning for analyzing accident causes, applying a CNN model to classify causes and deploying a combination of LSTM and attention mechanisms to enhance text classification of chemical accidents, respectively.

Furthermore, X. Luo et al. (2023) explored the use of NLP to automate the analysis of chemical accidents, categorizing risk factors to support decision making in risk analysis. Also, Macêdo et al. (2022) used BERT models for text mining to enhance quantitative risk analysis in oil refineries. Lastly, Song and Suh (2019) innovated in the detection of anomalies in accident reports by applying a text mining-based method to examine the narratives of accident reports.

Table 4. Application of AI and NLP techniques to enhance safety management in the chemical industry.

| Objective | Methodology | Results | Reference |
|--|---|--|-----------------------|
| Predict adverse events by learning from experience in the chemical industry. | NLP combined with Interpretive Structural Model (ISM) in a probabilistic approach | Identify critical factors that contribute to fire and explosion incidents, mainly management issues and lack of procedures and training. | (Kamil et al., 2023) |
| Analyze and improve the understanding of flare system failures in the oil and gas industry. | Fault Tree Analysis (FTA) and Dynamic Bayesian Network (DBN) approaches | A comprehensive and accurate assessment of flare system reliability is provided. | (Kabir et al., 2023) |
| Predicting and preventing incidents in aboveground onshore oil and refined products pipeline | Artificial Neural Networks (ANNs) use models to predict root causes and sub-causes using 108 incidents relevant attributes. | 80-92% accuracy range in predicting incident causes and sub-causes for aboveground onshore oil and refined products pipelines. | (Kumari et al., 2022) |
| Reduce occupational risks associated with confined spaces work by automatically extracting and classifying | BERT-BiLSTM-CRF and CNN models | Effective quantification and frequency estimation of the contributory factors contributing to risks associated with | (Wang & Zhao, 2022) |

| | | | |
|---|--|---|-----------------------|
| contributory factors from accident reports. | | work in confined spaces | |
| Improve hot work accident prevention in the chemical industry through an automated system that can classify and predict the causes, overcoming the limitations of manual analysis of unstructured accident records. | AAI and LLM models, such as the Latent Dirichlet Allocation (LDA) model for topic extraction and Convolutional Neural Networks (CNN) for cause prediction | F1 score of 0.89 in predicting key causes of hot work accidents in the chemical industry | (Xu et al., 2022) |
| Extracting information from free text chemical accident reports to enhance the prevention of occupational risks. | NLP and AI techniques combine word embedding and bidirectional long-short-term memory (LSTM) models with attention mechanisms. | The classification of accident causes, including unsafe acts, behaviors, equipment, material conditions, and management strategies, with identification of common trends, characteristics, causes, and high-frequency types of chemical accidents, had an average precision (p) of 73.1% and recall (r) of 72.5%. | (Jing et al., 2022) |
| Accident prevention in the chemical industry, using NLP to construct a knowledge graph of chemical accidents. | The NLP model is named entity recognition (NER), and it uses SoftLexicon and BERT-Transformer-CRF to structure and store accident knowledge in a Neo4j graph database. | Automatic extraction and categorization of risk factors from 290 Chinese chemical accident reports, outperforming previous models. | (Luo et al., 2023) |
| Enhance the early stages of quantitative | Text mining and fine-tuned trained | Identified potential accident outcomes | (Macêdo et al., 2022) |

| | | | |
|--|---|--|-------------------------|
| risk analysis (QRA) to prevent occupational risks associated with hazardous substances. | bidirectional encoder representations from transformers (BERT) models. | and ranked them by severity and probability, achieving mean accuracies of 97.42%, 86.44%, and 94.34%, respectively. User-friendly web-based app called HALO (hazard analysis based on language processing for oil refineries). | |
| Detection of anomalous conditions in accidents by mining text information from accident report documents. | AI and NLP, with text mining-based Local Outlier Factor (LOF) algorithm | Four major types of anomaly accidents in chemical processes were identified, and risk keywords were extracted and compared to provide a comprehensive view of the anomalous conditions. | (Song & Suh, 2019) |
| NLP application for unsupervised anomaly detection and efficient evaluation of chemical accident risk factors. | A Variational Autoencoder (VAE) is used for unsupervised anomaly detection in industrial accident reports. Doc2Vec is utilized as the 'Vector Space Model'. | Quantitative risk factors are extracted from narrative-based accident reports using an outlier factor (OF) function. The six most anomalous accident reports were identified. | (Rybak & Hassall, 2021) |

3.4. Transport System

The application of NLP and AI has the potential to enhance the accuracy and efficiency of risk assessment and safety management in Transport Systems (see **Table 5**). Thus, Hughes et al. (2019) used an AI-based model to extract and categorize terms from multilingual incident reports through the application of NLP techniques, achieving a high accuracy rate in categorizing safety incidents in public transport. Also, Valcamonico et al. (2022) also enhanced road safety analysis by integrating Hierarchical Dirichlet Processes and Doc2Vec with machine learning classifiers, showing how combined models can better balance accuracy and explainability in automated report classification. Moreover, Jidkov et al. (2020) focused on maritime risk assessment, employing deep learning and various NLP techniques to capture, process, and analyze data related to maritime safety events such as piracy, hijackings, and smuggling, improving incident classification and information extraction.

Meanwhile, Wang & Yin (2020) employed text mining and automatic association rules such as the FP-Growth algorithm to uncover key risk factors in China's transport sector, providing insights into systemic issues affecting safety. Additionally, Zhang et al. (2021) introduced the use of NLP and deep learning to analyze aviation accident reports with predictive purposes and safety management in aviation. More recently, Ricketts et al. (2022) proposed the use of NLP, rule-based phrase matching and a trained NER model to enhance hazard identification in HAZOP studies of aircraft subsystems, approaching the continuous model refinement and more efficient safety actions.

Specifically, in relation to railway safety and risk prevention, NLP and AI techniques have recently been used to innovate in incident prediction and management. For example, Hughes et al. (2018) developed a semi-automated classification system for close call reports in the GB railway industry, using NLP to associate incident reports with bow-tie accident causation models, with practical applications in categorizing a vast array of unstructured safety-related text. In contrast, Figueres-Esteban et al. (2016) used visual text analysis to extract safety information from the GB railways' Close Call System, highlighting its potential to identify risks despite the linguistic variation different reporter groups. Also, Wu et al. (2020) introduced NLP methods to improve subway accident decision-making processes in metro accidents with high precision in retrieving relevant past cases and advancing automated accident response systems. Moreover, Heidarysafa et al. (2018) applied deep learning to enhance the accuracy of accident labelling in the US railway sector and advanced the automatic classification of accident causes from narrative texts. Also, Ebrahimi et al. (2023) used NLP and Random Forest to develop a machine learning model capable of predicting evacuation needs following hazardous materials incidents on railways, mapping causal evacuation factors to improve emergency management. Furthermore, Hua et al. (2019) and Liu & Yang (2022) used text mining to improve risk identification in railway safety, extracting accident risk factors from Chinese railway accident reports through convolutional neural networks, and using deep learning techniques to quantify risk relationships in British railway incidents, respectively.

Table 5. Enhancement of risk assessment and safety management in transport systems through AI and NLP applications.

| Objective | Methodology | Results | Reference |
|---|---|--|----------------------------|
| Enhance occupational risk prevention in the transport system through the application of NLP and AI. | Text cleansing, tokenizing, tagging, and clustering, followed by analysis through NLP and a graph database to facilitate the querying of incident reports. | A true positive rate of 98.5% on a dataset of 5065 incident reports from the Swiss Federal Office of Transport, written in German, French, or Italian. | (Hughes et al., 2019) |
| Previous limitations in the expert interpretation of accident reports for road safety analysis have been overcome due to the voluminous nature of textual reports and the | NLP with textual report representations with Hierarchical Dirichlet Processes (HDPs) and Doc2vec, and ML-based classification by means of Artificial Neural Networks (ANNs), Decision | Accurate automatic extraction of the critical factors influencing road accident severity from accident reports. | (Valcamonico et al., 2022) |

| | | | |
|--|--|---|-----------------------|
| subjectivity of expert judgments. | Trees (DTs), and Random Forests (RFs), applied to a repository of road accident reports from the US National Highway Traffic Safety Administration | | |
| Development of a robust AI-based system capable of analyzing, categorizing, and extracting relevant information from unstructured maritime data sources, to assist in the prediction and prevention of maritime incidents. | DL and NLP are used to identify, classify and extract relevant maritime incident reports. NLP techniques include the bag-of-words approach, Named Entity Recognition (NER), and advanced word embeddings like Word2Vec, FastText, and BERT. ML models include convolutional neural networks (CNN), artificial neural networks (ANN), and long short-term memory (LSTM) networks optimized using Keras Tuner for hyperparameter tuning. | Accuracy up to 98.6% for binary incident classification. Incident date extraction achieved 61.8% accuracy | (Jidkov et al., 2020) |
| Assess and identify key risk factors in maritime accidents through text mining applied to accident reports. | Text mining and association rule mining using the FP-Growth algorithm | The main problems related to maritime accidents were unveiled, including overloading, poor navigational visibility, inadequate sailor competence, and insufficient government | (Wang & Yin, 2020) |

| | | | |
|---|---|--|---------------------------------|
| | | supervision of shipowners and shipping companies. Practical recommendations were made to government and regulatory bodies | |
| Predict traffic accidents by learning from textual data describing event sequences. | Data labelling from the National Transportation Safety Board (NTSB) accident investigation reports and Long Short-term Memory (LSTM) neural networks to predict adverse events. | Prototype query interface to predict and analyze traffic accidents from accident investigation reports. | (Zhang et al., 2021) |
| Automatic extraction of hazards, causes, and consequences from free-text occurrence reports to validate and refine safety measures for aircraft subsystems | NLP framework with rule-based phrase matching, combined with a spaCy Named Entity Recognition (NER) model. | Improved hazard identification system capable of reducing manual intervention to accurately determine causes, consequences, and hazards in HAZOP studies of aircraft transport systems. s. | (Ricketts et al., 2022) |
| Extraction of safety-related information from a large number of close call records in the GB railway industry, previously unfeasible for human analysis due to their sheer volume | NLP is applied to the analysis of free-text hazard reports and application to accident causation models, with categorization based on specific tokens. | Semi-automated technique for classifying close call reports in the GB railway industry. | (Hughes et al., 2018) |
| Extracting safety information from GB railways' Close Call System records, which accumulate | Visual text analysis techniques to extract safety information from GB railways' | The evaluation used 150 datasets covering incidents such as trespassing, slip/trip hazards, and level- | (Figueres-Esteban et al., 2016) |

| | | | |
|---|---|---|----------------------------|
| over150,000 text-based archives that are unmanageable using traditional methods | Close Call System records. | crossing issues. It showed that the method worked well with small and controlled data groups of data but not with larger datasets from different groups of people describing things in many different ways. | |
| Enhance the efficiency and accuracy decision making in metro accident response. | NLP techniques to automate the annotation of accident cases to facilitate information retrieval and Case-Based Reasoning (CBR) and Rule-Based Reasoning (RBR) to efficiently determine the most appropriate actions based on existing regulations and emergency plans | Average accuracy of 91%. | (Wu et al., 2020) |
| NLP application to the prevention of occupational risks avoiding railroad accidents in the United States. | NLP with advanced word embeddings like Word2Vec and GloVe. | Precise classification of accident causes from report narratives, with improved classification accuracy related to the increase in the number of reports analyzed. | (Heidarysafa et al., 2018) |
| Predicting the need for evacuation following railway incidents involving hazardous materials (hazmat) while simultaneously. | NLP and co-occurrence network analysis to scrutinize railway incident descriptions and supervised machine learning models, mainly Random | Elucidation of causal relationships through detailed network mapping of causes and contributing factors to emergencies in hazardous | (Ebrahimi et al., 2023) |

| | | | |
|--|--|---|--------------------|
| | Forest (RF), to evaluate the impact of different variables on evacuation prediction. | materials (hazmat) railway incidents. | |
| Analyze Chinese railway accident reports to better prevent future accidents. | NLP and text mining techniques, specifically a multichannel convolutional neural network (M-CNN) and a conditional random field (CRF) model are used to extract critical accident risk factors from text data. | Efficient extraction and summarization of risk factors. | (Hua et al., 2019) |
| Improvement of occupational risk prevention in railway safety. | Hidden Markov model, conditional random field (CRF) algorithm, bidirectional long short-term memory (Bi-LSTM), and Bi-LSTM-CRF deep learning network for named entity recognition of the reports. Random forest (RF) algorithm to standardize entity classification. Knowledge graph (KG) for railway hazard identification and risk assessment with a visual representation of the relationships between hazards, incidents, and accidents in the railway system. | The visualization and quantification of potential risk factors is needed to provide more effective railway risk prevention measures for railways. | (Liu & Yang, 2022) |

3.5. Other Sectors

As shown in **Table 6**, other sectors, such as nuclear energy, mining and the prevention of medical errors, also benefit from the integration of AI-based methods, particularly LLM and NLP. In relation to the application of NLP techniques to the nuclear power generation sector, Zhao et al. (2019; 2018) advanced the field by integrating NLP and multimodal data fusion to automatically identify causal relationships in event reports. They used a rule-based expert system, the Causal Relationship Identification (CaRI), to effectively capture causal associations with a success rate of 86%. On the other hand, Dalal & Bassu (2020) explored the development of “risk maps” by applying machine learning models to analyze data from sensors and computer vision systems to achieve a dynamic real-time capability to identify risks and prevent workplace accidents. The combination of NLP and AI methods in the field of occupational risk prevention has also recently led to several studies related to mine safety. Thus, Ganguli et al. (2021) carried out automatic data analysis from Mine Health and Safety Management Systems (HSMS) using NLP and Machine Learning (ML), specifically through the development of nine Random Forest (RF) models, demonstrating high accuracy and improved incident categorization.

In contrast, Shekhar and Agarwal (2021) applied text mining of fatality reports to enhance safety in Indian mines, identifying trends and patterns and highlighting the most vulnerable worker demographics and high-risk times periods. Furthermore, Qiu et al. (2021) combined text mining with complex network analysis to identify and quantify factors contributing to coal mine accidents, revealing complex interaction mechanisms and critical causal links, and providing a detailed map of accident causation pathways.

In the context of preventable medical errors, Cohan et al. (2017) proposed the use of NLP techniques to improve the identification and categorization of harmful events in patient care narratives, enhancing patients’ safety. For example, the complexity of clinicians’ reports represents a significant challenge, for which the authors used convolutional and recurrent neural networks, coupled with an attention mechanism, to provide an effective strategy for analyze the textual data to more accurately identify harmful events, categorizing their severity, and outperform existing approaches in detecting these events in large datasets of patient reports. Additionally, Denecke (2016) used NLP methods in the analysis of critical incident reports to address the underutilization of these reports in healthcare to enhance patient safety and quality of care, solving the difficulty and time-consuming nature of retrieving and analyzing these reports manually.

Table 6. Applications of AI, LLM, and NLP in enhancing safety across nuclear energy, mining, and healthcare sectors.

| Objective | Methodology | Results | Reference |
|--|---|--|---------------------|
| Enhance the safety and operation of nuclear power plants by automatically analyzing event reports, using NLP to efficiently extract and identify causal relationships. | The rule-based expert system, named Causal Relationship Identification (CaRI), has been augmented with a curated set of 11 keywords and 184 rules to identify causal relationships. | CaRI system successfully captures 86% of the causal relationships within the test data, surpassing inefficient manual procedures due to the immense volume and unstructured nature of these reports. | (Zhao et al., 2019) |
| Automated analysis of event reports from the nuclear power | Manual keyword identification is followed by the use of | 85% success rate in identifying causal relationships. | (Zhao et al., 2018) |

| | | | |
|---|--|--|---------------------------|
| generation sector, specifically focusing on the US Nuclear Regulatory Commission Licensee Event Report database. | Stanford CoreNLP for automated analysis and the identification of causal relationships. | | |
| Automate the analysis of Mine Health and Safety Management Systems (HSMS) data. | NLP and ML methods, with 9 Random Forest (RF) models developed to classify narratives from the Mine Safety and Health Administration (MSHA) database into nine different accident types | Models dedicated to individual categories outperformed those designed for multiple categories. 96% Successful automated classification, as confirmed through manual evaluation. | (Ganguli et al., 2021) |
| Prevention of fatal and non-fatal injuries through the automated analysis of Directorate General Mines Safety (DGMS) fatality reports for non-coal mines in Indian. | Data Acquisition from annual reports, followed by TM and NLP applications with Python libraries (Pandas, NumPy, and Sci-Kit Learn) to format the data, followed by Regular expressions (RegEx) to detect patterns. Later, NLP techniques were applied, tokenization was used using the SpaCy library, and part-of-speech (POS) tagging was used using Python's NLTK library. Finally, Python's Matplotlib for data analysis, using Seaborn libraries, along with | The most common accidents involve falling objects impacting workers aged between 28 and 32, specifically the 'mazdoor' (laborer) class. Most accidents occur between 10 AM and 2 PM. | (Shekhar & Agarwal, 2021) |

| | | | |
|---|--|---|----------------------|
| | Tableau, for visualization. | | |
| Automatic identification and quantification of the contributing factors in coal mine accidents, overcoming the limitations of human analysis methods | Text mining, association rule extraction, and network theory. Text mining to extract key accident causes, reduce dimensionality, and classify factors within the risk model. A priori algorithm to identify associations between causes, revealing core causes and critical causal pathways. | Fifty-two root causes were identified and categorized. | (Qiu et al., 2021) |
| Analyze complex narrative clinicians' reports to prevent medical errors and enhance patients' safety. | Convolutional and recurrent neural networks, coupled with an attention mechanism. NLP techniques to identify and categorize harm events in patient care narratives. | Improved medical error detection in large datasets, enhanced data analysis and root cause understanding, and better allocation of resources to address safety incidents have led to the prevention of patient's harm. | (Cohan et al., 2017) |
| Explore potential applications of NLP methods in the analysis of critical incident reports in healthcare to enhance patient safety and quality of care. | Faceted search for intuitive report retrieval and text mining to uncover relationships between reported events. Mapping of incident reports to the International Classification of Patient Safety (ICPS) to facilitate faceted searching and semantic annotation. | Requirements for automated processing include entity recognition, information categorization, event detection, and temporal analysis. | (Denecke, 2016) |

4. Conclusions

This review highlights the transformative potential of AI, particularly Large Language Models (LLMs) and NLP, in enhancing occupational risk prevention across various industrial sectors. The results show that AI and NLP methods are increasingly being integrated into occupational safety and health (OSH) practices, providing innovative solutions to longstanding challenges in workplace safety management. These technologies have demonstrated significant benefits in terms of automating the identification and classification of safety risks and risk factors, improving real-time hazard detection, and predicting occupational accidents, thereby enabling more proactive and effective risk management strategies.

In the aviation sector, AI and NLP have been successfully applied to analyze large volumes of safety reports, enhancing the detection of risk factors and support decision-making processes. These methods have enabled the identification of complex causal relationships and trends in aviation safety data that were previously difficult to identify using traditional approaches. Similarly, in the construction industry, AI-driven models have been employed to automate the extraction and classification of safety risks from unstructured textual data, such as accident reports, facilitating timely intervention and mitigation potential hazards. The chemical sector has benefited from the integration of AI and NLP to refine hazard analysis, enabling accurate prediction and prevention of accidents through the use of machine learning models and text-mining techniques.

However, several challenges and limitations remain that need to be addressed in order to fully realize the potential of these technologies in occupational risk prevention. Issues such as data quality, availability, and standardization, particularly in industries where safety data is unstructured or multilingual, remain significant barriers. Additionally, the transparency and interpretability of AI models, especially those involving complex machine learning algorithms, are critical concerns that require further attention to ensure trust and acceptance among OSH practitioners. Furthermore, the scarcity of domain-specific datasets for training AI models poses a challenge to achieving the desired levels of accuracy and reliability in real-world applications.

Despite these challenges, the review highlights the promising role of AI and NLP in advancing occupational risk prevention practices. The ability of these technologies to process large and complex datasets, identify patterns, and predict potential risks before they occur offers a significant advantage in promoting safer working environments. Future research should focus on developing more sophisticated AI models that can handle different types of data and address the limitations of current methods, such as enhancing model transparency and ensuring compatibility across different languages and contexts.

In conclusion, AI, LLMs, and NLP offer significant opportunities to revolutionize occupational risk prevention. By overcoming existing challenges and optimizing these technologies for specific industrial applications, organizations can achieve more efficient, accurate, and proactive approaches to managing workplace hazards, ultimately contributing to improved worker safety and health outcomes worldwide. Continued research and development in this area is essential to fully realise the benefits of AI-driven methodologies in the evolving occupational risk management landscape.

Author Contributions: N. Orviz-Martínez, E. Pérez-Santín: Writing - review & editing, Writing - original draft, Validation, Investigation, Formal analysis. J.I. López Sánchez: Writing - review & editing, Methodology, Validation, Supervision, Conceptualization.

Funding: This research received funding from the research project «Estudio de los nuevos escenarios y riesgos laborales asociados al proceso de transición de la economía lineal a la economía circular (ERL-ELEC)». Funding Entity: UNIR (2022-2024).

Data Availability Statement: No new data were created or analyzed in this study.

Conflicts of 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

1. Ahadh, A., Binish, G. V., & Srinivasan, R. (2021). Text mining of accident reports using semi-supervised keyword extraction and topic modeling. *Process Safety and Environmental Protection*, 155, 455-465. <https://doi.org/10.1016/j.psep.2021.09.022>
2. Badri, A., Boudreau-Trudel, B., & Souissi, A. S. (2018). Occupational health and safety in the industry 4.0 era: A cause for major concern? *Safety Science*, 109, 403-411. <https://doi.org/10.1016/j.ssci.2018.06.012>
3. Baker, H., Hallowell, M. R., & Tixier, A. J.-P. (2020). Automatically learning construction injury precursors from text. *Automation in Construction*, 118, 103145. <https://doi.org/10.1016/j.autcon.2020.103145>
4. Ben Abbes, S., Temal, L., Arbod, G., Lanteri-Minet, P.-L., & Calvez, P. (2022). Combining Ontology and Natural Language Processing Methods for Prevention of Falls from Height. En B. Villazón-Terrazas, F. Ortiz-Rodriguez, S. Tiwari, M.-A. Sicilia, & D. Martín-Moncunill (Eds.), *Knowledge Graphs and Semantic Web* (pp. 47-61). Springer International Publishing. https://doi.org/10.1007/978-3-031-21422-6_4
5. Cheng, M.-Y., Kusoemo, D., & Gosno, R. A. (2020). Text mining-based construction site accident classification using hybrid supervised machine learning. *Automation in Construction*, 118, 103265. <https://doi.org/10.1016/j.autcon.2020.103265>
6. Cohan, A., Fong, A., Ratwani, R. M., & Goharian, N. (2017). Identifying Harm Events in Clinical Care through Medical Narratives. *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, 52-59. <https://doi.org/10.1145/3107411.3107485>
7. Dalal, S., & Bassu, D. (2020). Deep analytics for workplace risk and disaster management. *IBM Journal of Research and Development*, 64(1/2), 14:1-14:9. IBM Journal of Research and Development. <https://doi.org/10.1147/JRD.2019.2945693>
8. de Vries, V. (2020). Classification of Aviation Safety Reports using Machine Learning. *2020 International Conference on Artificial Intelligence and Data Analytics for Air Transportation (AIDA-AT)*, 1-6. <https://doi.org/10.1109/AIDA-AT48540.2020.9049187>
9. Denecke, K. (2016). Automatic Analysis of Critical Incident Reports: Requirements and Use Cases. *Studies in Health Technology and Informatics*, 223, 85-92.
10. Dong, T., Yang, Q., Ebadi, N., Luo, X. R., & Rad, P. (2021). Identifying Incident Causal Factors to Improve Aviation Transportation Safety: Proposing a Deep Learning Approach. *Journal of Advanced Transportation*, 2021, e5540046. <https://doi.org/10.1155/2021/5540046>
11. Ebrahimi, H., Sattari, F., Lefsrud, L., & Macciotta, R. (2023). A machine learning and data analytics approach for predicting evacuation and identifying contributing factors during hazardous materials incidents on railways. *Safety Science*, 164, 106180. <https://doi.org/10.1016/j.ssci.2023.106180>
12. Fan, H., & Li, H. (2013). Retrieving similar cases for alternative dispute resolution in construction accidents using text mining techniques. *Automation in Construction*, 34, 85-91. <https://doi.org/10.1016/j.autcon.2012.10.014>
13. Fang, W., Luo, H., Xu, S., Love, P. E. D., Lu, Z., & Ye, C. (2020). Automated text classification of near-misses from safety reports: An improved deep learning approach. *Advanced Engineering Informatics*, 44, 101060. <https://doi.org/10.1016/j.aei.2020.101060>
14. Figueres-Esteban, M., Hughes, P., & van Gulijk, C. (2016). Visual analytics for text-based railway incident reports. *Safety Science*, 89, 72-76. <https://doi.org/10.1016/j.ssci.2016.05.009>
15. Fonseca, L. M. (2018). Industry 4.0 and the digital society: Concepts, dimensions and envisioned benefits. *Proceedings of the International Conference on Business Excellence*, 12(1), 386-397.
16. Gadekar, H., & Bugalia, N. (2023). Automatic classification of construction safety reports using semi-supervised YAKE-Guided LDA approach. *Advanced Engineering Informatics*, 56, 101929. <https://doi.org/10.1016/j.aei.2023.101929>
17. Ganguli, R., Miller, P., & Pothina, R. (2021). Effectiveness of Natural Language Processing Based Machine Learning in Analyzing Incident Narratives at a Mine. *Minerals*, 11(7), 776. <https://doi.org/10.3390/min11070776>
18. Garvin, T., & Kimbleton, S. (2021). Artificial intelligence as ally in hazard analysis. *Process Safety Progress*, 40(3), 43-49. <https://doi.org/10.1002/prs.12243>

19. Gomes-Miranda, L., & Gonçalves, F. (2024). The impact of Industry 4.0 on occupational health and safety: A systematic literature review. *Journal of Safety Research*, 90, 254-271. <https://doi.org/10.1016/j.jsr.2024.04.009>
20. Heidarysafa, M., Kowsari, K., Barnes, L., & Brown, D. (2018). Analysis of Railway Accidents' Narratives Using Deep Learning. 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), 1446-1453. <https://doi.org/10.1109/ICMLA.2018.00235>
21. Hua, L., Zheng, W., & Gao, S. (2019). Extraction and Analysis of Risk Factors from Chinese Railway Accident Reports. 2019 IEEE Intelligent Transportation Systems Conference (ITSC), 869-874. <https://doi.org/10.1109/ITSC.2019.8917094>
22. Hughes, P., Robinson, R., Figueres-Esteban, M., & van Gulijk, C. (2019). Extracting safety information from multi-lingual accident reports using an ontology-based approach. *Safety Science*, 118, 288-297. <https://doi.org/10.1016/j.ssci.2019.05.029>
23. Hughes, P., Shipp, D., Figueres-Esteban, M., & van Gulijk, C. (2018). From free-text to structured safety management: Introduction of a semi-automated classification method of railway hazard reports to elements on a bow-tie diagram. *Safety Science*, 110, 11-19. <https://doi.org/10.1016/j.ssci.2018.03.011>
24. ILO. (2023). A call for safer and healthier working environments – International Labour Organization. <https://researchrepository.ilo.org/esploro/outputs/report/995343988202676>
25. Jiao, Y., Dong, J., Han, J., & Sun, H. (2022). Classification and Causes Identification of Chinese Civil Aviation Incident Reports. *Applied Sciences*, 12(21), Article 21. <https://doi.org/10.3390/app122110765>
26. Jidkov, V., Abielmona, R., Teske, A., & Petriu, E. (2020). Enabling Maritime Risk Assessment Using Natural Language Processing-based Deep Learning Techniques. 2020 IEEE Symposium Series on Computational Intelligence (SSCI), 2469-2476. <https://doi.org/10.1109/SSCI47803.2020.9308441>
27. Jing, S., Liu, X., Gong, X., Tang, Y., Xiong, G., Liu, S., Xiang, Shuguang., & Bi, Rongshan. (2022). Correlation analysis and text classification of chemical accident cases based on word embedding. *Process Safety and Environmental Protection*, 158, 698-710. <https://doi.org/10.1016/j.psep.2021.12.038>
28. Kabir, S., Taleb-Berrouane, M., & Papadopoulos, Y. (2023). Dynamic reliability assessment of flare systems by combining fault tree analysis and Bayesian networks. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 45(2), 4305-4322. <https://doi.org/10.1080/15567036.2019.1670287>
29. Kamil, M. Z., Khan, F., Halim, S. Z., Amyotte, P., & Ahmed, S. (2023). A methodical approach for knowledge-based fire and explosion accident likelihood analysis. *Process Safety and Environmental Protection*, 170, 339-355. <https://doi.org/10.1016/j.psep.2022.11.074>
30. Khairuddin, M. Z. F., Hasikin, K., Abd Razak, N. A., Lai, K. W., Osman, M. Z., Aslan, M. F., Sabanci, K., Azizan, M. M., Satapathy, S. C., & Wu, X. (2022). Predicting occupational injury causal factors using text-based analytics: A systematic review. *Frontiers in Public Health*, 10, 984099. <https://doi.org/10.3389/fpubh.2022.984099>
31. Kierszbaum, S., Klein, T., & Lapasset, L. (2022). ASRS-CMFS vs. RoBERTa: Comparing Two Pre-Trained Language Models to Predict Anomalies in Aviation Occurrence Reports with a Low Volume of In-Domain Data Available. *Aerospace*, 9(10), Article 10. <https://doi.org/10.3390/aerospace9100591>
32. Kierszbaum, S., & Lapasset, L. (2020). Applying Distilled BERT for Question Answering on ASRS Reports. 2020 New Trends in Civil Aviation (NTCA), 33-38. <https://doi.org/10.23919/NTCA50409.2020.9291241>
33. Kim, Y., Park, J., & Park, M. (2016). Creating a Culture of Prevention in Occupational Safety and Health Practice. *Safety and Health at Work*, 7(2), 89-96. <https://doi.org/10.1016/j.shaw.2016.02.002>
34. Kuhn, K. D. (2018). Using structural topic modeling to identify latent topics and trends in aviation incident reports. *Transportation Research Part C: Emerging Technologies*, 87, 105-122. <https://doi.org/10.1016/j.trc.2017.12.018>
35. Kumari, P., Wang, Q., Khan, F., & Kwon, J. S.-I. (2022). A unified causation prediction model for aboveground onshore oil and refined product pipeline incidents using artificial neural network. *Chemical Engineering Research and Design*, 187, 529-540. <https://doi.org/10.1016/j.cherd.2022.09.022>
36. Leesakul, N., Oostveen, A.-M., Eimontaite, I., Wilson, M. L., & Hyde, R. (2022). Workplace 4.0: Exploring the Implications of Technology Adoption in Digital Manufacturing on a Sustainable Workforce. *Sustainability*, 14(6), Article 6. <https://doi.org/10.3390/su14063311>

37. Liu, C., & Yang, S. (2022). Using text mining to establish knowledge graph from accident/incident reports in risk assessment. *Expert Systems with Applications*, 207, 117991. <https://doi.org/10.1016/j.eswa.2022.117991>
38. Luo, X., Feng, X., Ji, X., Dang, Y., Zhou, L., Bi, K., & Dai, Y. (2023). Extraction and analysis of risk factors from Chinese chemical accident reports. *Chinese Journal of Chemical Engineering*, 61, 68-81. <https://doi.org/10.1016/j.cjche.2023.02.026>
39. Luo, Y., & Shi, H. (2019). Using lda2vec Topic Modeling to Identify Latent Topics in Aviation Safety Reports. 2019 IEEE/ACIS 18th International Conference on Computer and Information Science (ICIS), 518-523. <https://doi.org/10.1109/ICIS46139.2019.8940271>
40. Macêdo, J. B., das Chagas Moura, M., Aichele, D., & Lins, I. D. (2022). Identification of risk features using text mining and BERT-based models: Application to an oil refinery. *Process Safety and Environmental Protection*, 158, 382-399. <https://doi.org/10.1016/j.psep.2021.12.025>
41. Madeira, T., Melício, R., Valério, D., & Santos, L. (2021). Machine Learning and Natural Language Processing for Prediction of Human Factors in Aviation Incident Reports. *Aerospace*, 8(2), Article 2. <https://doi.org/10.3390/aerospace8020047>
42. Marev, K., & Georgiev, K. (2019). Automated Aviation Occurrences Categorization. 2019 International Conference on Military Technologies (ICMT), 1-5. <https://doi.org/10.1109/MILTECHS.2019.8870055>
43. Milea, A., & Cioca, L.-I. (2024). Work evolution and safety and health at work in Industry 4.0 / Industry 5.0. *MATEC Web of Conferences*, 389, 00074. <https://doi.org/10.1051/mateconf/202438900074>
44. Miraz, M. H., Hasan, M. T., Sumi, F. R., Sarkar, S., & Hossain, M. A. (2022). Industry 5.0: The Integration of Modern Technologies. In *Machine Vision for Industry 4.0*. CRC Press.
45. Miyamoto, A., Bendarkar, M. V., & Mavris, D. N. (2022). Natural Language Processing of Aviation Safety Reports to Identify Inefficient Operational Patterns. *Aerospace*, 9(8), Article 8. <https://doi.org/10.3390/aerospace9080450>
46. Mollaei, N., Fujao, C., Rodrigues, J., Cepeda, C., & Gamboa, H. (2023). Occupational health knowledge discovery based on association rules applied to workers' body parts protection: A case study in the automotive industry. *Computer Methods in Biomechanics and Biomedical Engineering*, 26(15), 1875-1888. <https://doi.org/10.1080/10255842.2022.2152678>
47. Perboli, G., Gajetti, M., Fedorov, S., & Giudice, S. L. (2021). Natural Language Processing for the identification of Human factors in aviation accidents causes: An application to the SHEL methodology. *Expert Systems with Applications*, 186, 115694. <https://doi.org/10.1016/j.eswa.2021.115694>
48. Pishgar, M., Issa, S. F., Sietsema, M., Pratap, P., & Darabi, H. (2021). REDECA: A Novel Framework to Review Artificial Intelligence and Its Applications in Occupational Safety and Health. *International Journal of Environmental Research and Public Health*, 18(13), Article 13. <https://doi.org/10.3390/ijerph18136705>
49. Posse, C., Matzke, B., Anderson, C., Brothers, A., Matzke, M., & Ferryman, T. (2005). Extracting information from narratives: An application to aviation safety reports. 2005 IEEE Aerospace Conference, 3678-3690. <https://doi.org/10.1109/AERO.2005.1559673>
50. Qiu, Z., Liu, Q., Li, X., Zhang, J., & Zhang, Y. (2021). Construction and analysis of a coal mine accident causation network based on text mining. *Process Safety and Environmental Protection*, 153, 320-328. <https://doi.org/10.1016/j.psep.2021.07.032>
51. Ricketts, J., Barry, D., Guo, W., & Pelham, J. (2023). A Scoping Literature Review of Natural Language Processing Application to Safety Occurrence Reports. *Safety*, 9(2), Article 2. <https://doi.org/10.3390/safety9020022>
52. Ricketts, J., Pelham, J., Barry, D., & Guo, W. (2022). An NLP framework for extracting causes, consequences, and hazards from occurrence reports to validate a HAZOP study. 2022 IEEE/AIAA 41st Digital Avionics Systems Conference (DASC), 1-8. <https://doi.org/10.1109/DASC55683.2022.9925822>
53. Roberts, B. (2015). The Third Industrial Revolution: Implications for Planning Cities and Regions. *Working Paper Urban Frontiers*, 1.
54. Robinson, S. D. (2016). Visual representation of safety narratives. *Safety Science*, 88, 123-128. <https://doi.org/10.1016/j.ssci.2016.05.005>

55. Rose, R. L., Puranik, T. G., & Mavris, D. N. (2020). Natural Language Processing Based Method for Clustering and Analysis of Aviation Safety Narratives. *Aerospace*, 7(10), Article 10. <https://doi.org/10.3390/aerospace7100143>
56. Rybak, N., & Hassall, M. (2021). Deep Learning Unsupervised Text-Based Detection of Anomalies in U.S. Chemical Safety and Hazard Investigation Board Reports. 2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), 1-7. <https://doi.org/10.1109/ICECCME52200.2021.9590834>
57. Sarkar, S., & Maiti, J. (2020). Machine learning in occupational accident analysis: A review using science mapping approach with citation network analysis. *Safety Science*, 131, 104900. <https://doi.org/10.1016/j.ssci.2020.104900>
58. Sharma, A., & Singh, B. J. (2020). Evolution of Industrial Revolutions: A Review. *International Journal of Innovative Technology and Exploring Engineering*, 9(11), 66-73. <https://doi.org/10.35940/ijitee.I7144.0991120>
59. Shekhar, H., & Agarwal, S. (2021). Automated Analysis through Natural Language Processing of DGMS Fatality Reports on Indian Non-Coal Mines. 2021 5th International Conference on Information Systems and Computer Networks (ISCON), 1-6. <https://doi.org/10.1109/ISCON52037.2021.9702346>
60. Shen, Q., Wu, S., Deng, Y., Deng, H., & Cheng, J. C. P. (2022). BIM-Based Dynamic Construction Safety Rule Checking Using Ontology and Natural Language Processing. *Buildings*, 12(5), Article 5. <https://doi.org/10.3390/buildings12050564>
61. Song, B., & Suh, Y. (2019). Narrative texts-based anomaly detection using accident report documents: The case of chemical process safety. *Journal of Loss Prevention in the Process Industries*, 57, 47-54. <https://doi.org/10.1016/j.jlp.2018.08.010>
62. Tanguy, L., Tulechki, N., Urieli, A., Hermann, E., & Raynal, C. (2016). Natural language processing for aviation safety reports: From classification to interactive analysis. *Computers in Industry*, 78, 80-95. <https://doi.org/10.1016/j.compind.2015.09.005>
63. Thompson, P., Yates, T., Inan, E., & Ananiadou, S. (2020). Semantic Annotation for Improved Safety in Construction Work. En N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the Twelfth Language Resources and Evaluation Conference* (pp. 1990-1999). European Language Resources Association. <https://aclanthology.org/2020.lrec-1.245>
64. Tian, D., Li, M., Shen, Y., & Han, S. (2023). Intelligent mining of safety hazard information from construction documents using semantic similarity and information entropy. *Engineering Applications of Artificial Intelligence*, 119, 105742. <https://doi.org/10.1016/j.engappai.2022.105742>
65. Van Gulijk, C. (2021). El desarrollo de una evaluación de riesgos dinámica y sus implicaciones para la salud y seguridad en el trabajo. *Agencia Europea para la Seguridad y la Salud en el Trabajo (EU-OSHA)*.
66. Valcamonico, D., Baraldi, P., Amigoni, F., & Zio, E. (2022). A framework based on Natural Language Processing and Machine Learning for the classification of the severity of road accidents from reports. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 1748006X221140196. <https://doi.org/10.1177/1748006X221140196>
67. Vinitha, K., Ambrose Prabhu, R., Bhaskar, R., & Hariharan, R. (2020). Review on industrial mathematics and materials at Industry 1.0 to Industry 4.0. *Materials Today: Proceedings*, 33, 3956-3960. <https://doi.org/10.1016/j.matpr.2020.06.331>
68. Wang, B., & Zhao, J. (2022). Automatic frequency estimation of contributory factors for confined space accidents. *Process Safety and Environmental Protection*, 157, 193-207. <https://doi.org/10.1016/j.psep.2021.11.004>
69. Wang, G., Liu, M., Cao, D., & Tan, D. (2021). Identifying high-frequency-low-severity construction safety risks: An empirical study based on official supervision reports in Shanghai. *Engineering, Construction and Architectural Management*, 29(2), 940-960. <https://doi.org/10.1108/ECAM-07-2020-0581>
70. Wang, Y., Chen, H., Liu, B., Yang, M., & Long, Q. (2020). A Systematic Review on the Research Progress and Evolving Trends of Occupational Health and Safety Management: A Bibliometric Analysis of Mapping Knowledge Domains. *Frontiers in Public Health*, 8. <https://doi.org/10.3389/fpubh.2020.00081>

71. Wang, Z., & Yin, J. (2020). Risk assessment of inland waterborne transportation using data mining. *Maritime Policy & Management*, 47(5), 633-648. <https://doi.org/10.1080/03088839.2020.1738582>
72. Westhoven, M. (2022). Requirements for AI Support in Occupational Safety Risk Analysis. *Proceedings of Mensch und Computer 2022*, 561-565. <https://doi.org/10.1145/3543758.3547576>
73. Westhoven, M., & Jadid, A. (2023). Using Natural Language Processing to Generate Risk Assessment Checklists From Workplace Descriptions. *Proceeding of the 33rd European Safety and Reliability Conference*, 2636-2637. https://doi.org/10.3850/978-981-18-8071-1_P452-cd
74. Wu, H., Zhong, B., Medjdoub, B., Xing, X., & Jiao, L. (2020). An Ontological Metro Accident Case Retrieval Using CBR and NLP. *Applied Sciences*, 10(15), Article 15. <https://doi.org/10.3390/app10155298>
75. Xu, H., Liu, Y., Shu, C.-M., Bai, M., Motalifu, M., He, Z., Wu, S., Zhou, P., & Li, B. (2022). Cause analysis of hot work accidents based on text mining and deep learning. *Journal of Loss Prevention in the Process Industries*, 76, 104747. <https://doi.org/10.1016/j.jlp.2022.104747>
76. Xu, N., Ma, L., Liu, Q., Wang, L., & Deng, Y. (2021). An improved text mining approach to extract safety risk factors from construction accident reports. *Safety Science*, 138, 105216. <https://doi.org/10.1016/j.ssci.2021.105216>
77. Yang, C., & Huang, C. (2023). Natural Language Processing (NLP) in Aviation Safety: Systematic Review of Research and Outlook into the Future. *Aerospace*, 10(7), Article 7. <https://doi.org/10.3390/aerospace10070600>
78. Yimyam, W., & Ketcham, M. (2022). Occupational Disease Risk Assessment System Using Artificial Intelligence System and Chatbot. *2022 International Conference on Cybernetics and Innovations (ICCI)*, 1-5. <https://doi.org/10.1109/ICCI54995.2022.9744143>
79. Zhang, C., & Yang, J. (2020). Second Industrial Revolution. En C. Zhang & J. Yang (Eds.), *A History of Mechanical Engineering* (pp. 137-195). Springer. https://doi.org/10.1007/978-981-15-0833-2_5
80. Zhang, F., Fleyeh, H., Wang, X., & Lu, M. (2019). Construction site accident analysis using text mining and natural language processing techniques. *Automation in Construction*, 99, 238-248. <https://doi.org/10.1016/j.autcon.2018.12.016>
81. Zhang, X., Srinivasan, P., & Mahadevan, S. (2021). Sequential deep learning from NTSB reports for aviation safety prognosis. *Safety Science*, 142, 105390. <https://doi.org/10.1016/j.ssci.2021.105390>
82. Zhao, Y., Diao, X., Huang, J., & Smidts, C. (2019). Automated Identification of Causal Relationships in Nuclear Power Plant Event Reports. *Nuclear Technology*, 205(8), 1021-1034. <https://doi.org/10.1080/00295450.2019.1580967>
83. Zhao, Y., Diao, X., & Smidts, C. (2018, septiembre 1). Preliminary Study of Automated Analysis of Nuclear Power Plant Event Reports Based on Natural Language Processing Techniques.

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