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-P5041-cd

Improving the Predictability of Hazardous Scenarios by Natural Language Processing. The case of accidents during lifting operations on ships and offshore platforms.

Ibsen Chivata Cardenas

University of Southern Denmark. E-mail: ibsenccardenas@sdu.dk

Igor Kozine

University of Southern Denmark. E-mail: igko@sdu.dk

The completeness and high predictability of hazardous scenarios by hazard identification methods are issues in risk analyses. A way to the improvement is to carry out both an exhaustive - to the extent possible - post-accident and predictive accident analysis. Currently, Natural Language Processing (NLP) allows quick processing of many accident reports. In combination with graphical tools, it is now even possible to automatically output causal diagrammatic models of accidents and visualize them on a multi-scenario accident diagram. A step forward is the application of NLP to support predictive analysis. Predictive accident analysis focuses on identifying deviations from expected or normal conditions, the subsequent events following these deviations, and their interactions leading to an accident. The expected or normal conditions are typically outlined in specifications and procedures. This paper demonstrates how NLP can assist hazard identification and predictive accident analysis during lifting operations on ships and offshore platforms.

Keywords: hazard identification, predictive accident analysis, natural language processing.

1. Introduction

Hazard identification methods are often portrayed as completely developed and flawlessly functional. However, even the most advanced method, HAZOP, is not without flaws. Primarily, due to challenges in maintaining focus during analyses and gaps in knowledge, the completeness of high-quality HAZOP studies ranges between 85% and 95% (Taylor 2018). This implies that 5% to 15% of accident events are missed. Other methods tend to perform even worse. Continued research is still necessary (Taylor 2018).

One way to improve the quality of hazard identification is by maximizing the use of accident reports. Post-accident analysis is a method that utilizes repositories of accident case histories, indexed in detail to enable quick and contextual retrieval of relevant knowledge. It is obvious that the quality of post-accident analysis depends on the thoroughness of the accident records and the reporter's choices regarding what was important to report. However, it is possible to develop a more complete analysis by combining post-accident analysis and predictive analysis in conjunction with the use of expert judgment (Cardenas et al. 2025).

A predictive analysis identifies deviations from expected or normal conditions in the execution of an operation, the subsequent events following these deviations, and their interactions leading to an accident. Specifications and procedures typically outline the expected or normal conditions (Cardenas et al. 2025). Some issues might arise when using the predictive analysis. It could potentially lead to the identification of some unrecorded and potentially non-credible accident events. Although this potential bias can be reduced by involving highly skilled experts, other complementary means to increase hazard identification credibility are desirable. One option is validation by exhausting information from the majority of currently available sources. Some available extensive sources are, for example, the European Maritime Safety Agency (EMSA), which has reported since 2011 more than 37000 events including accidents, incidents, and near-misses. The Health and Safety

Executive (HSE) has also made available more than 10000 events associated with oil and gas activities in the period 1980-2005. Exploiting all such information thoroughly appears to be a challenge. However, recent developments in computer and information science, mathematics, and linguistics that include Natural Language Processing (NLP), promise to be beneficial in tackling this challenge.

NLP focuses on the development of algorithms and models that enable computers to process and generate useful text (Chowdhary 2020, Kang et al. 2020). NLP encompasses a variety of tasks, including, text classification, machine translation, conversion of spoken language into text, extraction of knowledge, and text generation (Khurana et al. 2023).

The use of NLP in accident analysis has experienced significant growth in the last 50 years. More than 850 papers have been published in this period according to the Scopus database. Despite this considerable body of knowledge, the use of predictive accident analysis or similar approaches to predict accidents in conjunction with NLP remains underreported. The use of NLP for accident analysis is often focused on the classification of hazards (e.g., Jia et al. 2024, Kumi et al. 2024, Nurduhan and Kuleyin 2024) and the identification of the most frequent hazards and associated consequences (e.g., Kutela, et al. 2024, Venkatesh et al., 2024, Yang et al. 2021). The analysis of inaccurate, incomplete, or unstructured accident reports is also highly explored (e.g., Gangadhari et al. 2024, Guo et al. 2024, Ramos et al. 2024, Zhang et al. 2024). Few sources consider the improvement of the completeness of hazard identification using NLP, these include Hu et al. (2024) and Li et al. (2022) publications. Unfortunately, the latter authors did not use predictive or equivalent accident analysis. To bridge this gap, this paper's objective is to improve predictive accident analysis using relevant tools from NLP. To meet this objective, useful NLP tools are identified and coupled to develop a customized approach to predictive accident analysis.

The remainder of the paper is as follows. The next section describes the approach by providing details about the predictive analysis and useful NLP tools. Section 3 demonstrates aspects of the proposed approach by analyzing accidents involving injuries to people on ships and offshore platforms. The last section contains the discussion and conclusions of the research.

2. Description and implementation of the proposed approach

2.1 *Predictive accident analysis*

Predictive accident analysis ('predictive analysis' in short) aims to identify deviations from normal or acceptable conditions that are necessary for an accident to occur. Once these deviations are identified, the subsequent events that may ultimately lead to accidents can be determined. Potential deviations from expected conditions are identified by examining existing specifications or procedures. These documents typically break down the activity into specific tasks and the associated expected or acceptable conditions. Accident events could also arise from deviations from specified safety measures. A safety measure is considered as an action, procedure, or artefact designed to lower the occurrence or risk of injury, loss and danger to persons, property, or the environment (adapted from Collins English dictionary). Aspects of the predictive analysis have been used by Callesen et al. (2019) and Kozin and Taylor (2022). Predictive analysis can be summarized as follows:

(i) define the boundaries of the system, process, or activity to be analyzed (e.g., physical, operational, design/construction, managerial, or organizational)
(ii) obtain a description of the system process or activity to be analyzed
(iii) identify the prescribed expected and acceptable events and conditions and organize them sequentially
(iv) from the sequence of the events and conditions, select one for the analysis
(v) establish the potential deviations for the event and condition selected
(vi) identify the events following the deviations
(vii) repeat steps (iv-vi) for every event or condition identified
(viii) determine the causal links among the identified events including their consequences.

The following definitions are used:

- Event: A change in the state of a system, workplace, person, or machine
- Condition: The state of a system, workplace, person, or machine

2.2 *NLP tools*

In Figure 1, a flow diagram shows the steps and NLP tools required for the implementation of the proposed approach. As illustrated in Figure 1, the approach enables the use of various accident databases, as well as requires input from procedures and specifications. The output is a validated and credible set of deviations from events and conditions.

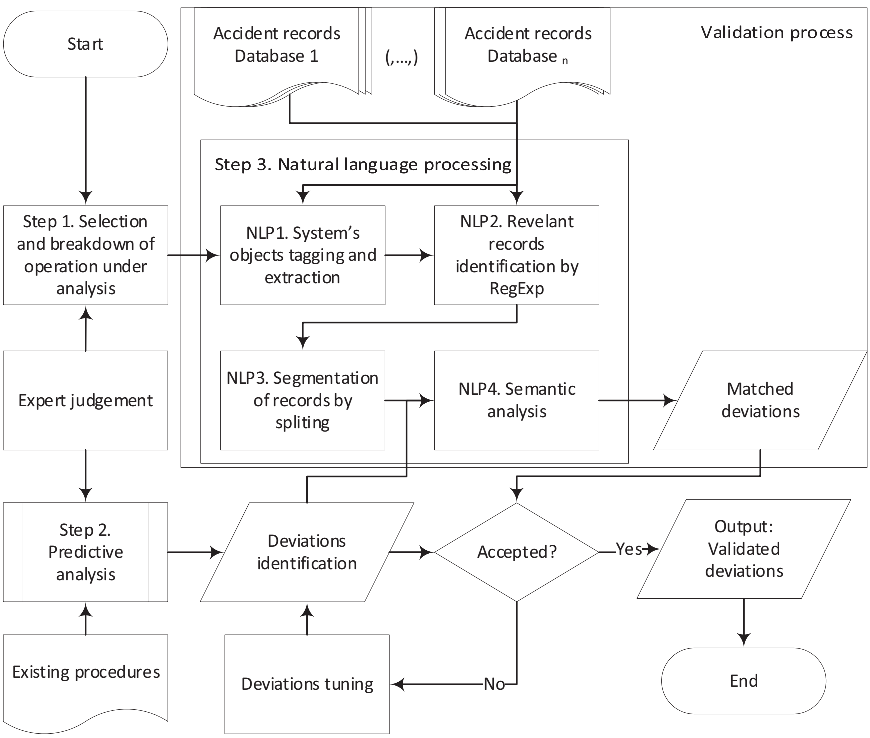

Fig. 1. A flow diagram that displays the steps in the implementation of the proposed approach.

In the implementation of the approach, expert judgment is required to supervise the breakdown into tasks of the system or operation under analysis (step 1 in Figure 1) and the predictive analysis (step 2, section 2.1). Experts also review the outputs (decision box).

The following briefly describes the NLP tools that can be coupled to support predictive analysis. NLP software packages, e.g., SpaCy or Natural Language Toolkit provide these tools. The tools can be customized using programming languages such as Python.

NLP1: Once step 1 is completed, the entities or objects that mutually interact in the execution of a task of interest can be identified. This is supported mainly by the Parts-Of-Speech (POS) tagging tool. It labels each word in a given text with its POS tag, e.g., noun, verb, adjective, determiner, and preposition (Zhang 2024). Note that nouns are of particular interest as they identify the interacting objects for a task under analysis.

NLP2: Based on the objects that have been identified by NLP1, the selection of relevant records from the input databases can take place. This is achieved by using Regular Expressions (RegExp) filters. In general, RegExp(s) detect, extract, replace, and match sequences of characters (Kozen 1977). Different objects linked to a task are queried using RegExp filters. Subsequently, relevant documents from the databases are identified. A typical RegExp used in NLP2 is the following:

$$[\hat{}.]*(\text{object 1}|\text{object 2}|\text{object 3}|.|.|.).*?(\text{object 1}|\text{object 2}|\text{object 3}|.|.|.)[\hat{}.]* \quad (1)$$

Expression (1) identifies documents that contain text fragments where an object queried, e.g., 'object 1' appears in conjunction with any other specified object of interest, e.g., 'object 2' or other objects enclosed within vertical bars '|'. Expression (1) exhausts all combinations of the set of objects under analysis, e.g., the set (object 1,object 2,object 3,…). Expression (1) also highlights the text between the matched objects, which likely provides the details of the interaction among the queried objects.

NLP3: This tool splits large documents into segments or fragments of text so they can be processed reliably by the subsequent NLP tool in the approach. NLP3 process is called segmentation or splitting. Detailed algorithms have been developed for this function, e.g., Gao et al. (2024).

NLP4: At the core of NLP4 is semantic analysis. It identifies text fragments that are associated with the queried specific objects, events, or deviations of interest. Semantic analysis establishes relationships between words, phrases, and sentences in a given context. The analysis resolves ambiguity that is typically found in unstructured texts in which words and sentences could have multiple meanings. Using pre-trained sentence models such as SBERT, words, sentences, or text fragments semantically equivalent to the queried specific objects, events, or deviations of interest can be quickly identified. A pre-trained sentence model is a neural network model that has been trained to learn patterns and features from extensive text data (Reimers and Gurevych 2019). NLP4 receives input from both Step 2 and NLP3 in the form of the queried specific objects, events, and deviations and renders as output equivalent deviations that have been recorded in the input databases. Subsequently, the experts will determine whether to accept the deviations that have been matched by NLP (decision box) or if the input deviation needs to be better specified before being reprocessed by NLP.

Other NLP tools are necessary for pre-processing the raw documents including tokenisation, stemming, and lemmatisation (Antons et al. 2020). These tools are not described in this paper as they are routinely used in any NLP.

The following section shows more details of the approach by an example using data from repositories that record accidents including injuries to people on ships and platforms.

3. Example

In this section, the proposed approach is further illustrated by analyzing lifting operations that are carried out on platforms or ships. Based on the authors' experience using the repositories from EMSA, HSE, and the Global Offshore Wind Health and Safety Organisation (G+), lifting operations are prone to accidents.

In the following, details about critical inputs, steps, and tools of the approach are given.

3.1. *Input*

Although any number of repository sources can be deployed when using the developed approach, for this example, HSE and G+ repositories were used. Both sources totalize 23085 records including accidents, incidents, and near-misses.

3.2. *Task identification*

Lifting procedures have been regulated by, e.g., the Department for Transport UK (2006). Table 1 shows a generic procedure that has been derived from the regulation and is considered enough for the proposed demonstration.

3.3. *Deviations identification*

Using steps (i) to (v) of the predictive analysis (subsection 2.1.), deviations from procedures were identified by this paper's authors. Some deviations are displayed in Table 2.

3.4. *Natural language processing*

For this process, Orange data mining software version 3.37 (Demsar et al. 2013) was used. The NLP tools described in this paper are available from this software. For the NLP4 tool, the software uses the pre-trained sentence model - SBERT (Reimers and Gurevych 2019). Although the recommendation is to pre-train a customized model for the analysis of accidents to obtain more accurate matches, SBERT model is considered sufficient for the demonstration.

Table 1. Derived generic procedure for lifting operations on a ship or platform.

#	Task
1	Ensure lifting operation is properly planned
2	Ensure appropriate supervision by a competent person
3	Ensure adequate and effective coordination between operators when two or more pieces of work equipment are used simultaneously to lift a single load
4	If lifting persons is required, use lifting equipment designed for lifting persons
5	Select appropriate accessories for lifting, considering the load, gripping points, loose gear, atmospheric conditions, and slinging configuration
6	Test lifting equipment, accessories, and loose gear after manufacture, installation, repair, or modification
7	Clearly mark lifting equipment and accessories with their safe working loads
8	Store accessories in conditions that prevent damage or degradation
9	Organize work so that when a worker is attaching or detaching a load by hand, the operation can be carried out safely, ensuring the worker retains direct or indirect control of the work equipment
10	Ensure lifting equipment can maintain its hold on the load in the event of a complete or partial power failure
11	Verify that lifting equipment will remain stable during use
12	Take measures to prevent the load from striking anything or any person
13	Ensure all parts of the load and attachments will hold together during the operation
14	Include measures to avoid collisions between the loads or the equipment
15	Ensure that permanently installed lifting equipment is positioned to minimize risks such as striking workers, load drifting, or unintentional release
16	As far as reasonably practicable, avoid carrying or suspending loads over areas occupied by workers
17	If the operator cannot observe the full path of the load, ensure a responsible person has appropriate means of communication to guide the operation
18	Conduct thorough examinations and inspections at regular intervals and after exceptional circumstances
19	Halt lifting operations if meteorological conditions deteriorate to the point that they could affect the safe use of the lifting equipment or expose persons to danger
20	Do not resume operations unless the procedures and safety measures are applied

```
[^.]*?(lift|lower|pull|remove|load|hook|winch).*?( finger| thumb| neck| bod| arm| foot| feet| head| ankle| face| wrist| hand| leg| deck| toe| forearm| back| shoulder| sea)[^.]*
```
(2)

As stated before, the focus of the example is on injuries during lifting operations. Accordingly, the objects extraction by NLP 1 started with the search of operations such as ‘lifting’ and objects related to injuries such as ‘finger’ or ‘arm’. Initial queries provided matches using NLP2. When analyzing the matched text in the initial searches, new objects and associated operations were identified. After a number of trials, eventually, the deployed RegExp filter was (Expression 2):

Expression (2) also matched the text between the specified objects in (2) and revealed more objects and interactions that provided increased details about the accident events. Note that Expression (2) has been used in Orange software. Python-based scripts in other platforms might require a slightly different but equivalent syntax.

Using Expression (2), NLP2 identified 3684 out of the 23085 records available from the input databases. Subsequently, NLP3 is enabled and processed the 3684 records to make them ready to be input to NLP4.

Once the processing and identification of objects and relevant documents were completed by the tools NLP 1 to 3 and the deviations identified, both the processed text of the relevant documents alongside the deviations in Table 2 were fed to the NLP 4 tool. NLP4 processed the information and rendered a list of matches as candidate deviations (see Figure 2).

Table 2. Some deviations from the procedure for lifting operations on a ship or platform.

Task #	Abbreviated description of deviations and NLP accepted matches (in parenthesis)
1	Operation unplanned (2)
2	Lack of supervision by a competent person (1)
3	Inadequate and ineffective coordination (4)
4	Inappropriate equipment for lifting persons (2)
5	Inappropriate equipment load, gripping points, loose gear, atmospheric conditions, and slinging configuration (29)
6	Untested lifting equipment, devices, accessories, and loose gear (1)
7	Lifting equipment and accessories not marked (4)
8	Damage or degradation of equipment, devices, or accessories (28)
9	A worker is attaching or detaching a load by hand unsafely, worker does not have direct or indirect control equipment (22)
10	Lifting equipment cannot maintain its hold on the load (11)
11	Lack of verification of equipment stability (7)
12	Load striking anything or any person (479)
13	Load and attachments do not hold together (442)
14	Collisions between the loads or the equipment (10)
15	Incorrect equipment installation (9)
16	Carrying or suspending loads over areas occupied by workers (7)
17	Operator cannot observe the full path of the load (3)
18	No regular or incidental examination or inspection (3)
19	Meteorological conditions deteriorate and lifting operations not halted under unsafe conditions (2)
20	Operations resumed without the procedures and safety measures applied (2)

3.5. *Output*

This paper's authors went through the list of deviations rendered by the processing machine and selected consistent deviations to the ones queried. This is illustrated in Figure 2. The selection of the deviations is informed by a score, which is computed as the maximum cosine similarity (Reimers and Gurevych 2019) between the matched deviations and the input deviation. Based on experience with NLP4, a threshold for the maximum cosine similarity above 0,60 should be chosen to obtain satisfactory output.

4. Analysis and interpretation of the results

Figure 2 shows the deviations matched by NLP4, when the deviation 12 'Load striking anything or any person' is the input. It produced 479 matches with a score above 0,60 (see Table 2). This means that this deviation has significant support from recorded events. The same can be said about deviation 13. For the rest of the deviations in Table 2, it was found that some deviations have small support, e.g., deviation 1 (2 accepted matches) or deviation 2 (1 match), consequently, these deviations are dubious and need further investigation. Note that most of the recorded- high-support deviations concentrate on the physical or operational aspects of lifting activities.

Low support is also found for deviations from safety measures that provide redundancy or that are designed to counteract unwanted occurring events, for example, deviations 6, 11, 18, 19, and 20. This observation suggests that accidents arising once these measures have been deviated are considerably less frequent than accidents arising from other types of safety measures. This is an expected outcome.

Another possible explanation of the results for the small support deviations lies in the integrity of the input records that rarely report on issues associated with design, managerial, or organizational aspects or distant causes of accidents.

5. Discussion and conclusions

A predictive accident analysis approach has been implemented using Natural Language Processing (NLP) tools and applied to analyze accidents that could involve injuries to people on ships and platforms during lifting operations. As output, validated sets of deviations potentially leading to accidents were provided. This is achieved by matching records from extensive databases.

The focus of the reported approach is on validation of the deviations identified by the predictive analysis using NLP. Although Large Language Models (LLMs) can be considered to 'generate' deviations, this research does not advocate that possibility. Caution should be exercised with the outputs generated by LLMs, as they tend to produce the most probable sequences of text. In the context of accident analysis, this means generating the most likely events, which is not beneficial when aiming to improve the completeness of accident analysis, where low-probability accidents are also crucial. Furthermore, LLMs are known to be highly resource-intensive (e.g., Gendron et al., 2024) and not environmentally friendly. Conversely, in the proposed approach, only the necessary NLP tools are coupled, to more efficiently make use of resources.

Further research can consider the pre-training of small and customised language models learnt from the existing accident databases and using the identified NLP tools. This will enable a more accurate validation of deviations.

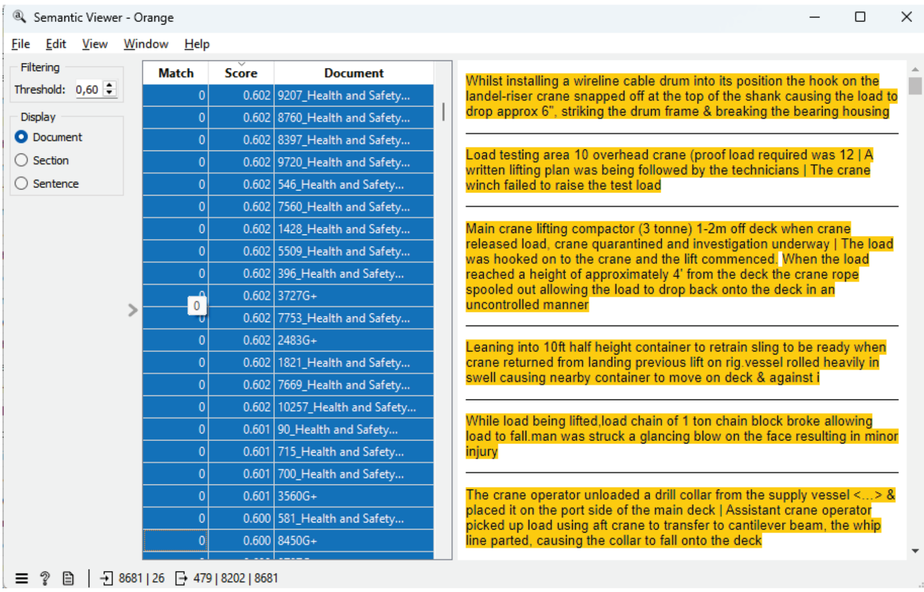

Fig. 2. Output of the NLP4

Acknowledgement

The SDU Maritime Research Platform supported this work with a donation from the A/S Dampskibsselskabet Orient's Foundation. The authors are thankful to the Global Offshore Wind Health and Safety Organisation for allowing access to the G+ database.

References

Antons, D., Grünwald, E., Cichy, P., Salge, T. O. (2020). The application of text mining methods in innovation research: current state, evolution patterns, and development priorities. *R & D Management, 1(1)*, 1-18

Callesen, F. G., Blinkenberg-Thrane, M., Taylor, J. R., & Kozine, I. (2019). Container ships: fire-related risks. *Journal of Marine Engineering & Technology, 20(4)*, 262–277.

Cardenas, I. C., Taylor, R., Kozine, I., and Fenn, A. (2025, January) Accident Analysis Reinforced by Natural Language Processing. Working paper, University of Southern Denmark

Chowdhary, K. R. (2020) *Fundamentals of Artificial Intelligence*. Springer.

Department for Transport UK (2006). The Merchant Shipping and Fishing Vessels (Lifting Operations and Lifting Equipment) Regulations 2006. (Revised version). Department for Transport, Ministry of State UK.

Demsar J, Curk T, Erjavec A, Gorup C, Hocevar T, Milutinovic M, Mozina M, Polajnar M, Toplak M, Staric A, Stajdohar M, Umek L, Zagar L, Zbontar J, Zitnik M, Zupan B (2013) Orange: Data Mining Toolbox in Python, Journal of Machine Learning Research 14(Aug): 2349−2353.

Gangadhari, R.K., Rabiee, M., Khanzode, V., Murthy, S.; Kumar Tarei, P. (2024) From unstructured accident reports to a hybrid decision support system for occupational risk management: The consensus converging approach. *Journal of Safety Research, Volume 89,* 91-104.

Gao, Y., Yihong, G., Yuting, C., and Yongqiang, C. (2024) Application of Large Language Models to Intelligently Analyze Long Construction Contract Texts. *Construction Management and Economics, October*, 1–17.

Gendron, G., Rožanec, J. M., Witbrock, M., Dobbie, G. (2024). Counterfactual Causal Inference in Natural Language with Large Language Models. *arXiv preprint arXiv:2410.06392*.

Guo, A., Zhou, Y., Tian, H., Fang, C., Sun, Y., Sun, W., Gao, X., Luu, A.T., Liu, Y., Chen, Z. (2024) SoVAR: Build Generalizable Scenarios from Accident Reports for Autonomous Driving Testing *Proceedings - 2024 39th ACM/IEEE International Conference on Automated Software Engineering*, *ASE*.

Hu, J., Zhou, W., Zheng, P., Liu, G. (2024) A Novel Approach for the Analysis of Ship Pollution Accidents Using Knowledge Graph. *Sustainability (Switzerland), 16(13),* 5296.

Jia, Q., Fu, G., Xie, X., Xue, Y., Hu, S. (2024) Enhancing accident cause analysis through text classification and accident causation theory: A case study of coal mine gas explosion accidents. *Process Safety and Environmental Protection, Volume 185,* 989-1002.

Kang, Y., Zhao C., Chee-Wee T., Qian H., and Hefu L. (2020) Natural language processing (NLP) in management research: A literature review. *Journal of Management Analytics 7, no. 2*, 139-172.

Khurana, D., Aditya, K., Kiran, K., and Sukhdev, S. (2023) Natural language processing: state of the art, current trends and challenges. *Multimedia tools and applications 82, no. 3*, 3713-3744.

Kozen, D.C. (1977). Kleene Algebra and Regular Expressions. In *Automata and Computability. Undergraduate Texts in Computer Science*, pp. 55–60. Springer.

Kozin, I., & Taylor, J. R. (2022). Operations Error Analysis of the Use of Electronic Chart Display and Information System (ECDIS). *In M. C. Leva, E. Patelli, L. Podofillini, & S. Wilson (Eds.), Proceedings of the 32nd European Safety and Reliability Conference. Research Publishing Services, Singapore*

Kumi, L., Jeong, J., Jeong, J. (2024) Data-driven automatic classification model for construction accident cases using natural language processing with hyperparameter tuning. *Automation in Construction, Volume 164*, 105458.

Kutela, B., Salum, J.H., Seif, S.R., Das S., Kidando E. (2024) Navigating the blame game: Investigating automated vehicle fault in collisions under mixed traffic conditions. *Robotics and Autonomous Systems, Volume 182,* 104831.

Li, S., You, M., Li, D., Liu, J. (2022) Identifying coal mine safety production risk factors by employing text mining and Bayesian network techniques. *Process Safety and Environmental Protection, Volume 162,* 1067-1081.

Nurduhan, M., Kuleyin, B. (2024) Cluster-based Visualization of human element interactions in marine accidents. *Ocean Engineering Volume 298,* 117153.

Ramos, P.M.S., Macedo, J.B., Maior, C.B.S., Moura, M.C., Lins, I.D. (2024) Combining BERT with numerical variables to classify injury leave based on accident description *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, 238(5)*, 945-956.

Reimers, N., Gurevych, I. (2019) Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *In Proceedings of the 2019 Conference on Empirical Methods in NLP.*

Taylor, J. R. (2018) Lecture notes: System Safety and Reliability Engineering, 42171. Hazard Identification. Department of Management Engineering, Technical University of Denmark.

Venkatesh, R. K., Siddharth R., Yuvaraj S., Ramesh Kumar K.A. (2024) An Artificial Intelligence based automated case-based reasoning (CBR) system for severity investigation and root-cause analysis of road accidents – Comparative analysis with the predictions of ChatGPT. *Journal of Engineering Research (Kuwait), Volume 12, Issue 4,* 895-903.

Yang, Y., Kai Z., Houbing S., and Dahai L. (2021) Machine Learning-enabled Adaptive Air Traffic Recommendation System for Disaster Evacuation. *In 40th Digital Avionics Systems Conference, DASC 2021 - Proceedings. IEEE.*

Zhang, X. (2024). Disentangling syntactics, semantics, and pragmatics in natural language processing. Doctoral thesis, Nanyang Tech. University.

Zhang, X., Tian, D., Ren, Q., Li, M., Shen, Y., Han, S. (2024) A hybrid deep semantic mining method considering fuzzy expressions for the automatic recognition of construction safety hazard information. *Advanced Engineering Informatics, Volume 61,* 102507.