



Causation analysis of crane-related accident reports by utilizing ChatGPT and complex networks

Yifan Wang¹, Junyu Chen², Bo Xiao^{1,*}, Shane T. Mueller³, Jingjing Guo⁴

¹Department of Civil, Environmental, and Geospatial Engineering, Michigan Technological University, Houghton, MI 49931, USA.

²Department of Building and Real Estate, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China.

³Department of Cognitive and Learning Science, Michigan Technological University, Houghton, MI 49931, USA.

⁴College of Civil Engineering, Hunan University, Changsha, 410082, Hunan, China.

***Correspondence to:** Bo Xiao, Department of Civil, Environmental, and Geospatial Engineering, Michigan Technological University, Houghton, MI 49931, USA. E-mail: boxiao@mtu.edu

Received: April 11, 2025 **Accepted:** May 19, 2025 **Published:** May 22, 2025

Cite this article: Wang Y, Chen J, Xiao B, Mueller ST, Guo J. Causation analysis of crane-related accident reports by utilizing ChatGPT and complex networks. *J Build Des Environ*. 2025;3:202535. <https://doi.org/10.70401/jbde.2025.0009>

Abstract

This study integrates ChatGPT and complex network (CN) techniques into an accident analysis framework designed to reduce manual effort in accident causation analysis. The proposed framework supports construction stakeholders in extracting causal factors (CFs) from accident reports and identifying both critical CFs and key causal paths. A multistep research design was adopted to develop and validate this novel framework for analyzing crane-related construction accident reports using ChatGPT and CN techniques. First, ChatGPT was prompted to extract CFs from a database of crane-related accident reports. Second, evaluation metrics and an expert questionnaire survey were developed to assess ChatGPT's performance in CF extraction. Finally, CN analysis was conducted to explore the relationships among CFs and to identify critical causal paths. A total of 95 crane-related accidents from Hong Kong (2011-2020) were analyzed using the proposed framework. The critical CFs identified included: "carelessness", "operation error", "crane unbalanced", "machine failure", "parts of a crane fall", "object strike", "worker fall", "trapping", "collapse of crane", and "load drop". The critical path identified was: "broken/failed rope" → "load drop" → "object strike". The primary contribution of this study lies in developing an AI-driven framework that combines the contextual understanding of ChatGPT with the structural analysis capabilities of CN methods—offering a novel and scalable approach to accident causation analysis in the construction industry. Safety managers and practitioners can leverage this framework to improve the automation, consistency, and interpretability of construction accident reporting.

Keywords: ChatGPT, large language model, complex network, accident causation analysis, construction safety, generative artificial intelligence

1. Introduction

The construction industry plays a vital role among the major industrial sectors in many countries, contributing significantly to national gross domestic product^[1]. However, it is also persistently threatened by occupational injuries and safety-related incidents. According to the International Labor Organization^[2], construction workers are three to six times more likely to experience fatal accidents compared to workers in other sectors. Therefore, enhancing safety management in the construction industry is an urgent priority. Learning from past accidents is essential for understanding their causes, reducing injury risk, identifying effective prevention measures, and improving overall safety on construction sites^[3]. For example, on 7 September 2022, a large luffing jib tower crane collapsed at a construction site in Sau Mau Ping, Hong Kong, resulting in three fatalities and six injuries^[4]. The incident shocked the public and highlighted a complex interplay of mechanical failure, procedural lapses, and organizational shortcomings.

Construction accident cases are typically reported and documented as free-text narratives, which highlights both the importance and the challenge of systematically extracting essential information, such as causal factors (CFs) from such reports. Generally, two primary methods are employed for extracting information from construction accident reports. The first is a manual approach, in which domain experts read and summarize data from unstructured texts, as demonstrated by Chen *et al.*^[5] in their analysis of 95 crane-related accidents in Hong Kong. The second approach leverages natural language processing (NLP) to reduce the subjectivity and time consumption associated with manual efforts. For example, Kim and Chi^[6] developed NLP models to retrieve accident cases and extract knowledge from textual reports, demonstrating that NLP can automate key components of the analytical process.



Although NLP has facilitated the automation of accident analysis, it still encounters challenges related to generalizability, limited sample sizes, and restricted semantic understanding. First, most existing models rely on keyword-based approaches and require retraining for each new task, which limits their applicability across diverse contexts. Second, the effectiveness of these models is constrained by insufficient training data. For example, Zhong *et al.*^[7] developed a model that performed well in classification accident cases, but its performance was limited by a lack of diverse data from additional sources. Moreover, considerable manual effort is still needed for data labeling and rule-based extraction. These limitations reduce the ability of NLP tools to comprehend and reason over complex narratives. In essence, traditional NLP models are effective at pattern recognition but lack the capacity to interpret nuanced, context-dependent information in accident descriptions.

Recently, large language models (LLMs) have emerged as a promising solution to these limitations. Pre-trained on extensive datasets, these models demonstrate strong generalization capabilities and can automate complex, context-sensitive tasks with minimal additional training^[8]. For example, ChatGPT, developed by OpenAI, excels at interpreting and generating human-like text in response to prompts^[9]. This advancement has sparked growing interest in exploring its applications, benefits, and challenges. Evidence suggests that ChatGPT can outperform humans in tasks with specific objectives, such as essay generation^[10], and criteria-based feedback on students writing^[11]. Unlike traditional NLP methods, ChatGPT can understand linguistic context and infer implicit causal relationships, making it particularly well suited for analyzing unstructured accident narratives. Its ability to function effectively with few-shot or zero-shot examples offers a distinct advantage in safety-related domains, where labeled datasets are limited and scenarios highly variable.

While ChatGPT is capable of extracting a wide range of CFs from accident narratives, it does not inherently support quantitative analysis or the systematic identification of relationships among those factors. To address this limitation, we incorporate complex network (CN) analysis^[12], a method widely used to model interactions and dependencies within various systems. In the context of accident causation, CN enables the construction of directed networks that illustrate interdependencies among CFs, quantify their significance, and identify critical causal paths^[12]. This combined use of ChatGPT for linguistic understanding and CN for structural analysis offers a more comprehensive and interpretable approach to accident causation analysis.

This study aims to investigate the innovative integration of ChatGPT into the analysis of construction accident causation. Specifically, we propose a novel framework that combines ChatGPT's linguistic comprehension with CN analysis to improve the extraction and interpretation of CFs. First, structured prompts are designed for ChatGPT to extract CFs from narrative reports. Subsequently, evaluation metrics are employed to assess ChatGPT's performance. Then, CN methods are utilized to analyze the relationships among the extracted factors, demonstrating the scientific validity of the outputs generated by ChatGPT. Finally, the framework is validated using a dataset of crane-related accidents, and practical recommendations are provided based on the critical CFs and causal pathways identified. By leveraging ChatGPT's contextual reasoning capabilities, this framework enhances both the generalizability and automation in construction accident analysis.

2. Literature Review

This section begins with a comprehensive review of the literature on construction accident analysis. It then discusses the applications of NLP within the construction industry and reviews the advancements in LLMs. Finally, the section identifies existing research gaps and outlines the study's objectives to clarify its motivation.

2.1 Accident analysis in construction

Research on accident analysis in construction relies heavily on reliable and sufficient data. Available accident cases typically originate from official databases and individual research datasets. Official databases are maintained and updated by organizations or institutions, such as those in the United States^[13] and South Korea^[14]. In contrast, personal datasets are developed by researchers to fill gaps in official data, exemplified by the European fatal accident database compiled by Morris *et al.*^[15] and the subway construction accident database created by Zhou *et al.*^[16].

Research methods for construction accident analysis can be broadly categorized into statistical analysis and conceptual studies. Statistical approaches rely on large datasets to compare and analyze accident characteristics such as type, timing, site conditions, casualties, and economic losses^[17]. Common techniques include principal component analysis, ordinary least squares regression, among others. For example, Irumba^[18] identified spatial correlations between accidents in Kampala, Uganda, and surrounding regions, while Zhang *et al.*^[19] examined the frequency of CFs and accidental severities based on 41 tower-crane incidents in China. However, this approach can be limited by significant manual effort required for data collection, and unscientific sampling of the dataset may weaken the conclusions.

Conceptual studies summarize theoretical models based on practical accident cases and expert knowledge to investigate the causes and evolution mechanisms of accidents. Common approaches involve accident causation models that explain why accidents occur from various theoretical perspectives, including human factors, management, unsafe behaviors, system theory, and cognitive neuroscience. Notable examples include Leveson's systems theoretic accident model and processes^[20], Hollnagel's functional

resonance analysis method^[21], and the human factors analysis and classification system developed by Shappell and Wiegmann^[22]. However, these models require extensive validation through analysis of numerous real-world accident cases over time.

Recently, there has been growing interest in applying CN analysis to integrate quantitative and qualitative approaches, thereby combining the strengths of both statistical and conceptual methods. For example, drawing on systems thinking, Zhang *et al.*^[12] developed a CN model to identify critical CFs and pathways in tower-crane accidents in China. Their analysis involved 195 empirical accident cases and employed a causation model comprising six subsystems and 34 CFs. However, extracting information from accident reports remains a labor-intensive process. Since investigative reports are the predominant format for accident documentation, CN analysis continues to rely heavily on manual interpretation of unstructured text. In particular, the extraction of CFs requires advanced cognitive skills, including textual comprehension, criterion-based judgment, logical reasoning, and causal inference. Consequently, there is an urgent need to improve the level of automation in CN analysis to reduce manual workload and enhance overall efficiency.

2.2 Natural language processing in construction

NLP enables machines to understand and generate human languages by analyzing textual data, thereby effectively replacing the need for manual operations^[23]. With recent advancements in artificial intelligence (AI), particularly in machine learning (ML), and deep learning (DL), NLP techniques have seen growing adoption in the construction industry^[24], especially in areas such as document management related to accident analysis.

Document management is one of the most direct applications of NLP in the construction industry, involving tasks such as knowledge retrieval, document classification, and question-answering systems. For example, Sun *et al.*^[25] developed a framework to extract information from monthly construction reports, visualizing the content using keyword networks and tag clouds. Hassan and Le^[26] applied ML-based models to classify contract documents and accurately identify relevant text segments. Zhong *et al.*^[27] combined DL and NLP to develop a chatbot capable of automatically responding to building regulation inquiries. Notably, the performance of their approach was influenced by the size of the training datasets.

Another key area of NLP application in construction is safety management, where accident reports constitute the primary data source. NLP-based analysis in this domain generally fall into three categories: (1) Classification of accident information: For example, Fang *et al.*^[28] employed DL models to classify near-miss incidents reported in safety reports; (2) Extraction of safety information: Xu *et al.*^[29] proposed a rule-based NLP approach to extract key elements related to construction safety management from textual data, including accident types, occurrence times, causes, and consequences; and (3) Retrieval of semantic knowledge: Zou *et al.*^[30] applied NLP techniques to retrieve similar cases from accident databases to support risk assessment in construction projects. However, these methods still rely on manual pre-processing, rule definition, data labeling, and model training, and face limitations in generalization and sample size.

2.3 The development of large language models

LLMs are a significant recent advancement in AI. Studies highlight their vast potential in handling NLP tasks such as automatic classification^[31] and machine translation^[32]. ChatGPT, developed by OpenAI, has gained attention due to its outstanding performance. GPT-3.5, launched in November 2022, became the fastest-growing consumer software application by January 2023. Trained on a 570GB dataset from the internet^[9], ChatGPT demonstrates the potential to address limitations of earlier NLP models, such as restricted training data scale and issues with generalizability. The successful training strategy positions ChatGPT to outperform traditional ChatBot systems in content awareness and text generation.

In the construction industry, only a limited number of studies have examined the applications of ChatGPT, with a focus on project scheduling, work progress, and daily report generation. Prieto *et al.*^[8] investigated the potential of ChatGPT in automating project scheduling and found that it could forecast accurate timeline with an average precision of 81.3%. However, the study was limited by a small sample size. You *et al.*^[33] proposed RoboGPT, which leverages ChatGPT's reasoning capabilities to automate sequence planning in robot-based assembly tasks. Their evaluation, which included two case studies and 80 trials, demonstrated that RoboGPT effectively guided robots through complex operations. Xiao *et al.*^[34] developed a method that automates daily report generation from construction videos by combining ChatGPT with computer vision. The method achieved 91.13% accuracy in productivity analysis and received a report quality score of 4.23 out of 5.0. Notably, ChatGPT's advanced capabilities in interpreting construction accident reports were initially explored in our previous study^[35], although that study did not incorporate any causation analysis techniques, which limited its ability to further examine the underlying causes of construction accidents.

2.4 Research gaps and research objectives

As discussed earlier, analyzing unstructured text in accident reports is essential for understanding the causes of accidents and implementing effective preventive measures. Developing an automated and scalable framework has the potential to significantly enhance both the efficiency and reliability of construction safety management. However, the current body of research reveals three critical methodological gaps that remain unresolved:

Research Gaps:

Limited automation of accident report analysis: Existing approaches in construction accident analysis still depend heavily on human effort, resulting in a high workload and subjectivity. Although NLP-based methods show promise, they continue to require substantial manual intervention for tasks such as pre-processing, rule definition, data labeling, and model training. There is an urgent need for innovative methodologies, such as the application of LLMs, to reduce manual effort and enable automated, end-to-end analysis workflows.

2. Poor generalizability and limited data scalability: Most NLP methods are developed for narrowly defined tasks and exhibit limited transferability across diverse accident scenarios due to their dependence on small, highly domain-specific datasets. Additionally, high costs associated with data acquisition and the presence of data monopolies result in insufficient training sample sizes, which severely restrict model scalability. Consequently, there is an urgent need for innovative approaches that can harness the extensive training capabilities of LLMs to overcome these data limitations and improve model generalization across a wide range of construction accident cases.

3. Lack of interpretability in causation analysis using LLM outputs: Although recent studies have applied ChatGPT and other LLMs to tasks such as scheduling, work progress monitoring, and document summarization within construction, no research to date has systematically investigated their potential for structured causation analysis in safety management. Furthermore, existing work has not addressed the challenge of converting the unstructured outputs generated by LLMs into interpretable and scientifically valid models for causation analysis. Developing an integrated methodology that combines LLM-driven information extraction with structured analysis techniques such as CN modeling constitutes a novel and unexplored research direction.

To fill these gaps, three research objectives have been proposed:

Research objectives:

1. Applying LLMs to minimize manual work: Leveraging the advanced language understanding and generation capabilities of LLMs, this study aims to design structured prompts that enable ChatGPT to autonomously extract CFs from narrative construction accident reports. The structured extraction process is intended to significantly reduce human intervention and enhance consistency in accident data analysis.

2. Evaluating ChatGPT's performance for CF extraction: To address the limitations of generalizability and data scarcity inherent in traditional NLP methods, this study will establish evaluation metrics and validation strategies to assess ChatGPT's effectiveness as an alternative approach for CF extraction.

3. Integrating ChatGPT with CN modeling for interpretability: To overcome the current lack of interpretability in LLM-based analyses, this study proposes constructing a causation CN model based on CFs extracted by ChatGPT. Analysis of the CN model will reveal network topological features, offering structured and visual insights into causal relationships, thereby demonstrating a novel methodology that integrates LLM-driven text analysis with systematic safety analysis frameworks.

3. Methodology

The proposed framework for automated causation analysis consists of five key steps: data collection, factor extraction, evaluation and correction, network establishment, and network analysis. The overall structure is illustrated in [Figure 1](#). First, in the data collection module, a suitable database of accident reports is selected. Once an accident report is input, the factor extraction module utilizes ChatGPT to identify CFs from the accident descriptions using predefined prompts. In the evaluation and correction module, the extracted results are reviewed and refined through a questionnaire-based expert survey to incorporate domain-specific knowledge. The validated CFs are then passed to the network establishment module, where an accident causation CN model is constructed. Finally, the network model is analyzed and visualized, with particular emphasis on its topological characteristics. This framework facilitates automated causation analysis based on unstructured textual reports.

3.1 Data collection

The data collection module identifies the accident reports used as input for the proposed framework. In this study, a total of 95 crane-related accident reports from Hong Kong, spanning the period from 2011 to 2020, were gathered for causation analysis^[5]. These reports were manually collected from reliable sources, primarily retrieved from the official website of the Labour Department and supplemented by reputable mass media publications. To mitigate subjectivity and ensure data quality, two research assistants independently reviewed the dataset, retaining only those reports that contained sufficient detail for analysis. The resulting database is recognized for its transparency, adequate narrative length, and non-redundant content, making it well-suited for application within the proposed framework. Furthermore, this dataset has also been cited in previous studies on crane operation hazards^[36] and construction worker fatigue detection^[37].

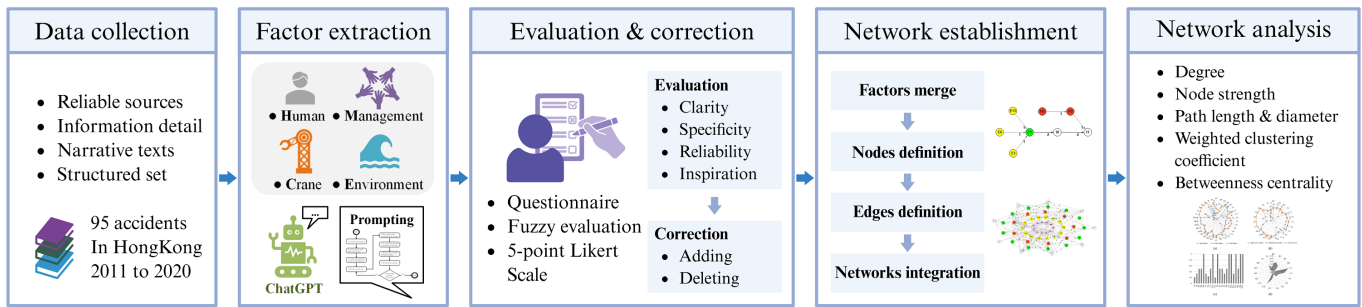


Figure 1. Framework of the proposed method.

3.2 Factor extraction

The factor extraction module employs ChatGPT to identify CFs from the collected accident reports, based on clearly defined task requirements and well-structured prompts.

Task requirements are clarified by designing specific dimensions for CF identification. This study adopts the Man-Machine-Environmental Systems Engineering theory as the theoretical foundation for dimension design. Established in 1981 under the guidance of Xuesen Qian, represents an advanced interdisciplinary field that leverages systems science and engineering to analyze interactions among humans, machines, and the environment^[38]. For crane accidents in this study, considering cranes as the main machine involved and the importance of on-site construction safety management, four CF identification dimensions are designed: human factors, crane factors, environmental factors, and management factors (H-C-E-M dimensions). Accordingly, ChatGPT categorizes the extracted factors into these classes based on 95 accident reports.

Effective instructions, or prompts, are natural language descriptions of task requirements provided to LLMs by humans. In this study, a structured prompting process consisting of eight steps is proposed to create precise and machine-understandable prompts that enable satisfactory output performance from ChatGPT. As illustrated in Figure 2, the process begins with the researcher clarifying the task purpose, followed by drafting specific prompts that avoid open-ended or overly broad language. These drafted prompts are then carefully reviewed and revised. The revised prompts are input into ChatGPT, with an emphasis on encouraging extended responses. After experimenting with various prompts and questioning styles, the outputs generated by ChatGPT are recorded. Finally, these outputs are assessed to determine whether they meet expectations. If not, the process returns to the prompt revision step. By following this iterative procedure, effective prompts for extracting CFs from accident reports can be developed.

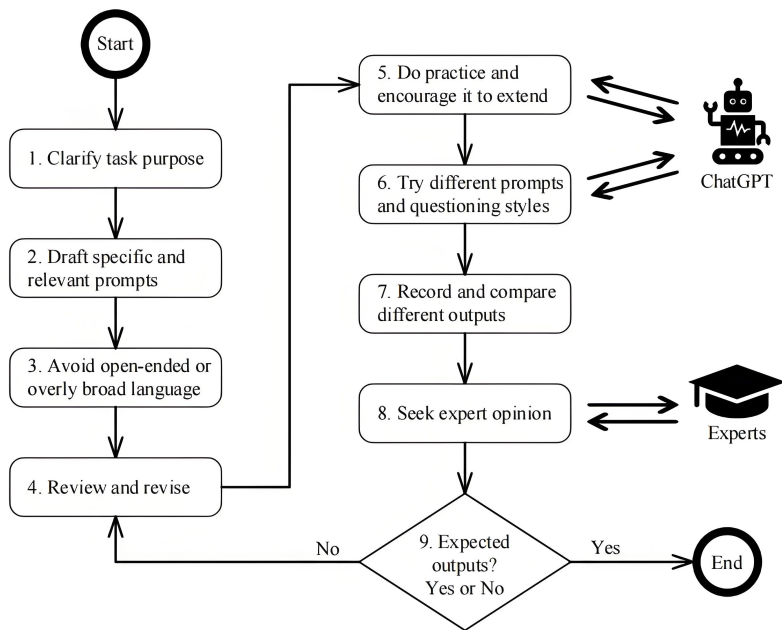


Figure 2. Structured process for obtaining effective prompts.

3.3 Evaluation and correction

The evaluation and correction module assesses ChatGPT’s performance in factor extraction for accident causation analysis and prepares validated input for the network establishment module. This module consists of two stages: (1) evaluating ChatGPT’s factor extraction and (2) correcting the extracted results based on expert experience.

In the first stage, participants evaluate ChatGPT’s output using four metrics: clarity, specificity, reliability, and inspiration^[39]. After reviewing the original accident report, they rate the output on a five-point Likert Scale, where 1 = “strongly disagree”, 2 = “disagree”, 3 = “neutral”, 4 = “agree”, 5 = “strongly agree”^[40].

- Clarity: The output is clear and easy to read.
- Specificity: The extracted factors are distinct, independent, and not duplicated.
- Reliability: The factors are directly supported by the content of the original report.
- Inspiration: The factors are reasonably inferred from the report, even if not explicitly stated.

In the second stage, experts revise the extracted factors based on their professional knowledge and experience. Specifically, they add any relevant factors that were not identified by ChatGPT or remove those that are not supported by the original report. The corrected set of factors is then used as input for the network establishment module.

3.4 Network establishment

This module constructs a directed weighted CN model for accident causation analysis, highlighting the practical value of the extracted factors. CN theory can effectively illustrate correlations among factors and system regularities^[24]. The establishment of the directed weighted CN is guided by two assumptions: (1) nodes represent accident CFs extracted from reports, and (2) directed edges represent the causal relationships between these factors. Accordingly, the network establishment module involves four steps: factor merging, node definition, edge definition, and network integration.

In the first step, synonymous factors are merged to improve clarity and facilitate analysis. Extracted factors that have the same meaning but are expressed differently are renamed using consistent terminology. For example, “dark conditions”, “dark environments”, and “poor lighting and visibility conditions at night” are uniformly renamed as “poor visibility”. Through this process, 239 raw factors were consolidated into 45 CFs and 9 outcome factors (OFs). Two examples of this merging process are provided in Table 1.

Table 1. Two examples of merging factors.

Merged factors	Raw factors	Report No.
Improper inspection and maintenance (CF)	Failure to ensure proper inspection and maintenance of the crane and its components (M)	12
	Failure to properly inspect the wire rope before use (M)	21
	Lack of qualified inspection of the bridge girder erection machine (H)	22
	Failure to repair the defective iron hook (H)	25
	old truck-mounted crane killed another operator (M)	35
	Lack of maintenance or inspection (+)	42
	The crane was too old to repair (C)	79
Worker fall(OF)	Lack of proper inspection and maintenance procedures for shackles (M)	83
	The driver fell to the ground with the cab (+)	6
	The lifted worker unbalanced and fell (+)	11
	The worker unbalanced and fell (+)	13
	Fall from the height (+)	22, 40, 52, 68, 90, 93
	The worker fell from the platform (H)	49, 70, 78
	Loss of balance of the worker (H)	54
	The operator fell out of the cab (+)	81, 84

CF: causal factors; OF: outcome factors; H: human-related factors; C: crane-related factors; E: environment-related factors; M: management-related factors; +: new factors added by experts that failed to be extracted from the original report.

Next, nodes are defined into two categories, as detailed in Table S1. The first category consists of 45 accident factors derived from the merged CFs, which are further grouped into human-related, crane-related, environment-related, and management-related categories. These factors are labeled using initials and numbers (e.g., H1, H2, C1, E3). The second category comprises 9 incident outcomes based on merged OFs, labeled as I1, I2, I3, and so forth.

Thirdly, directed edges are defined to represent the relationships among nodes. Based on the descriptions in the accident reports, these edges illustrate the influences between CFs. Two examples of the directed edge identification process are provided in Table 2. The first column lists the report serial numbers, the second contains the accident descriptions, the third presents the extracted factors with their merged labels, and the fourth depicts the causal relationships. The network model is constructed separately for each of the 95 accident reports using this approach.

Table 2. Two examples of identifying relationship.

Report No.	Description of accident	Factor	Relationship
62	A worker was operating a medium-sized crawler-mounted mobile crane. During this period, he was working near a hole with an area of about 8 meters by 8 meters and a depth of more than 10 meters. The crane suddenly lost balance and turned over. A worker nearby was crushed by the crane boom, got hit on his feet, and almost fell to the platform together with the crane. The worker was sent to the hospital for treatment and there were multiple fractures in both feet.	Large hole side operation (E13); Crane unbalanced (C1); Crane turned over (I8); Worker crushed (I1);	$E13 \rightarrow C1 \rightarrow I8 \rightarrow I1$;
64	When the operator was driving the excavator on a slope toppled sideways, it was suspected that the soil was too soft. Consequently, the excavator was out of balance and suddenly rolled over. The operator jumped out of the excavator to escape but failed and was overwhelmed by the excavator to death. The Labor Department has issued suspension notices to the contractors involved to suspend. The contractors cannot resume the work until the LD is satisfied that measures to abate the relevant risk have been taken.	Inadequate safety awareness and response of the operator in emergencies (H5); Operator error of self-rescue (H3); Lack of stability of the excavator (C1); Softness and instability of the soil (E3); Slope (E6); Crane rolled over (I8); Operator crushed (I1);	$(E3+E6) \rightarrow C1 \rightarrow I8$; $H5 \rightarrow H3 \rightarrow I1$;

Finally, the 95 individual network models are integrated into a comprehensive network. During this integration, duplicate nodes are merged, and duplicate edges are represented by weighted values indicating their frequency. Figure 3 illustrates this integration process using an example of networks 62 and 62 from Table 2. Both accident cases include the causal relationships between nodes “C1”, “I8”, and “I1”, resulting in a causal chain weight of 2.

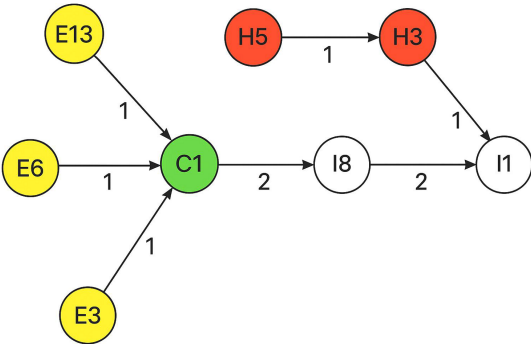


Figure 3. An example of integrating networks.

In this way, all 95 causal networks were integrated into an overall directed weighted CN model, as shown in Figure 4. Using Pajek, a software specialized in the analysis and visualization of large networks, the directed weighted crane accident causation network can be modeled, visualized, and analyzed^[41]. This model consists of 56 nodes, color-coded according to different factor types, and 145 edges, whose weights and thickness represent the frequency of the causal relationships.

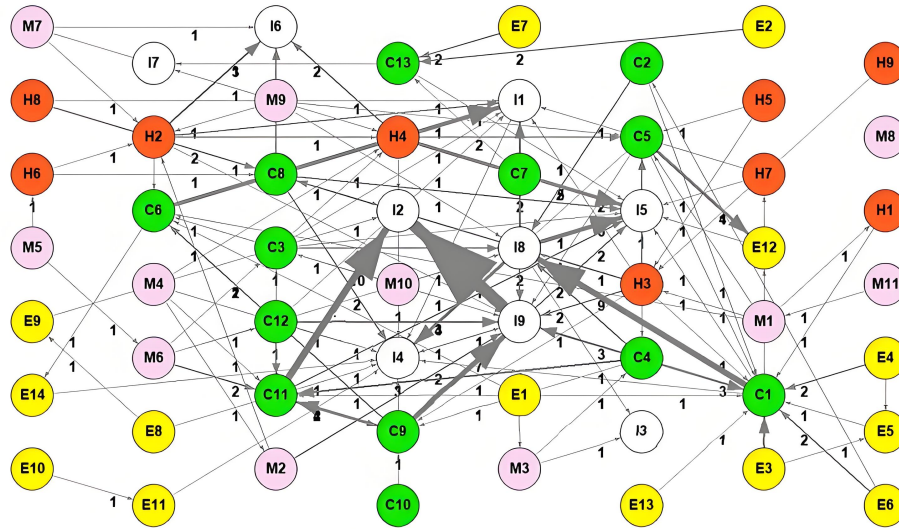


Figure 4. The directed weighted crane accident causation network.

3.5 Network analysis

The directed weighted CN model is effective in analyzing natural and man-made systems^[42]. Computing its topological properties provides a comprehensive analysis of accident propagation and evolution mechanisms^[43]. This study selects six indicators for evaluation: degree, node strength, average path length, diameter, weighted clustering coefficient, and betweenness centrality.

A directed weighted network G with N nodes can be mathematically represented by an $N \times N$ adjacency matrix A with element, as shown in Eq. (1), where a_{ij} takes the value 1 if node i points to node j and 0 otherwise, and ω_{ij} is the weight value of the links from node i to node j .

$$A_{ij} = \begin{cases} a_{ij} \cdot \omega_{ij} & \text{if node } i \text{ points to node } j, \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

(1) Degree

The degree k_i of a node i in a network reflects its importance based on the number of links connected to it. In a directed network, the total degree k_i^{total} of a node i is the sum of its output degree $k_i^{\text{out}} = \sum_j a_{ij}$ (outgoing links), and input degree $k_i^{\text{in}} = \sum_j a_{ji}$ (incoming links)^[44].

(2) Node Strength

Node strength s_i reflects the importance of a node in a weighted network by accounting for both the number of links and their corresponding weights^[42]. In a directed weighted network, the total node strength s_i^{total} of a node i is divided into output node strength $s_i^{\text{out}} = \sum_j \omega_{ij}$ and input node strength $s_i^{\text{in}} = \sum_j \omega_{ji}$.

(3) Shortest Path Length, Average Path Length and Diameter

In a directed weighted network, the shortest path length d_{ij}^{ω} between a pair of nodes (i, j) is defined as the minimal total inverse weights along the edges connecting them^[21], as expressed in Eq. (2). The average path length L of the network is the mean shortest path length calculated over all pair of nodes (i, j) , as described in Eq. (3). The diameter of the network is the longest among all the shortest path lengths^[43]. These measures reflect the efficiency of information or energy transmission within the network.

$$L = \frac{1}{N(N-1)} \sum_{i,j \in N, i \neq j} d_{ij}^{\omega} \quad (2)$$

$$L = \frac{1}{N(N-1)} \sum_{i,j \in N, i \neq j} d_{ij}^{\omega} \quad (3)$$

(4) Weighted Clustering Coefficient

In a weighted network, the weighted clustering coefficient C_i^{ω} is defined without considering edge direction, as shown in Eq. (4). Here, $\omega'_{ij} = \omega_{ij} + \omega_{ji}$ represents the combined weight of the edge between node i and node j , and $a'_{ij} = 1$ (if $a_{ij} = 1$ or $a_{ji} = 1$). This coefficient reflects both the number of closed triplets around a node and the total weight of these triplets relative to the vertex's strength^[42].

$$C_i^{\omega} = \frac{1}{s_i(s_i - 1)} \sum_{j,h} \frac{\omega'_{ij} + \omega'_{ih}}{2} a'_{ij} a'_{ih} a'_{jh} \quad (4)$$

(5) Betweenness Centrality

Betweenness centrality b_i quantifies the importance of node i by counting the number of shortest paths that passing through it^[44]. It is defined in Eq. (5), where j and h represent two non-adjacent nodes, n_{jh} is the total number of shortest paths between nodes j and h , and $n_{jh}(i)$ denotes the number of those paths that pass through node i .

$$b_i = \sum_{j,h \in N, j \neq h} \frac{n_{jh}(i)}{n_{jh}} \quad (5)$$

4. Results for the Performance of ChatGPT

4.1 Factor extraction results

This subsection presents the outcomes of the factor extraction module, including the finalized prompt wording and an example output generated by ChatGPT. The GPT-3.5 model was selected due to its widespread use and significant influence. Unless otherwise specified, “ChatGPT” in this section refers to GPT-3.5. Following the proposed prompting process and incorporating two rounds of revisions alongside expert feedback, the prompt wording was finalized as follows:

“I would like you to act as an expert in safety accident analysis in the construction field. I have collected some case descriptions of construction crane accidents and would like to extract and analyze the causes of these accidents. Your task is to extract, summarize, and list, in ‘concise, precise, and professional phrases, not sentences’, all the causes (if any) that directly or indirectly contributed to the accidents, and divide the extracted factors into four areas: human-related, crane-related, environment-related, and management-related/InlineParagraph>management-related/InlineParagraph> factors, without explanation. Note: please keep in mind that the causes listed are based solely on the case description provided, and not on any outside knowledge or information. My first/next case description of a construction crane accident is __ (Accident description text is pasted here)”.

By using ChatGPT, researchers efficiently extracted CFs by inputting the accident report text and recording the generated outputs. Figure 5 presents an example in which ChatGPT clearly lists the extracted CFs, demonstrating its comprehension of the task requirements.

 **Case 3:**
A green hand worker while operating a bulldozer for earthmoving work in a reclamation site, fell into the sea together with the bulldozer out of control. Because the driver did not carry a life-saving device with him, he was trapped and unable to swim ashore on his own. According to a witness at the scene, the subcontractor involved often used novices to work, and it was estimated that it was a cost consideration. Besides, the contractor involved has not yet arranged to pick up the bulldozer that fell into the sea. Labour Department has issued suspension notices to the contractors involved to suspend. The contractors cannot resume the work until the LD is satisfied that measures to abate the relevant risk have been taken.

 **ChatGPT's output:**

Human-related factors: ● Inexperienced worker operating heavy machinery ● Lack of proper training and supervision ● Failure to wear personal protective equipment ● Lack of life-saving device on the worker ● Subcontractor frequently using novice workers	Crane-related factors: None Environment-related factors: ● Proximity of the work site to the sea Management-related factors: ● Lack of arrangements for the bulldozer retrieval ● Failure to provide safety measures for workers
---	--

Figure 5. An example of ChatGPT's output.

4.2 Evaluation and correction results

This subsection presents the outcomes of the evaluation and correction module. This evaluation was conducted through a face-to-face questionnaire involving nine participants, all of whom had over five years of experience in construction management research. To ensure fairness and efficiency, the 95 accident cases were randomly divided into three groups A, B, and C with 32, 32, and 31 cases, respectively. Each group was assigned to three participants. Participants performed fuzzy evaluations of ChatGPT’s factor extraction outputs based on the original accident reports. To minimize subjectivity, participants were first briefed on the research objectives and evaluation criteria. A pretest involving three randomly selected accident cases from the database was then conducted to establish consensus on evaluation and correction standards. After completing the pretest individually, participants engaged in a discussion about ChatGPT’s outputs and reached agreement on the following points: (1) there is no standardized answer for this questionnaire; (2) participants are encouraged to rely on their professional experience for independent judgment; and (3) factors extracted in Task 1 must be clearly justified within the accident reports.

Figure 6 presents the evaluation results of ChatGPT’s performance in extracting accident factors. The 95 accident cases were divided into groups A, B, and C, represented by each column, while ChatGPT’s performance was assessed across four criteria: clarity, specificity, reliability, and inspiration, shown in each row. The vertices of the radar chart correspond to case numbers, and the vertical axis represents the scores assigned by the experts. Different polygon colors indicate different experts, with larger polygon areas suggesting higher performance ratings. The results indicate that ChatGPT was rated positively for clarity and specificity, neutrally for reliability and negatively for inspiration. Notably, the lack of inferred factors beyond the provided information led to lower scores in inspiration. Furthermore, the closer alignment of polygons reflects higher agreement among experts, thereby enhancing the credibility of the evaluation. However, expert G in group C rated clarity significantly differently from other experts, due to a misunderstanding of task requirements.

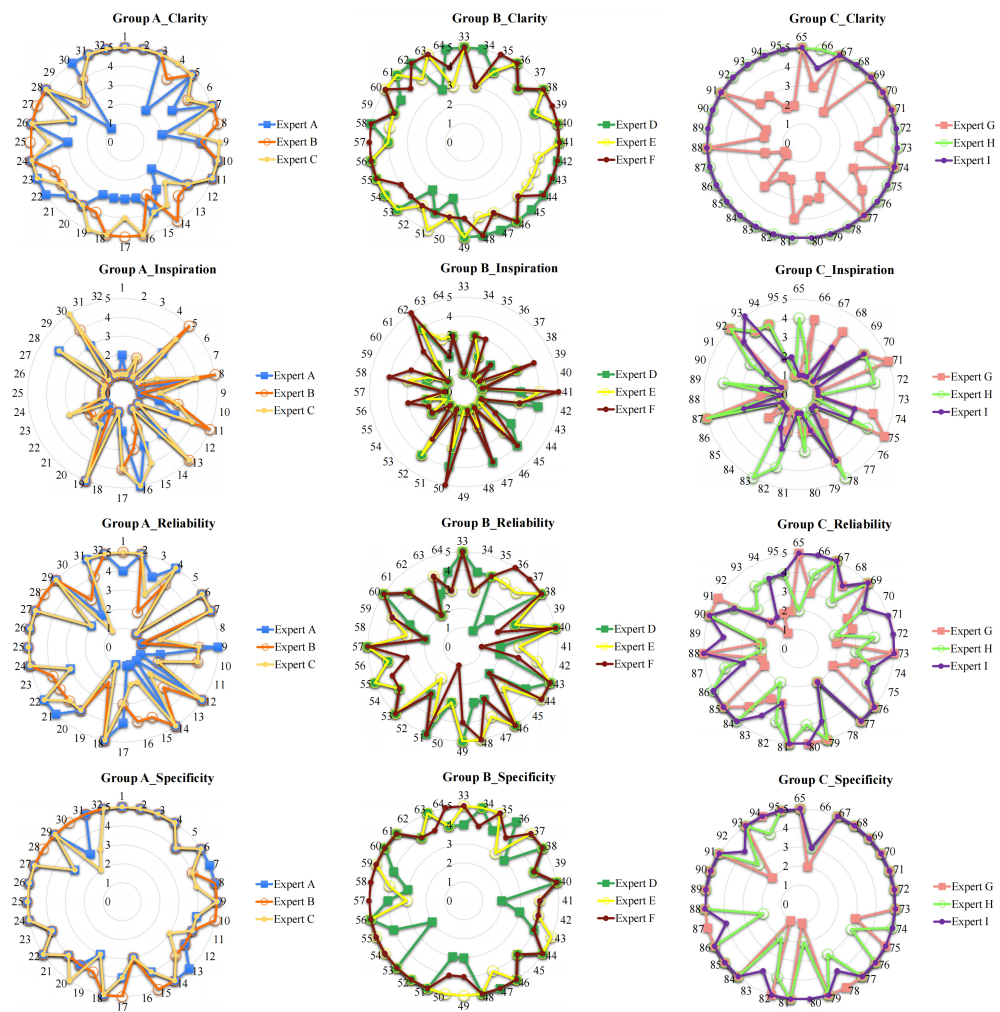


Figure 6. Evaluation results of the ChatGPT’s performance for factor extraction.

5. Results for Causation Network Analysis

This section presents the results of the network analysis module, as shown in **Figure 7**. The analysis includes six indicators: degree, node strength, average path length, diameter, weighted clustering coefficient, and betweenness centrality. These calculations provide a deeper understanding of the network by identifying key nodes, interactions, and critical paths.

(1) Degree

Figure 7a illustrates the distribution of nodes with a total degree greater than 2. Notably, I4 (Worker fall) and I5 (Trapping) show the highest input degree values of 16 and 13, respectively, highlighting their important roles as frequently affected incident outcomes in crane-related accidents. Regarding output degree, H3 (Operation error), H2 (Carelessness), and C3 (Machine failure) have the highest values of 9, 8, and 8, respectively, indicating these factors as primary causes of such accidents. Overall, the nodes with the highest total degree are C1 (Crane unbalanced) and I4 (Worker fall), suggesting that crane imbalance is a critical factor and preventing worker falls should be a key safety priority.

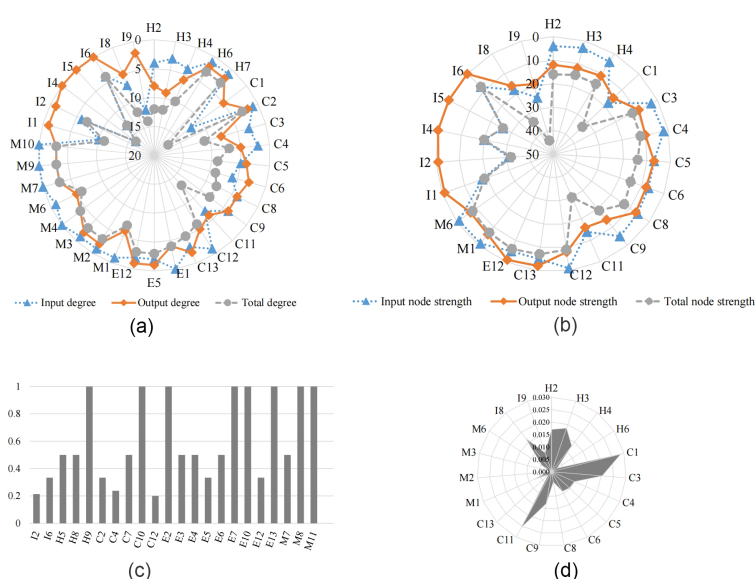


Figure 7. Measurement criteria of the directed weighted crane accident causation network: (a) Nodes with a total degree greater than 2; (b) Nodes with a total node strength greater than 5; (c) Nodes with a weighted clustering coefficient greater than 0.2; (d) Nodes with a betweenness centrality greater than 0.001.

(2) Node Strength

Figure 7b presents nodes with total strength exceeding 5. Accident factors C1 (Crane unbalanced) and C11 (Parts of a crane fall) exhibit the highest total node strengths of 33 and 30, respectively, indicating that these factors require stringent attention and effective control measures during the production process. Among incident outcomes, I9 (Load drop), I8 (Collapse of crane), and I2 (Object strike) have the highest total node strengths of 44, 32, and 30, respectively, indicating the need to focus outcomes to prevent secondary accidents.

(3) Shortest Path Length, Average Path Length and Diameter

In the directed weighted crane accident causation network, the average path length is 0.410, indicating that the typical connection between two accident CFs is less than one step. The network's diameter, defined as the longest among all calculated shortest paths, is 3.9. The diameter value is substantially greater than the average path length, highlighting potential causal relationships that may often be overlooked. Furthermore, the 10 shortest paths ending at incident outcome nodes share the sequence C9 (Broken/failed rope) → I9 (Load drop) → I2 (Object strike), suggesting that these factors and their causal links should be prioritized in accident prevention and investigation.

(4) Weighted Clustering Coefficient

Nodes with weighted clustering coefficients ranging from 0.2 and 1.0 are shown in **Figure 7c**. To prevent ripple effects, factors with larger coefficients require greater control. Among the 22 factors with coefficients above 2, nine are environment-related and five are crane-related, indicating that these nodes tend to form clusters. Moreover, the network's average weighted clustering coefficient of 0.2952 is significantly higher than that of a random network of the same size. This suggests that most nodes are susceptible to the influence of others, thereby increasing the complexity of accident prevention for both workers and managers.

(5) Betweenness Centrality

Nodes with a betweenness centrality greater than 0.001 are shown in Figure 7d. Among them, eight nodes exceed a centrality value of 0.01: C1 (Crane unbalanced) at 0.0286, C11 (Parts of a crane fall) at 0.0255, C3 (Machine failure) at 0.0204, H3 (Operation error) at 0.0185, H2 (Carelessness) at 0.0170, I8 (Collapse of crane) at 0.0169, H4 (In an unsafe position) at 0.0131, and C9 (Broken/failed rope) at 0.0130. These results indicate that such factors frequently serve as key bridges connecting various CFs in crane-related accidents. Therefore, special attention should be paid to these nodes by workers and managers to help prevent the propagation and escalation of accidents.

6. Discussion

This section discusses the implications of the study’s findings, focusing first on the evaluation and correction of ChatGPT, followed by the identification of critical CFs and causal path. The limitations of the study are also discussed.

6.1 Evaluation and correction of ChatGPT

This subsection discusses the results of the evaluation and correction module. Firstly, ChatGPT demonstrated strong performance in producing clear and specific extraction results. As shown in Figure 6, ChatGPT consistently received positive feedback on clarity and specificity from all nine expert evaluators, underscoring its potential applicability in construction accident analysis. Furthermore, ChatGPT was explicitly instructed to extract only the CFs mentioned in the original reports. Interestingly, its relatively low scores in the inspiration category reflect greater precision, as no speculative or imaginative content was generated. This finding suggests that ChatGPT may also be well suited for other structured information extraction tasks, such as accident report classification or knowledge retrieval, although further validation across different domains is warranted.

Secondly, the evaluation revealed occasional inconsistencies caused by “hallucination”^[45], a known issue in LLMs wherein the model generates plausible yet inaccurate information. This issue arises from limitations in pre-training data and the absence of real-time verification mechanisms. To address this challenge, our study employed an expert evaluation and correction process, in which domain experts systematically reviewed ChatGPT’s outputs to remove inaccurately inferred CFs and to supplement any omitted factors. This human-in-the-loop approach, illustrated in Figure 8, proved effective in enhancing output reliability and provides a practical method for mitigating model uncertainty in safety-critical contexts. Other strategies discussed in the literature, such as supervised fine-tuning with domain-specific datasets^[46], automated fact-checking^[47], and continuous user feedback, may further improve robustness in future applications. Our findings indicate that expert feedback plays a crucial role in improving the stability and reliability of ChatGPT’s output in construction safety research and broader industrial contexts.

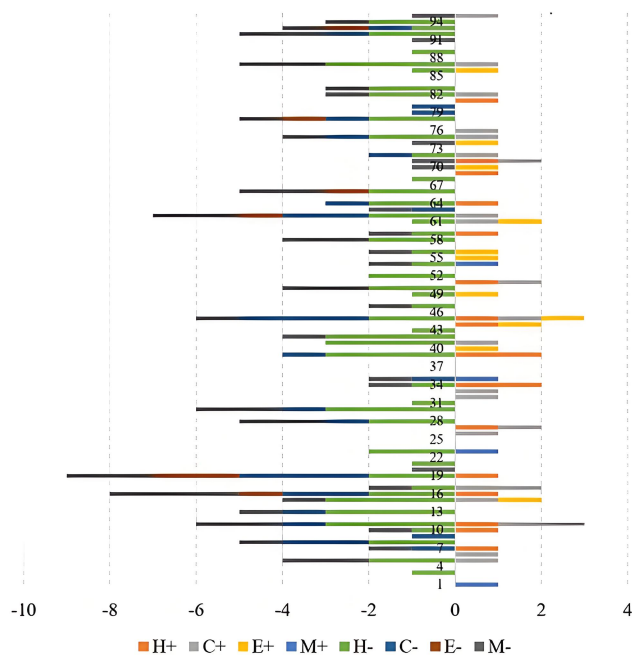


Figure 8. Evaluation results of the ChatGPT’s performance for factor extraction. H+: added human-related factors; H-: deleted human-related factors; C+: added crane-related factors; C-: deleted crane-related factors; E+: added environment-related factors; E-: deleted environment-related factors; M+: added management-related factors; M-: deleted management-related factors.

Thirdly, a comparative analysis of ChatGPT and human expert extractions revealed distinct patterns. ChatGPT tended to overgeneration, producing redundant CFs rather than omitting key ones. Domain experts refined these outputs by adjusting the identified factors across the H-C-E-M dimensions. As shown in Figure 8, deletions (negative values) and additions (positive values) were represented using distinct colors. ChatGPT more readily identified human- and management-related factors compared to human experts, whereas human evaluators were more likely to supplement additional human- and crane-related factors. For environment-related factors, the extraction results from ChatGPT and human experts were largely consistent, likely due to the relatively small proportion of environment-related entries in the dataset.

6.2 Critical causal factors and path analysis

Based on the preceding analysis of six indicators used to evaluate the directed weighted crane accident causation network, certain nodes and paths were found to play more prominent roles within the network model. As a result, critical CFs and causal paths related to crane accidents can be identified, as summarized in Table 3. Considering input degree, output degree, and total degree, the nodes with significantly higher values include: I4, I5, H3, H2, C3, and C1. Similarly, C1, C11, I9, I8, and I2 are notable for their high node strength. Among the top ten shortest paths terminating at incident outcome nodes, the sequence C9 → I9 → I2 was identified as a critical causal path. In addition, the nodes with the highest betweenness centrality are C1, C11, and C3. After removing duplicates, the final set of critical factors for crane-related accidents comprises H2 (Carelessness), H3 (Operation error), C1 (Crane unbalanced), C3 (Machine failure), C11 (Parts of a crane fall), I2 (Object strike), I4 (Worker fall), I5 (Trapping), I8 (Collapse of crane), and I9 (Load drop). These CFs and their associated relationships require increased attention in crane accident safety management.

Table 3. Critical causal factors and paths analysis.

Measured indicators	Critical causal factors	Critical causal paths
Degree	I4, I5, H3, H2, C3, C1	-
Node strength	C1, C11, I9, I8, I2	-
Shortest path	-	C9 → I9 → I2
Betweenness centrality	C1, C11, C3	-
Integrated analysis	H2, H3, C1, C3, C11, I2, I4, I5, I8, I9	C9 → I9 → I2

Regarding incident outcome nodes, namely I2 (Object strike), I4 (Worker fall), I5 (Trapping), I8 (Collapse of crane), and I9 (Load drop), these represent the crane-related accident types characterized by high frequency and severity. Consequently, construction safety managers must prioritize the prevention of these particular accident types to improve overall construction safety performance. In terms of causal factor nodes, human-related and crane-related factors are especially critical in crane accidents. Key factors include H2 (Carelessness), H3 (Operation error), C1 (Crane unbalanced), C3 (Machine failure), and C11 (Parts of a crane fall). Therefore, it is essential to implement a comprehensive training program for crane operators, riggers, and site personnel, focusing on safe operating procedures, hazard recognition, and emergency response protocols. Additionally, a rigorous schedule for regular inspection and maintenance of cranes and associated equipment should be established to ensure optimal operational condition. Moreover, disrupting the causal chain of accidents is recognized as an effective accident prevention strategy. This study identifies the causal pathway C9 (Broken/failed rope) → I9 (Load drop) → I2 (Object strike) as critical in crane-related accidents, necessitating focused control measures. To mitigate risks associated with this pathway, enhanced practices for rope inspection and maintenance are imperative. Specifically, crane ropes should be inspected regularly to detect signs of wear, corrosion, or damage, and any compromised ropes must be replaced promptly to prevent failure. Specifically, a meticulous maintenance protocol involving regular inspections of crane ropes is warranted to detect early signs of wear, corrosion, or damage. Timely identification of such issues facilitates prompt replacement of worn or damaged ropes, thereby preventing potential failures and subsequent accidents.

6.3 Contributions and limitations

This study makes three primary contributions to the field of construction safety and accident analysis. First, it proposes and validates a novel framework that integrates ChatGPT with CN techniques, offering an automated solution for extracting and analysis CFs from unstructured accident reports. Unlike Zhang *et al.*^[12], who identified critical CFs and causal paths of tower crane accidents in China using CN techniques but relied on manual extraction from narrative texts, our approach significantly reduces manual effort by leveraging ChatGPT for automated CF extraction. Second, this study introduces a set of tailored evaluation metrics and incorporates expert validation to systematically assess ChatGPT’s performance. This provides a methodological foundation for future applications of LLMs in accident analysis. Compared to Prieto *et al.*^[8] who explored the use of ChatGPT for construction project scheduling and reported strong performance in terms of efficiency and clarity, our study addresses the overgeneralization of evaluation metrics.

Prieto's nine evaluation indicators were found to be overly redundant for capturing the specific nuances of construction accident analysis, whereas our streamlined and targeted metrics better reflect the task-specific requirements of CF extraction. Third, by applying CN techniques to the outputs generated by ChatGPT, this study provides a structured, network-based perspective to identify key CFs and causal paths. Unlike the work of Wang *et al.*^[35], who evaluated ChatGPT's outputs but did not pursue further structured analysis, our approach integrates CN modeling to systematically reveal the relationships and interdependencies among the extracted CFs. This integrated methodology delivers actionable insights for construction safety managers and practitioners aiming to enhance on-site safety, and demonstrates the practical potential of combining LLM-driven extraction with structured network analysis for accident causation studies.

This study has three primary limitations that should be addressed in future research. First, the scope of the analyzed accident reports was limited to crane-related incidents and did not include visual data. Although such accidents are relatively rare, they tend to be severe. For instance, only 95 crane accidents were recorded in Hong Kong between 2011 and 2020^[5]. This study focused exclusively on textual data, which continues to dominate construction accident documentation. However, incorporating a more diverse dataset, including visual elements such as accident scene images, could provide additional insights and enhance the robustness of the analysis. Second, the proposed causation analysis framework is not fully automated. While significant progress has been made in automating the extraction of CF using ChatGPT, researchers were still required to manually document outputs and construct the CN model. Moreover, human expertise remains indispensable for investigating accidents and compiling comprehensive reports, as the verification of causes and collection of on-site evidence are currently beyond the capabilities of generative AI. Third, this study employed an existing version of ChatGPT rather than a domain-specific LLM. Future research should explore fine-tuning LLMs with construction safety datasets and incorporating reinforcement learning with human feedback to improve the precision and reliability of CF extraction. Advancing these approaches may reduce the manual workload and enhance the applicability of LLMs in specialized industrial domains.

7. Conclusion and Future Work

This study proposed a novel framework for causation analysis of construction accident reports by integrating ChatGPT and CN techniques. A dataset of 95 crane-related accident reports from Hong Kong was utilized. ChatGPT was employed to automatically extract CFs, with expert evaluation used to refine and validate the outputs. A CN model was then constructed to explore the internal relationships among the extracted CFs. The experimental results revealed key CFs associated with crane accidents, including H2 (Carelessness), H3 (Operation error), C1 (Crane unbalanced), C3 (Machine failure), C11 (Parts of a crane fall), I2 (Object strike), I4 (Worker fall), I5 (Trapping), I8 (Collapse of crane), and I9 (Load drop). Critical causal paths were also identified, such as C9 (Broken/failed rope) → I9 (Load drop) → I2 (Object strike).

Future research should aim to expand the dataset to cover a wider variety of accident types and incorporating multimodal data, such as images, to enrich the analytical depth and robustness. Additionally, developing domain-specific LLMs through fine-tuning and reinforcement learning could further improve accuracy and reduce dependence on manual intervention. The application of LLMs in specialized industries such as construction is anticipated to become an emerging trend. Effectively leveraging these models in domains including education, management, and decision-making will be a critical focus for both future research and practical implementation.

Supplementary materials

The supplementary material for this article is available at: [Supplementary materials](#).

Declarations

Authors contribution

Wang Y: Data analysis and interpretation, manuscript drafting, article conception and design.

Chen J: Data acquisition, manuscript revision.

Xiao B, Mueller S, Guo J: Article conception and design, manuscript revision.

Conflicts of interest

The authors declare that there are no conflicts of interest.

Ethical approval

Not applicable.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

Some or all data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

Funding

None.

Copyright

© The Author(s) 2025.

References

1. Statista. Value Added to Gross Domestic Product by the Construction Industry in the United States from 2000 to 2022 [Internet]. Statista; 2023 [cited 2025 Apr 10]. Available from: <https://www.statista.com/statistics/785445/value-added-by-us-construction>
2. International Labour Organization. Safety and Health in the Construction Sector-Overcoming the Challenges [Internet]. Geneva: International Labour Organization; 2014 [cited 2025 Apr 10]. Available from: https://www.ilo.org/empent/Eventsandmeetings/WCMS_310993/lang-en/index.htm
3. Zhou Z, Li Q, Wu W. Developing a versatile subway construction incident database for safety management. J Constr Eng Manag. 2012;138:1169-1180. [DOI]
4. Dahm A. Hong Kong Tower Crane Collapse Kills 3 Injures 6 [Internet]. Power Progress; 2022 [cited 2025 May 11]. Available from: <https://www.powerprogress.com/news/hong-kong-tower-crane-collapse-kills-3-injures-6/8023152.article>
5. Chen J, Chi HL, Du Q, Wu P. Investigation of operational concerns of construction crane operators: An approach integrating factor clustering and prioritization. J Manage Eng. 2022;38(4):04022020. [DOI]
6. Kim T, Chi S. Accident case retrieval and analyses: Using natural language processing in the construction industry. J Constr Eng Manag. 2019;145(3):04019004. [DOI]
7. Zhong B, Pan X, Love PED, Ding L, Fang W. Deep learning and network analysis: Classifying and visualizing accident narratives in construction. Autom Constr. 2020;113:103089. [DOI]
8. Prieto SA, Mengiste ET, García De Soto B. Investigating the use of ChatGPT for the scheduling of construction projects. Buildings. 2023;13:857. [DOI]
9. Koubaa A, Boulila W, Ghouti L, Alzahem A, Latif S. Exploring ChatGPT capabilities and limitations: A critical review of the NLP game changer. Preprints [Preprint]. 2023 [cited 2025 Apr 10]. Available from: <https://doi.org/10.20944/preprints202303.0438.v1>
10. Herbold S, Hautli-Janisz A, Heuer U, Kikteva Z, Trautsch A. A large-scale comparison of human-written versus ChatGPT-generated essays. Sci Rep. 2023;13:1-11. [DOI]
11. Steiss J, Tate T, Graham S, Cruz J, Hebert M, Wang J, et al. Comparing the quality of human and ChatGPT feedback of students' writing. Learn Instr. 2024;91:101894. [DOI]
12. Zhang W, Xue N, Zhang J, Zhang X. Identification of critical causal factors and paths of tower-crane accidents in China through system thinking and complex networks. J Constr Eng Manag. 2021;147:04021174. [DOI]
13. Hinze J, Pedersen C, Fredley J. Identifying root causes of construction injuries. J Constr Eng Manag. 1998;124:67-71. [DOI]
14. Choi J, Gu B, Chin S, Lee JS. Machine learning predictive model based on national data for fatal accidents of construction workers. Autom Constr. 2020;110:102974. [DOI]
15. Morris A, Brace C, Reed S, Fagerlind H, Bjorkman K, Jaensch M, et al. The development of a European fatal accident database. Int J Crashworthiness. 2010;15:201-209. [DOI]
16. Zhou Z, Irizarry J, Zhou J. Development of a database exclusively for subway construction accidents and corresponding analyses. Tunn Undergr Space Technol. 2021;111:103852. [DOI]
17. Mannering FL, Shankar V, Bhat CR. Unobserved heterogeneity and the statistical analysis of highway accident data. Anal Methods Accid Res. 2016;11:1-16. [DOI]
18. Irumba R. Spatial analysis of construction accidents in Kampala, Uganda. Saf Sci. 2014;64:109-120. [DOI]
19. Zhang X, Zhang W, Jiang L, Zhao T. Identification of critical causes of tower-crane accidents through system thinking and case analysis. J Constr Eng Manag. 2020;146:04020071. [DOI]
20. Leveson N. A new accident model for engineering safer systems. Saf Sci. 2024;42:237-270. [DOI]
21. Hollnagel E. FRAM: The functional resonance analysis method: modelling complex socio-technical systems. 1st ed. Boca Raton: CRC

- Press;2017. [DOI]
22. Shappell SA, Wiegmann DA. The Human Factors Analysis and Classification System—HFACS [Internet]. Daytona Beach (FL): Embry-Riddle Aeronautical University Scholarly Commons; 2000 [cited 2025 Apr 10]. Available from: <https://commons.erau.edu/publication/737/>
 23. Wu C, Li X, Guo Y, Wang J, Ren Z, Wang M, et al. Natural language processing for smart construction: Current status and future directions. *Autom Constr.* 2022;134:104059. [DOI]
 24. Ding Y, Ma J, Luo X. Applications of natural language processing in construction. *Autom Constr.* 2022;136:104169. [DOI]
 25. Sun J, Lei K, Cao L, Zhong B, Wei Y, Li J, et al. Text visualization for construction document information management. *Autom Constr.* 2020;111:103048. [DOI]
 26. Hassan F, Le T. Automated requirements identification from construction contract documents using natural language processing. *J Leg Aff Dispute Resolut Eng Constr.* 2020;12(2):04520009. [DOI]
 27. Zhong B, He W, Huang Z, Love PED, Tang J, Luo H. A building regulation question answering system: A deep learning methodology. *Adv Eng Inf.* 2020;46:101195. [DOI]
 28. Fang W, Luo H, Xu S, Love PED, Lu Z, Ye C. Automated text classification of near-misses from safety reports: An improved deep learning approach. *Adv Eng Inf.* 2020;44:101060. [DOI]
 29. Xu N, Ma L, Wang L, Deng Y, Ni G. Extracting domain knowledge elements of construction safety management: Rule-based approach using Chinese natural language processing. *J Manage Eng.* 2021;37:04021001. [DOI]
 30. Zou Y, Kiviniemi A, Jones SW. Retrieving similar cases for construction project risk management using Natural Language Processing techniques. *Autom Constr.* 2017;80:66-76. [DOI]
 31. Menon S, Vondrick C. Visual classification via description from large language models. *ArXiv: 2210.07183 [Preprint]*. 2022 [cited 2025 Apr 10]. Available from: <https://doi.org/10.48550/arXiv.2210.07183>
 32. Wang L, Lyu C, Ji T, Zhang Z, Yu D, Shi S, et al. Document-level machine translation with large language models. *ArXiv: 2304.02210 [Preprint]*. 2023 [cited 2025 Apr 10]. Available from: <https://doi.org/https://doi.org/10.48550/arXiv.2304.02210>
 33. You H, Ye Y, Zhou T, Zhu Q, Du J. Robot-enabled construction assembly with automated sequence planning based on ChatGPT: RoboGPT. *ArXiv: 2304.11018 [Preprint]*. 2023 [cited 2025 Apr 10]. Available from: <https://doi.org/10.48550/arXiv.2304.11018>
 34. Xiao B, Wang Y, Zhang Y, Chen C, Darko A. Automated daily report generation from construction videos using ChatGPT and computer vision. *Autom Constr.* 2024;168:105874. [DOI]
 35. Wang Y, Chen J, Xiao B, Zhang Y, Chen Y, Li Q. Investigating the potential of ChatGPT in construction management: A study of interpreting construction crane-related accident reports. In: Li D, Zou PXW, Yuan J, Wang Q, Peng Y, editors. *Proceedings of the 28th International Symposium on Advancement of Construction Management and Real Estate (CRIOCM 2023)*. Singapore: Springer; 2024. p. 327-340. [DOI]
 36. Hu S, Fang Y, Moehler R. Estimating and visualizing the exposure to tower crane operation hazards on construction sites. *Saf Sci.* 2023;160:106044. [DOI]
 37. Park S, Seong S, Ahn Y, Kim H. Real-time fatigue evaluation using ecological momentary assessment and smartwatch data: An observational field study on construction workers. *J Manage Eng.* 2023;39:04023008. [DOI]
 38. Guo X, Du J, Pu Y, Liu Q, Wang Y, Li J. 40-year development of man-machine-environment system engineering from scientific papers. In: Long S, Dhillon BS, editors. *Man-Machine-Environment System Engineering: Proceedings of the 21st International Conference on MMESE*. Singapore: Springer; 2022. p. 21-28. [DOI]
 39. Lahat A, Shachar E, Avidan B, Shatz Z, Glicksberg BS, Klang E. Evaluating the use of large language model in identifying top research questions in gastroenterology. *Sci Rep.* 2023;13(1):4164. [DOI]
 40. Joshi A, Kale S, Chandel S, Pal DK. Likert scale: Explored and explained. *Br J Appl Sci Technol.* 2015;7:396-403. [DOI]
 41. Nooy WD, Mrvar A, Batagelj V. *Exploratory social network analysis with Pajek*. 3rd ed. Cambridge: Cambridge university press;2018. [DOI]
 42. Zhou J, Xu W, Guo X, Ding J. A method for modeling and analysis of directed weighted accident causation network (DWACN). *Physica A.* 2015;437:263-277. [DOI]
 43. Zhou Z, Irizarry J, Li Q. Using network theory to explore the complexity of subway construction accident network (SCAN) for promoting safety management. *Saf Sci.* 2014;64:127-136. [DOI]
 44. Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang DU. Complex networks: Structure and dynamics. *Phys Rep.* 2006;424:175-308. [DOI]
 45. Ji Z, Lee N, Frieske R, Yu T, Su D, Xu Y, et al. Survey of hallucination in natural language generation. *ACM Comput Surv.* 2023;55(12):1-38. [DOI]
 46. Huang L, Yu W, Ma W, Zhong W, Feng Z, Wang H, et al. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Trans Inf Syst.* 2025;43(2):1-55. [DOI]
 47. Augenstein I, Baldwin T, Cha M, Chakraborty T, Ciampaglia GL, Corney D, et al. Factuality challenges in the era of large language models and opportunities for fact-checking. *Nat. Intell.* 2024;6(8):852-863. [DOI]

Publisher's Note

Science Exploration remains a neutral stance on jurisdictional claims in published maps and institutional affiliations. The views expressed in this article are solely those of the author(s) and do not reflect the opinions of the Editors or the publisher.