buildings

MDPI

Article

Explainable Emergency Response Recommendation for Foundation Pit Collapse Accidents: An LLM-Enabled E/B/R Ontology-Driven Hybrid Reasoning Framework

Tengyao Wang and Jue Li *

School of Transportation, Changsha University of Science and Technology, Changsha 410114, China
* Correspondence: lij93@csust.edu.cn

Abstract

Foundation pit collapse accidents require rapid identification of accident conditions and appropriate emergency measures. Accident investigation reports contain valuable response knowledge, but their unstructured format limits direct reuse. This study develops an LLM-enabled E/B/R ontology-driven hybrid reasoning framework for explainable emergency response recommendations. Accident knowledge is organized as Event (E), Bearing (B), and Response (R). The resulting ontology contains 51 subclasses, and 69 foundation pit collapse reports were retained from 1110 collected construction collapse reports to construct the case base. For a newly imported report, Qwen3.6-Plus extracts candidate E/B features within the ontology label space, while code validation and filtering remove invalid labels. The validated features are processed separately by rule-based reasoning (RBR) and case-based reasoning (CBR), and their responses are fused with source labels. On a 39-sample comparison set, Qwen3.6-Plus achieved a micro-F1 score of 0.821 for E/B extraction. Applications to three target cases showed that RBR provided rule-supported measures, whereas CBR contributed additional responses from related historical cases. A questionnaire assessment of the earlier prototype outputs, based on 19 valid responses, yielded an overall mean score of 5.97 on a seven-point scale. The framework provides a traceable means of reusing accident-report knowledge for construction emergency decision support while restricting the LLM to constrained extraction and result explanation.

Keywords: foundation pit collapse; construction safety; emergency response recommendation; E/B/R ontology; large language model; rule-based reasoning; case-based reasoning; hybrid reasoning

Check for updates

Academic Editor: Pramen P. Shrestha

Received: 28 August 2026
Revised: 9 October 2026
Accepted: 10 October 2026
Published: 13 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Foundation pit collapse accidents pose a serious risk to construction projects because they may occur suddenly and result in casualties, secondary collapse, damage to surrounding structures, traffic disruption, and other social impacts. Their occurrence is commonly associated with interacting geological, hydrological, structural, construction, and management factors, including weak soil, rainfall or seepage, insufficient support, inappropriate excavation, monitoring failure, and unsafe construction practices [1–4]. Once a collapse occurs, emergency personnel must rapidly identify the principal accident features and determine appropriate measures, such as emergency reporting, evacuation, on-site rescue, temporary reinforcement, professional rescue, medical assistance, and command coordination. Effective post-accident response therefore depends on both timely situation

identification and the appropriate reuse of established safety knowledge and historical accident experience.

Accident investigation reports contain valuable information about accident processes, site conditions, causal deficiencies, emergency actions, and lessons learned. Text mining and natural language processing have been applied to these reports to identify accident causes, risk factors, and construction safety knowledge [5,6]. However, the extracted information is generally used for accident classification, causal analysis, or risk prevention and is rarely transformed into structured knowledge for emergency response recommendations. Moreover, foundation pit collapse is a specialized accident type, and publicly available reports containing sufficient process, cause, and response information are limited. These conditions make it difficult to rely solely on large-scale supervised learning or purely data-driven prediction. A knowledge-driven approach that combines structured representation, explicit rules, and historical case reuse is therefore suitable for this data-limited construction safety problem.

Ontology-based modeling provides a formal means of organizing fragmented safety knowledge by explicitly defining domain concepts and their relationships [7]. Construction safety ontologies have been applied to automated safety checking, safety knowledge modeling, and risk knowledge management [8,9]. Case-based reasoning (CBR), meanwhile, supports decision-making by retrieving and reusing experience from similar previous cases [10] and has been applied to construction accident pre-control and safety risk management [11,12]. However, historical cases must first be represented through a consistent feature structure before they can be compared and reused. Rule-based reasoning (RBR) can complement CBR by encoding explicit relationships between identified accident features and predefined response measures. Integrating ontology representation, RBR, and CBR therefore provides a potential means of combining rule-supported safety knowledge with experience recorded in historical accidents.

Large language models (LLMs) provide new capabilities for processing unstructured accident reports, including information extraction, summarization, and natural-language explanation. Nevertheless, directly using an LLM to generate emergency responses may produce unsupported or insufficiently grounded outputs in safety-critical contexts [13–15]. Knowledge-graph- and ontology-enhanced approaches have consequently been proposed to improve the grounding and controllability of LLM applications [13,14]. In emergency response recommendation, the role of the LLM should therefore be constrained: it can identify candidate accident features within a predefined ontology and explain the results generated by explicit reasoning, but it should not independently determine safety-critical response measures.

Recent studies have explored structured knowledge and intelligent reasoning in different engineering applications. Knowledge graphs have been used to improve the usability of risk information in construction safety planning [16]; domain knowledge has been integrated with multimodal LLMs for construction safety inspection [17]; ontology-guided LLM extraction has supported the construction of task-oriented knowledge graphs from engineering documents [18]; and hybrid CBR–RBR reasoning has been applied to parametric bridge BIM modeling [19]. These studies provide important foundations for knowledge-driven engineering decision support. However, limited research has connected accident-report processing, ontology-constrained LLM extraction, explicit rule reasoning, historical case retrieval, response fusion, and explanation generation within a unified framework for post-accident emergency response in foundation pit collapse scenarios.

To address this gap, this study proposes an LLM-enabled E/B/R ontology-driven hybrid reasoning framework for emergency response recommendation in foundation pit collapse accidents. Based on the public safety triangle theory, accident knowledge is

organized into three dimensions: Event (E), representing the accident event and directly observable triggering or site conditions; Bearing (B), representing deficiencies or weak links in the engineering system; and Response (R), representing corresponding emergency measures. In the recommendation process, an LLM extracts candidate E/B features within the predefined ontology label space. After validation and normalization, these features are processed separately by RBR and CBR. RBR generates rule-supported responses, while CBR retrieves related historical cases and provides case-supported response measures. The two outputs are fused while retaining the source of each response, and the LLM subsequently explains the reasoning results without modifying the generated response categories.

The principal contributions of this study to construction safety and emergency decision support are as follows:

(1) An E/B/R ontology is developed to organize foundation pit collapse knowledge by connecting accident events, engineering-system deficiencies, and emergency responses within a shared semantic structure.
(2) An LLM-constrained extraction process is established to transform unstructured accident reports into validated E/B features for subsequent knowledge-driven reasoning, without assigning the LLM responsibility for directly generating safety-critical responses.
(3) An RBR–CBR hybrid reasoning mechanism is implemented to combine responses supported by predefined rules and related historical cases and to preserve the reasoning source of each response.
(4) A prototype decision-support system is implemented and evaluated through feature-extraction tests, separate RBR and CBR outputs, fused recommendations for three target cases, and questionnaire-based evaluation, providing evidence of the framework's preliminary applicability.

The remainder of this paper is organized as follows. Section 2 reviews knowledge-driven construction safety, emergency decision support, LLM-constrained information extraction, and hybrid reasoning. Section 3 describes the E/B/R ontology, accident-feature extraction and normalization, RBR–CBR reasoning, response fusion, and prototype implementation. Section 4 presents the framework application and validation results. Section 5 discusses the implications for construction safety and emergency decision support, together with the practical applications, limitations, and transferability of the framework. Section 6 presents the conclusions.

2. Related Work

2.1. Knowledge-Driven Construction Safety and Emergency Decision Support

Foundation pit and underground construction collapse accidents are commonly associated with interacting geological, hydrological, structural, construction, and management factors. Previous studies have investigated these accidents through forensic analysis, risk-factor identification, Bayesian networks, catastrophe or evidence-theory-based assessment, and knowledge-graph-based hazard coupling analysis [1–4,20]. These studies have improved the understanding of collapse mechanisms and risk propagation. However, they have primarily focused on pre-accident risk assessment, hazard identification, and causal analysis. Comparatively less attention has been given to transforming accident features into structured emergency response recommendations after a collapse has occurred.

Accident investigation reports provide an important source of construction safety knowledge because they document accident processes, site conditions, causal deficiencies, emergency actions, and lessons learned. Earlier studies applied machine-learning methods to classify accident narratives, demonstrating the potential of these reports for safety knowledge discovery [21]. Subsequent studies have used text mining, natural language

processing, deep learning, and network analysis to classify accident causes, extract risk factors, and identify relationships among contributing factors [5,6,22–25]. Although these approaches convert unstructured narratives into analyzable information, their outputs generally remain focused on classification or causal analysis and are rarely connected directly with emergency response generation.

Ontology-based knowledge modeling provides a formal means of organizing fragmented construction safety information. Ontologies explicitly define domain concepts and relationships and thereby support knowledge sharing, reuse, and reasoning [7]. In construction safety, ontology-based approaches have been applied to automated safety checking, risk knowledge management, and safety information organization [8,9,26–29]. Linked data, knowledge graphs, and ontological information systems have also been used to support safety planning, information sharing, and accident knowledge reuse [16,30–32]. Nevertheless, most existing applications focus on hazard identification, safety checking, risk management, or accident causation. Response-oriented knowledge representations that explicitly connect accident events, Bearing (B) deficiencies, and emergency measures for foundation pit collapse accidents remain limited.

Case-based reasoning (CBR) and rule-based reasoning (RBR) provide two different mechanisms for reusing structured safety knowledge. CBR addresses a new problem by retrieving and reusing experience from similar historical cases [10] and has been applied to accident pre-control, safety decision-making, and automated construction safety risk management [11,12]. Its effectiveness, however, depends on structured case representation and interpretable similarity assessment. In foundation pit collapse accidents, related cases may share specific E/B features, such as slope instability, seepage, insufficient support, personnel entrapment, or continued collapse risk, even when their complete accident descriptions are different.

RBR can encode explicit relationships between accident features and corresponding emergency measures. SWRL provides a rule representation mechanism for ontology-based inference [33], allowing predefined safety knowledge to be expressed as traceable reasoning rules. RBR is suitable for generating rule-supported baseline measures, whereas CBR can supplement these measures with responses recorded in related historical accidents. Hybrid RBR–CBR approaches have been explored in other engineering applications [19], but their combined use for foundation pit collapse emergency response recommendation remains limited.

Overall, existing research provides important foundations in collapse-risk analysis, accident-report mining, ontology-based safety knowledge representation, and rule- and case-based reasoning. However, three connected gaps remain. First, accident-report analysis generally stops at cause classification or risk-factor extraction rather than using the extracted information to recommend emergency responses. Second, existing construction safety ontologies mainly support risk prevention and knowledge management rather than response-oriented reasoning. Third, RBR and CBR are commonly studied separately, leaving limited support for combining rule-based safety knowledge with historical response experience. These gaps motivate the proposed E/B/R ontology-driven hybrid reasoning framework for explainable emergency response recommendation.

2.2. LLM-Constrained Information Extraction and Hybrid Reasoning

Large language models (LLMs) have demonstrated capabilities in processing unstructured text, including information extraction, summarization, question answering, and natural-language explanation. These capabilities are relevant to the architecture, engineering, and construction (AEC) sector because accident reports, safety documents, emergency plans, and historical case records are commonly stored as long and unstructured texts.

Reviews of LLM applications in AEC have identified opportunities in document processing, information retrieval, knowledge interaction, and decision assistance, while also highlighting challenges related to hallucination, domain adaptation, data privacy, and output reliability [34,35].

Recent studies have examined the integration of LLMs with structured engineering knowledge. Domain knowledge has been combined with multimodal LLMs to support construction safety inspection [17], while ontology-guided LLM extraction has been used to construct task-oriented knowledge graphs from engineering documents [18]. These studies indicate that explicit domain knowledge can help define the scope and expected structure of LLM outputs. For construction accident reports, this creates an opportunity to use LLMs as an interface between unstructured narratives and ontology-based safety knowledge.

The use of LLMs in emergency decision support nevertheless requires particular caution. Directly generated emergency recommendations may contain unsupported information because the model output is not necessarily grounded in verified accident knowledge [15]. Knowledge graph–LLM emergency decision frameworks have therefore been proposed to combine language processing capabilities with structured emergency knowledge [13]. More broadly, knowledge graph-enhanced LLM research suggests that external structured knowledge can improve factual grounding and reduce unsupported generation [14]. These findings support limiting the role of LLMs in safety-critical applications and retaining explicit knowledge-based mechanisms for the core reasoning process.

Ontology-constrained extraction provides one means of implementing this division of responsibilities. Instead of asking an LLM to generate emergency responses directly, the model can be restricted to identifying candidate features within a predefined ontology label space. Structured output requirements and subsequent code validation can then be used to remove invalid labels before the extracted information is passed to the reasoning modules. The LLM can also explain the reasoning results after the response categories have been generated, provided that it does not alter those categories. This arrangement allows the language model to support text processing and result interpretation while keeping the response-generation process traceable to explicit rules and historical cases.

Existing studies have advanced the integration of LLMs with ontologies and knowledge graphs, but limited attention has been given to their use within a complete response-recommendation process for a specific type of construction accident. Current construction safety applications commonly focus on inspection, document processing, knowledge extraction, or question answering. Few studies connect ontology-constrained extraction from accident reports with SWRL-based RBR, historical case retrieval, response fusion, and explanation of the resulting recommendations.

Accordingly, this study positions the LLM as a constrained E/B extraction component and a subsequent explanation component rather than an autonomous emergency decision maker. The core responses are generated by an E/B/R ontology-driven RBR–CBR process: RBR provides explicit rule-supported measures, while CBR contributes measures recorded in related historical cases. This division of roles connects unstructured accident-report information with traceable and explainable emergency response recommendations.

3. Materials and Methods

3.1. Framework Overview

This study proposes an LLM-enabled E/B/R ontology-driven hybrid reasoning framework for emergency response recommendation in foundation pit collapse accidents. The framework integrates ontology-based knowledge representation, LLM-constrained extraction, rule-based reasoning (RBR), case-based reasoning (CBR), response fusion, and LLM-based explanation. The LLM is used to extract candidate E/B features within the pre-

defined ontology label space and to explain the final reasoning results. The core emergency response recommendations are generated by the knowledge-driven RBR and CBR modules rather than directly by the LLM.

As shown in Figure 1, the framework consists of two stages: offline knowledge construction and online emergency response recommendation. In the offline stage, publicly available accident investigation reports are collected and screened to construct the foundation pit collapse accident case base. Based on the public safety triangle theory, the accident information is organized into Event (E), Bearing (B), and Response (R) dimensions. The extracted information is manually reviewed, normalized into ontology concepts, and organized by accident case. The resulting E/B/R ontology provides the semantic basis for constructing the SWRL rule base and the structured historical case base.

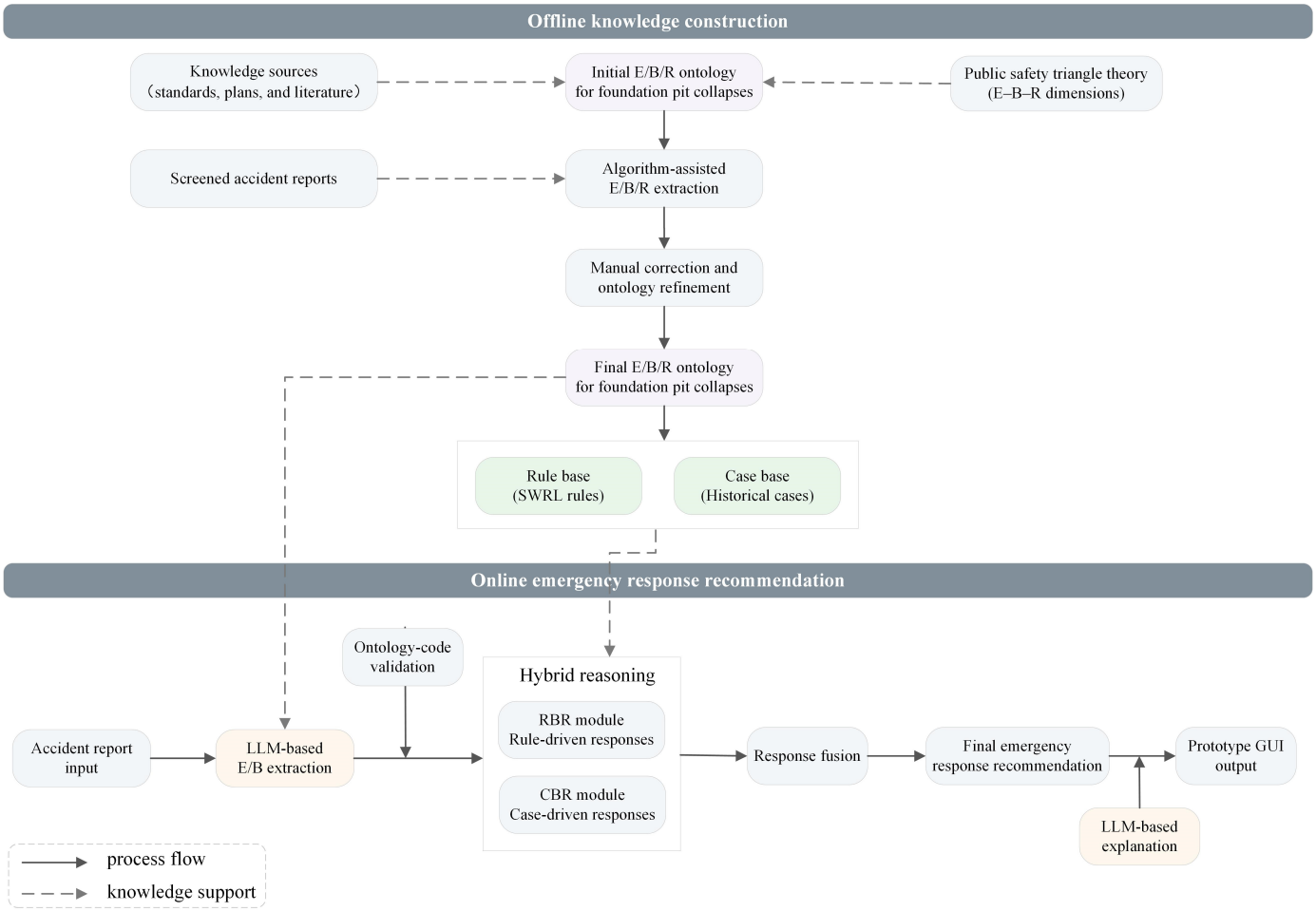


Figure 1. Overall workflow of the proposed LLM-enabled E/B/R ontology-driven hybrid reasoning framework.

In the online stage, a newly imported foundation pit collapse accident report is processed as the target case. Using the predefined ontology label space and a fixed prompt, the LLM extracts candidate E/B features from the report. The extracted labels are then processed through JSON parsing, ontology-code validation, invalid-label filtering, duplicate removal, and label-name mapping. Only validated E/B features are passed to the reasoning modules. Response (R) labels are not generated during this extraction stage because they are the outputs of the subsequent reasoning process.

The validated E/B features are provided separately to the RBR and CBR modules. RBR infers rule-supported responses using predefined SWRL rules, whereas CBR retrieves related historical cases and extracts their recorded response measures. The two outputs are

then fused while retaining their reasoning sources. Responses generated by both modules are identified as consensus responses, while responses generated by only one module are retained as RBR-exclusive or CBR-exclusive responses.

Finally, the fused responses and their supporting evidence are passed to the LLM-based explanation module. The LLM explains the relationships among the extracted E/B features, reasoning evidence, and recommended responses but does not add, remove, or modify the response categories. The results are presented through the prototype graphical user interface, completing the workflow from accident-report input to structured feature extraction, dual-source reasoning, response fusion, and explanation output.

3.2. Public Safety Triangle-Based Ontology Representation

To provide a structured knowledge basis for emergency response recommendation, an E/B/R ontology for foundation pit collapse accidents was developed based on the public safety triangle theory. Ontologies provide explicit representations of domain concepts and their relationships and support knowledge sharing, reuse, and reasoning [7]. The public safety triangle links emergencies, disaster-bearing entities, and emergency management [36]. In this study, these dimensions are adapted into Event (E), Bearing (B), and Response (R) for the specific task of emergency response recommendation. B describes deficiencies associated with the affected engineering and organizational system, while R represents response measures within the broader scope of emergency management.

E, B, and R are used as a decision-oriented representation of emergency scenarios rather than as a strict causal chain. As summarized in Table 1, Event (E) represents the accident event and its directly observable triggering or site conditions. Bearing (B) represents deficiencies or weak links in the engineering system that reduce its capacity to withstand or control the accident. Response (R) represents emergency measures used to prevent further risk expansion and support rescue operations. This structure was selected because a response recommendation requires information about what is happening, which system weaknesses may hinder risk control, and which measures may address the situation. Separating these dimensions allows E/B features to serve as inputs to rule matching and case retrieval, while R categories form the recommendation outputs.

Table 1. E/B/R representation adapted from the public safety triangle theory.

Dimension	Definition in this Study	Role in the Proposed Framework
Event (E)	Accident events and directly observable triggering or site conditions	Provides scenario features for structured extraction and reasoning
Bearing (B)	Deficiencies or weak links in the engineering system's capacity to withstand or control the accident	Represents system vulnerabilities used by RBR and CBR
Response (R)	Emergency measures used to control risk expansion and support rescue operations	Provides the response categories generated by RBR and CBR

The E/B/R structure is related to established accident analysis frameworks through its representation of accident conditions and system deficiencies. Human error perspectives examine unsafe acts together with the working and organizational conditions that influence human performance. The Swiss cheese model explains how active failures and latent conditions can combine to undermine multiple layers of defence [37]. These perspectives help interpret the behavioral and organizational deficiencies represented within B, including unsafe operations, management deficiencies, and monitoring failures. B also includes physical deficiencies, such as inadequate support. Thus, E/B/R retains information relevant to

human and system failures, while organizing it for response recommendation rather than reconstructing the full sequence of defensive-layer failures.

Root cause analysis identifies underlying causes and supports corrective actions to prevent recurrence [38]. Its findings can inform the conditions and deficiencies recorded in E/B, but E/B labels do not themselves constitute a root cause analysis. The purpose of the present framework is to link these reported conditions and deficiencies to emergency response categories through rules and historical cases. The relationship is therefore complementary: established accident analysis approaches help explain failures, while E/B/R provides a common structure for organizing relevant findings and response knowledge for rule matching and case retrieval.

The ontology was constructed through a combined top-down design and bottom-up refinement process. In the top-down stage, the initial E/B/R structure was developed based on the public safety triangle theory and informed by foundation pit engineering standards and emergency management guidelines [39,40]. The main categories and their relationships were defined according to the information required to describe foundation pit collapse scenarios and corresponding emergency responses.

In the bottom-up stage, the screened accident investigation reports were processed to identify E/B/R-related information. The extracted expressions were manually reviewed and mapped to the initial ontology. Unmatched concepts, unclear semantic boundaries, and missing categories were then used to refine or supplement the ontology. This iterative process produced the final E/B/R ontology for foundation pit collapse accidents.

The resulting ontology provides the shared semantic foundation for the proposed framework. It defines the controlled label space used for LLM-constrained E/B extraction, supports the construction of SWRL rules in the RBR module, and provides the feature structure used to organize historical cases in the CBR module. In this way, the ontology connects accident-report processing, structured feature extraction, rule-based reasoning, case-based reasoning, and emergency response recommendation within a unified knowledge representation. The complete ontology structure, SWRL rule set, and ontology-constrained LLM prompt template are provided in Appendixes A–C, respectively.

3.3. Accident Feature Extraction and Normalization

Accident feature extraction was used to transform unstructured accident reports into structured E/B/R features. The process consisted of offline extraction for historical case-base construction and online extraction for emergency response recommendation. The offline process produced calibrated E/B/R labels for historical cases, whereas the online process extracted candidate E/B labels from a newly imported accident report as inputs to RBR and CBR. Table 2 summarizes the extraction and normalization process.

Table 2. Summary of accident feature extraction and normalization.

Part	Main Operation	Output
Offline extraction	Extract and manually calibrate E/B/R-related information from standardized accident reports	Calibrated E/B/R labels
Case-base construction	Organize the calibrated labels by accident case	Structured case–feature matrix
Online extraction	Extract and validate candidate E/B labels using a fixed prompt and ontology constraints	Validated E/B labels for RBR and CBR

In the offline process, the screened accident reports were first converted into a unified report structure. Because E/B/R-related information was mainly contained in the accident process and accident-cause sections, sentence extraction focused on paragraphs describing

the accident process, direct causes, indirect causes, and accident nature. Candidate sentences were identified using keyword matching, rule-based pattern matching, and semantic similarity matching.

For semantic matching, the sentence-transformer model paraphrase-multilingual-MiniLM-L12-v2 was used to calculate the similarity between an accident-report sentence and the definition of each ontology category:

$$Sim(s, c_i) = \cos(v_s, v_{c_i}) \quad (1)$$

where v_s and v_{c_i} denote the vector representations of sentence s and ontology category c_i, respectively. A similarity threshold of 0.60 was used, with category-specific keywords serving as supplementary matching cues. This threshold was applied during candidate screening, after which the extracted sentences were manually reviewed and calibrated.

The candidate sentences were refined into concise feature expressions and mapped to the corresponding E/B/R ontology categories. Mismatched labels, unclear semantic boundaries, and unmatched expressions were manually reviewed and used to revise or supplement the ontology. When one expression contained multiple accident features, multiple ontology labels were assigned. To examine the reliability of the manual calibration, a subset of the labeled sentences was rechecked by the same annotator after an interval. The corresponding consistency results are reported in Section 4.2.

The calibrated E/B/R labels were then organized by accident case to construct the structured case–feature matrix used by the CBR module. They also provided reference labels for prompt refinement and comparison of the candidate LLMs. Ten labeled samples were used during prompt refinement and were excluded from the formal model comparison. The fixed prompt was subsequently evaluated using an independent set of 39 samples, as described in Section 4.2.

In the online recommendation process, the selected LLM was used only as an ontology-constrained feature extraction component. After an accident report was imported, the model extracted candidate E/B labels using the predefined ontology label space and fixed prompt. Response labels were not extracted by the LLM because Response (R) was the target output generated by RBR and CBR.

The LLM output was required to follow a structured JSON format. It was subsequently processed through JSON parsing, ontology-code validation, invalid-label filtering, duplicate removal, and label-name mapping. Only labels corresponding to valid E/B ontology codes were retained and passed to the reasoning modules. The final prompt template used for LLM-constrained E/B extraction is provided in Appendix C.

3.4. Hybrid Reasoning and Emergency Response Recommendation

After the E/B features of a target accident are validated, they are provided separately to the RBR and CBR modules. RBR generates responses according to predefined emergency rules, whereas CBR retrieves related historical cases and provides responses recorded in those cases. The two modules therefore produce an RBR configuration output and a CBR configuration output from the same validated E/B input. Their outputs are subsequently fused to form the final recommendation while preserving the source of each response.

In the RBR module, SWRL rules were constructed to formalize emergency responses under typical foundation pit collapse conditions. The ontology and SWRL rules were implemented in Protégé. HermiT was used to check the consistency of ontology concepts and relationships, while Drools was used through SWRLTab to execute the rules and write the inferred results back to the ontology. Each rule uses one or more E/B features as its triggering conditions and infers a corresponding R category. The general rule form is expressed as follows:

$$E_i \wedge B_j \rightarrow R_k \tag{2}$$

where E_i denotes an Event feature, B_j denotes a Bearing-deficiency feature, and R_k denotes the inferred Response. A rule may also be triggered by a single critical E or B feature when the corresponding response is required. The rules primarily cover basic measures such as emergency reporting, on-site rescue, evacuation, initial reinforcement, professional rescue, and command coordination. Accordingly, the RBR output represents responses explicitly supported by the predefined rule base.

In the CBR module, validated E/B features are used to retrieve related historical cases represented as structured E/B/R feature matrices in Neo4j. Retrieved cases need not be identical to the target accident; instead, their similarity is evaluated through ontology-based feature matching, and their historical R labels provide case-supported response measures.

Case similarity is calculated using a hierarchy-aware Dice measure for the Event and Bearing dimensions separately. Within each dimension, exact label matches are paired first and assigned a weight of 1. The remaining labels are then paired within the same parent category and assigned a weight of 0.5. Each label can participate in only one pair. Unrelated labels do not contribute to the matching score.

$$S_D(T_D, H_D) = \frac{2\left(m_D + 0.5_{p_D}\right)}{|T_D| + |H_D|}, D \in \{E, B\} \tag{3}$$

Here, T_D and H_D denote the distinct labels of the target and historical cases in dimension D, respectively. m_D is the number of exact pairs, and p_D is the number of remaining pairs sharing the same parent category. For each parent category, the number of partial pairs is the smaller of the two remaining label counts. The denominator accounts for the label sets of both cases, so additional unmatched historical features reduce similarity rather than automatically increasing the score.

When the target case contains both E and B features, the overall similarity is as follows:

$$S(T, H) = \frac{S_E(T_E, H_E) + S_B(T_B, H_B)}{2} \tag{4}$$

When only one target feature dimension is available, its similarity is used as the overall score. If the historical case lacks a dimension present in the target case, that component receives a score of zero. R labels do not participate in similarity calculation. A historical case is eligible only if it contains at least one exact E or B match with the target case. Eligible cases are ranked by overall similarity, with case identifiers used to resolve ties consistently. The two highest-ranked cases are retained, and their historical R labels are collected for subsequent response fusion.

Foundation pit-specific conditions influence recommendation through both rule activation and case retrieval. The E/B representation captures reported conditions related to collapse mechanisms, including slope instability, support deformation, and seepage, together with deficiencies such as absent or insufficient support. In Appendix B, seepage combined with absent support activates RBR-11, slope instability combined with absent support activates RBR-12, and support deformation combined with insufficient support activates RBR-13. These combinations all lead to R1_3 InitialReinforcement, but retain different triggering conditions. Their shared output reflects the category-level resolution of the framework; it does not imply that the corresponding stabilization techniques are interchangeable.

Site conditions also affect the recommended response categories. Continued collapse risk activates the evacuation rule RBR-04. Personnel entrapment combined with a haz-

ardous rescue environment activates rules for professional command and multi-department coordination, while the additional presence of a constrained site activates the resource-allocation rule RBR-07. In CBR, the same E/B conditions contribute to feature similarity and therefore influence which historical cases and associated responses are retrieved. The contribution of site information consequently depends on whether it is documented and represented in the available features.

The framework thus uses mechanism-related and site-condition information to support the selection of response categories. It does not calculate detailed geotechnical failure processes, design stabilization works, or determine site-specific rescue procedures. These requirements extend beyond the information represented by the current ontology and require additional site assessment and professional judgment.

The RBR and CBR outputs are fused by retaining responses from both reasoning sources and recording their source composition. A response generated by both modules is classified as a consensus response. A response generated only by RBR is classified as an RBR-exclusive response and retained as a rule-supported baseline measure. A response generated only by CBR is classified as a CBR-exclusive response and retained as a case-supported supplementary measure. Table 3 summarizes the fusion strategy.

Table 3. Fusion strategy for rule-driven and case-driven responses.

Response Category	Definition	Role in the Fused Recommendation
Consensus	Generated by both RBR and CBR	Response supported by both predefined rules and historical cases
RBR-exclusive	Generated only by RBR	Rule-supported baseline response
CBR-exclusive	Generated only by CBR	Case-supported supplementary response

The fusion process does not discard an RBR response because it is absent from the CBR output. Similarly, a CBR-exclusive response is retained as a supplementary recommendation rather than treated as a mandatory rule-based measure. This strategy preserves the contributions of both reasoning sources and makes the basis of each recommended response traceable.

The fused responses are then provided to the LLM-based explanation module. The LLM does not add, remove, or modify the response categories generated by RBR and CBR. It generates an explanation based on the validated E/B features, the RBR output, the retrieved cases and their responses, and the fusion results. The explanation describes the relationship between the identified accident features and the recommended measures and provides supporting information for users to interpret the reasoning results.

Therefore, the final recommendation is generated through explicit rule- and case-based reasoning rather than direct LLM decision-making. The ontology, SWRL rules, and historical case base can be manually updated when new accident cases, revised ontology categories, or validated emergency rules become available, without changing the overall reasoning process.

3.5. Prototype Implementation

A prototype decision-support system was developed to demonstrate the implementation of the proposed E/B/R ontology-driven hybrid reasoning framework. Following the online workflow shown in Figure 1, the system imports a target accident report, extracts E/B features, executes RBR and CBR, fuses the two response outputs, and presents the final recommendations and their explanations through a graphical user interface. The main components of the prototype are summarized in Table 4.

Table 4. Main components of the prototype system.

Component	Implementation	Main Function
LLM component	Qwen3.6-Plus API, fixed prompt, and JSON Mode	Ontology-constrained E/B extraction and explanation of reasoning results
RBR engine	Protégé, OWL, SWRL, and Drools	Rule-driven emergency response generation
CBR engine	Neo4j case base	Historical case retrieval and case-supported response extraction
Response fusion module	Rule–case response integration	Source identification, duplicate removal, and fusion of RBR and CBR outputs
Prototype GUI	Graphical user interface	Accident-report import and presentation of extraction and recommendation results

The LLM component was implemented using Qwen3.6-Plus, with the model ID qwen3.6-plus, accessed through the Alibaba Cloud Model Studio API. Based on the E/B extraction comparison reported in Section 4.2, this model was adopted for the final prototype. It was operated in non-thinking mode with the temperature set to 0, a fixed prompt, and JSON Mode. The model output was processed through API retry handling, JSON parsing, ontology-code validation, invalid-label filtering, duplicate removal, keyword-based post-processing, and label-name fallback mapping. These procedures ensured that only valid ontology-defined E/B labels were provided to the reasoning modules.

The RBR engine was implemented using the ontology and SWRL rule base developed in Protégé 5.6.7 (Stanford University, Stanford, CA, USA). The ontology and rules were stored in Web Ontology Language (OWL) format, and Drools was used through SWRLTab to execute the rules and generate rule-supported responses. The CBR engine was implemented in Neo4j Community Edition 5.15.0 (Neo4j, Inc., San Mateo, CA, USA), where historical accident cases and their E/B/R relationships were stored to support feature-based case retrieval and reuse of historical responses. Protégé therefore supports ontology representation and rule reasoning, whereas Neo4j supports structured case storage and retrieval.

The response fusion module processes the outputs of the RBR and CBR engines according to the strategy described in Section 3.4. It removes duplicate responses and identifies consensus, RBR-exclusive, and CBR-exclusive responses. Consensus responses are supported by both reasoning modules, while exclusive responses retain their respective rule-based or case-based sources. The fused response list and its reasoning evidence are then passed to the LLM-based explanation module.

The LLM does not modify the response categories produced by RBR and CBR. It generates explanations using the validated E/B features, rule-supported responses, retrieved cases, case-supported responses, and fusion results. The prototype interface displays the imported accident report, extracted E/B features, RBR and CBR outputs, fused recommendations, response sources, and explanation results. The ontology, SWRL rules, and historical case base can be manually updated when validated knowledge or new accident cases become available.

Recommendation traceability is maintained through the structured outputs of the two reasoning modules. For RBR, the supporting evidence consists of the extracted E/B conditions and the corresponding rule that produces an R category; the rule antecedents and conclusions are provided in Appendix B. For CBR, the retrieval results retain the historical case identifiers, similarity scores, matched E/B features, and recorded R categories, allowing the case-based responses to be traced to their historical sources. After fusion, each response retains a source label indicating RBR, CBR, or support from both modules. The interface presents the rule-based outputs, related cases, matching evidence,

and response-source labels for inspection. These structured records provide the basis for tracing recommendations, while the LLM-generated narrative helps users interpret the results.

4. Results

4.1. Case-Base Construction and Feature Extraction Results

A foundation pit collapse accident case base was constructed from publicly available construction accident investigation reports. The original reports were collected through manual downloading and web crawling, and files in different formats were converted into a unified DOCX format for subsequent text processing. To identify foundation pit collapse accidents, latent Dirichlet allocation (LDA)-based topic classification was first applied to the collected reports, and the results were then corrected using category-specific keywords and manual screening. Foundation pit-related reports were identified using terms related to excavation, retaining structures, support systems, anchors, cut-off walls, dewatering wells, and enclosure structures. After removing irrelevant, duplicated, and incomplete reports, 69 reports were retained for E/B/R feature extraction and case-base construction. The main screening results are summarized in Table 5.

Table 5. Summary of accident-report collection and screening results.

Item	Result
Initially collected construction collapse accident reports	1110
Successfully converted reports into unified DOCX format	1107
Candidate foundation pit collapse reports identified by automatic classification	314
Final retained foundation pit collapse reports after manual screening	69

For the retained reports, sentence-level feature extraction was conducted on standardized accident-report sections. Since E/B/R-related information was mainly concentrated in the accident process and accident-cause sections, extraction focused on paragraphs describing the accident process, direct causes, indirect causes, and accident nature. The initial algorithm-assisted extraction identified candidate E-, B-, and R-related sentences, which were then manually refined into concise feature expressions. These expressions were matched with the initial E/B/R ontology and manually calibrated to obtain the final labels.

During calibration, the ontology was refined when extracted expressions could not be directly mapped to existing categories or when the semantic boundaries of categories were unclear. For example, repeated descriptions of soft or unstable soil led to the inclusion of E2_4 SoftSoil; descriptions of on-site personnel organizing immediate rescue were represented by R1_4 OnSiteOrganizedRescue; mechanical participation in rescue was refined from a narrow "excavator clearing" expression to R5_1 ExcavatorAssistedRescue; and command actions led by fire, housing-construction, public security, or other professional agencies were represented by R4_1 ProfessionalAgencyCommand. These refinements indicate that the final ontology was not a fixed taxonomy determined entirely in advance, but an adapted E/B/R knowledge structure formed through theory-driven design, accident-text extraction, algorithm-assisted matching, and manual calibration.

The final E/B/R ontology contains 51 subclasses, including 16 event-trigger subclasses, 18 bearing-deficiency subclasses, and 17 emergency response subclasses. The complete ontology structure is provided in Appendix A. Figure 2 presents the final ontology implemented in Protégé, showing the three main dimensions and representative subclasses used for feature normalization, rule construction, and case representation.

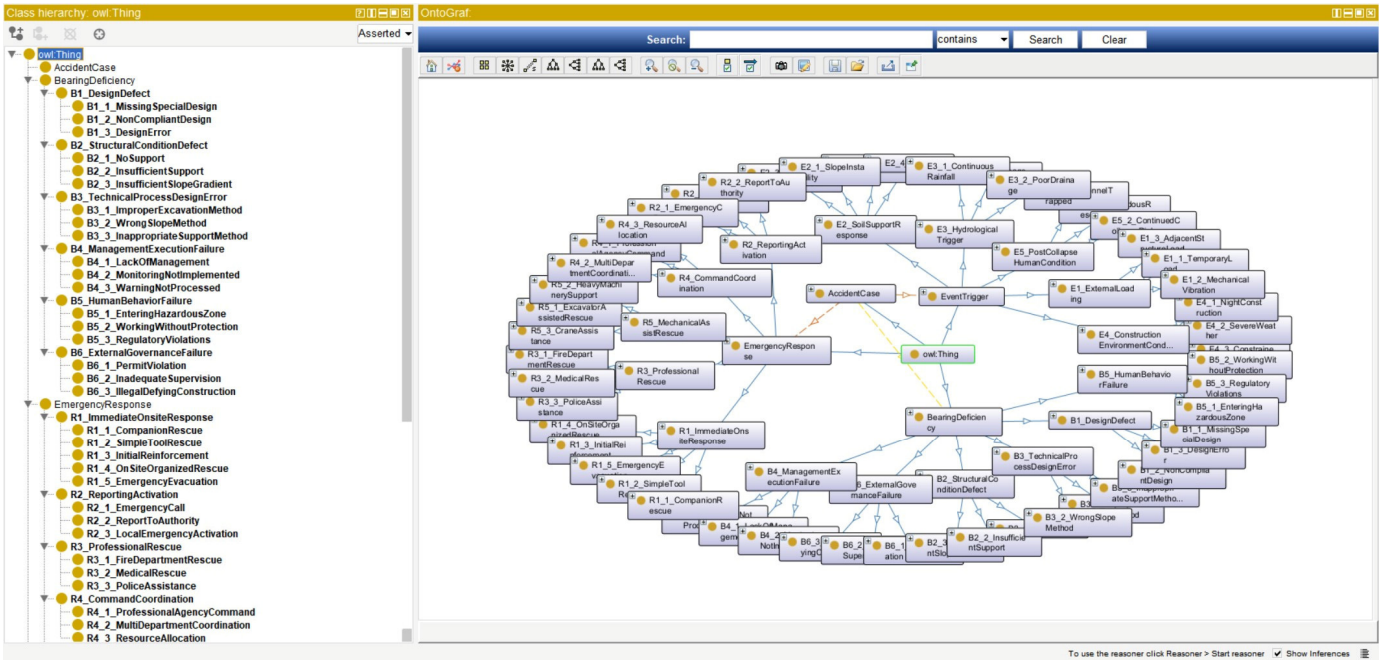


Figure 2. Final E/B/R ontology for foundation pit collapse accidents in Protégé.

After ontology calibration, the final E/B/R labels were grouped by accident case to construct a binary case–feature matrix. Each accident case was represented as one row, and each ontology subclass was represented as one feature column. A value of 1 indicates that the corresponding feature appears in the case, whereas 0 indicates that it was not coded and does not necessarily establish its actual absence. This matrix-based representation provides the input structure for CBR-based similarity retrieval.

The case–feature matrix was further imported into Neo4j to support structured storage and visual inspection. Figure 3 illustrates a representative case–feature graph, in which accident cases are connected with their corresponding E/B/R feature nodes. This graph representation makes the relationship between accident cases, structured features, and historical emergency responses traceable, and provides a structured basis for retrieving cases with the same or related E/B features during online recommendation.

Overall, the constructed case base provides a traceable and structured data foundation for the proposed hybrid reasoning framework. The retained reports provide accident-level cases, the calibrated E/B/R labels provide structured feature representations, and the Neo4j case graph supports subsequent similarity-based case retrieval.

To further examine the suitability of the E/B/R representation for emergency response recommendation, three historical accidents with contrasting documented conditions were analyzed using their original reports. The analysis compared response-relevant conditions with existing E/B categories and examined whether their combinations preserved meaningful scenario differences. Table 6 summarizes the source evidence, selected feature mappings, and representation boundaries.

The comparison shows how E/B combinations retain distinctions relevant to different response contexts. In the Shenzhen accident, the combination describes loose soil affected by vibration together with personnel entrapment. In the Zhuhai accident, support deformation, insufficient support, and further-collapse risk describe an evolving structural hazard associated with evacuation and stabilization. The Hefei accident combines personnel entrapment with drainage and support deficiencies, while also illustrating a specific surrounding-facility hazard that is not explicitly represented. Thus, shared features such as personnel entrapment do not imply that the remaining scenario conditions are identical.

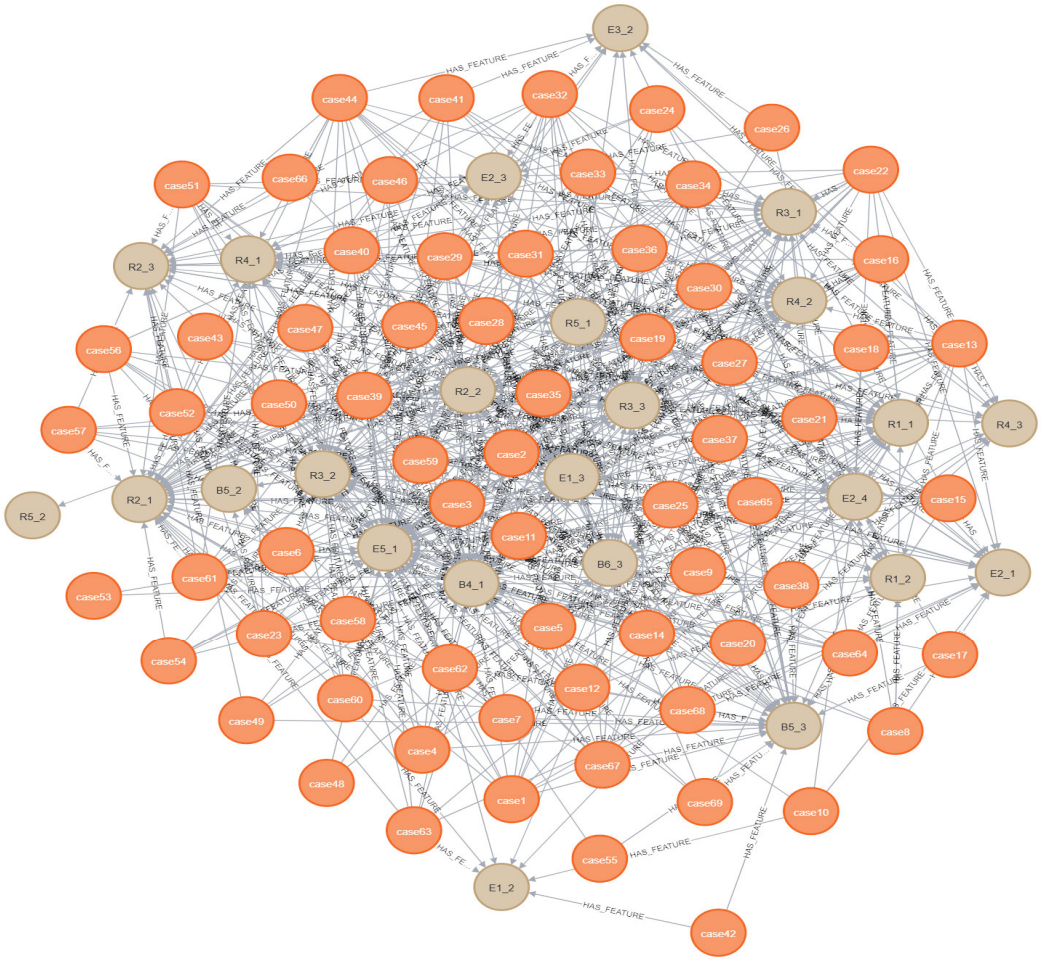


Figure 3. Neo4j representation of case–feature relationships in the foundation pit collapse case base.

These features are connected to response knowledge through case-level associations and explicit reasoning rules. Each historical case links its E/B features with the R measures documented in the report, while SWRL rules specify responses associated with particular E/B combinations. For example, support deformation combined with insufficient support is linked to initial reinforcement in the rule base. The representation therefore supports the use of joint conditions in reasoning and retrieval, although case-level associations do not themselves establish causal relationships or demonstrate that historical actions were optimal.

The analysis also distinguishes limitations of the source material from limitations of the representation. Accident reports vary in their descriptions of site conditions and emergency actions, restricting the information available for coding. However, some documented details, such as the threatened utility-pole foundation in the Hefei accident, are not explicitly captured by the current categories. The present representation consequently supports comparison at the response-category level without fully reproducing spatial relationships, event sequences, or detailed operational requirements. These historical examples provide case-specific evidence of its applicability and boundaries rather than comprehensive validation across all foundation pit collapse scenarios.

Overall, the constructed case base provides a structured foundation for the proposed hybrid reasoning framework. The E/B/R labels organize reported accident conditions and historical responses, the source-based analysis clarifies which scenario distinctions are retained, and the Neo4j case graph supports traceable case retrieval and knowledge reuse.

Table 6. Source-based analysis of response-relevant scenario distinctions in three historical accidents.

Historical Accident	Documented Conditions and Response Context	Selected E/B Features	Representation Boundaries
Shenzhen metro accident, 5 March 2016	Loose backfill was disturbed by traffic and drilling vibration and collapsed onto workers. On-site personnel organized their rescue and transport to hospital.	E1_2 MechanicalVibration; E2_4 SoftSoil; E5_1 PersonnelTrapped; B4_1 LackOfManagement	The representation captures vibration, loose soil, personnel entrapment, and management deficiencies, but does not specify burial depth, affected body parts, or rescue duration.
Zhuhai Zuguo Square accident, 6 May 1998	Support deformation and failure developed progressively, threatening workers and nearby buildings. Personnel were evacuated before the collapse expanded, and subsequent backfilling and stabilization aimed to prevent further sliding.	E2_3 SupportStructureDeformation; E5_2 ContinuedCollapseRisk; B2_2 InsufficientSupport; B3_1 ImproperExcavationMethod	The representation captures support-related deficiencies and further-collapse risk, but does not fully describe the locations of threatened buildings or the sequence of evacuation and stabilization.
Hefei trench accident, 28 August 2018	Inadequate drainage, insufficient slope grading, and absent support accompanied a collapse that buried a worker. Following rescue, the trench was backfilled to stabilize a nearby utility-pole foundation and prevent another accident.	E2_1 SlopeInstability; E3_2 PoorDrainage; E5_1 PersonnelTrapped; B2_1 NoSupport; B2_3 InsufficientSlopeGradient	The representation captures entrapment and unstable trench conditions, but does not explicitly identify the threatened utility-pole foundation or the specific object requiring stabilization.

4.2. Validation of Feature Extraction and Reasoning Components

Four validation checks were conducted to examine the reliability of the structured feature extraction and reasoning components: intra-annotator consistency of the manually calibrated E/B/R labels, comparison of candidate LLMs, repeated-run stability of online E/B extraction, and questionnaire-based evaluation of the SWRL rules. These checks were intended to determine whether the calibrated labels, constrained LLM extraction, and rule base could provide sufficiently consistent inputs and outputs for subsequent framework application.

First, an intra-annotator consistency check was conducted for the manually calibrated E/B/R labels. Approximately 15% of the manually labeled sentences were randomly selected and re-annotated after an interval under a blind setting. In total, 84 labeled sentences were included in the repeated annotation. Because the task involved fine-grained and multi-label annotation, agreement was evaluated at three levels. Strict text-level exact agreement was 52.4%, label-code set agreement was 72.6%, and coarse E/B/R-level agreement was 85.7%. These results indicate that the annotation was more consistent at the main E/B/R dimension level than at the fine-grained label level. The calibrated labels were therefore used as reference data for case-base construction and subsequent model comparison, while the remaining variation in fine-grained labels was treated as a limitation of manual annotation.

Second, Qwen3.6-Plus and DeepSeek-V3 were compared for ontology-constrained E/B extraction from Chinese accident-report texts. The comparison samples originated from the 84 sentences randomly selected for the intra-annotator consistency check, representing approximately 15% of the manually labeled sentences. After excluding sentences containing only R labels, 49 E/B-related samples remained. Ten were used to refine the prompt, ontology-label descriptions, and structured output requirements and were excluded from the formal comparison, leaving 39 test samples. Both models received the same test texts, ontology label space, finalized prompt, and output requirements. Only E/B extraction was

evaluated because emergency response labels were generated by the subsequent RBR and CBR modules.

Three indicators were used: label-code set agreement, E/B-level agreement, and micro-F1. Label-code set agreement required the predicted set of ontology codes to match the reference set. E/B-level agreement examined whether the main event-trigger and bearing-deficiency dimensions were correctly identified, whereas micro-F1 evaluated overall label-level extraction performance. The results are presented in Table 7.

Table 7. Comparison of candidate LLMs for ontology-constrained E/B extraction on the 39-sample independent test set.

Model	Label-Code Set Agreement	E/B-Level Agreement	Micro-F1
Qwen3.6-Plus	69.2%	97.4%	0.821
DeepSeek-V3	51.3%	89.7%	0.723

As shown in Table 7, Qwen3.6-Plus achieved higher label-code set agreement, E/B-level agreement, and micro-F1 than DeepSeek-V3. The corresponding exact label-code set results were 27 of 39 samples for Qwen3.6-Plus and 20 of 39 samples for DeepSeek-V3, while the E/B-level agreement results were 38 of 39 and 35 of 39 samples, respectively. Based on this comparison, Qwen3.6-Plus was selected for the online extraction module. In the prototype system, it was used in non-thinking mode with a fixed prompt, predefined ontology label space, and JSON Mode. Its role was limited to generating candidate E/B labels, which were subsequently checked through JSON parsing, ontology-code validation, invalid-label filtering, and duplicate removal before being passed to the reasoning modules.

Third, the repeated-run stability of the online E/B extraction module was examined. Six cases, including the three target cases and three additional cases, were processed five times with the temperature set to 0. The model, prompt, ontology label space, and extraction settings were kept unchanged. Across the 30 runs, no invalid ontology labels or output-formatting errors were observed, and E/B label-set consistency was 86.7%. These results indicate generally consistent E/B extraction under the tested settings.The repeated-run evaluation results are summarized in Table 8.

Table 8. Repeated-run stability of online E/B extraction.

Indicator	Result
Repeated runs	6 cases × 5 runs = 30
Invalid ontology labels or formatting errors	0/30
E/B label-set consistency	86.7%

Finally, the 21 SWRL rules used in the RBR module were evaluated through a questionnaire. The rules were developed based on relevant standards, emergency response plans, accident-case commonalities, and domain reasoning. They were organized into four groups: life-safety baseline rules, risk-control baseline rules, upgrade-triggered response rules, and extended candidate rules. The complete rule set is provided in Appendix B.

The questionnaire was completed by 20 respondents with engineering-related post-graduate backgrounds and relevant knowledge of construction safety, foundation pit engineering, or emergency response. Each rule was evaluated on a seven-point Likert scale in terms of rule rationality and the sufficiency of its triggering conditions. All 21 rules obtained mean scores above 5.0, indicating that the rule set was generally considered reasonable by the respondents. All rules were therefore retained, with minor wording

refinements made according to the open-ended feedback. Figure 4 presents a representative SWRL reasoning example.

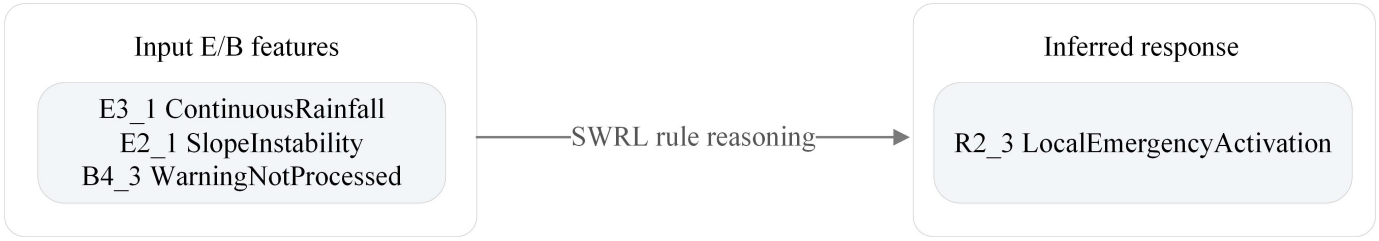


Figure 4. Example of SWRL-based rule reasoning for emergency response inference.

Overall, the validation results provide preliminary support for the components used in the subsequent target-case applications. The repeated annotation check examined the consistency of the calibrated E/B/R labels, the 39-sample comparison supported the selection of Qwen3.6-Plus for constrained E/B extraction, the repeated-run test examined E/B extraction consistency, and the questionnaire results supported the rationality of the SWRL rule set. These checks do not establish general performance across all foundation pit collapse scenarios but provide a basis for applying the framework to the selected target cases.

4.3. Application to Target Foundation Pit Collapse Cases

To examine the applicability of the proposed framework in realistic accident scenarios, three publicly available foundation pit collapse accident reports were selected as target cases and processed using the prototype system. The selected cases represented different accident conditions, including slope instability under soft-soil conditions, collapse associated with drainage and support deficiencies, and unsupported trench or small-pit collapse. Each target case was treated as a newly imported accident report. If a target case was already included in the historical case base, it was excluded from CBR retrieval to prevent self-matching.

For each target case, the system first extracted candidate E/B features using the ontology-constrained LLM extraction module. The extracted labels were then processed through JSON parsing, ontology-code validation, invalid-label filtering, and duplicate removal. The validated E/B features were subsequently provided to the RBR and CBR modules. RBR generated rule-driven baseline responses, whereas CBR retrieved historically related cases and provided case-supported response measures. The outputs of the two modules were then fused to generate the final emergency response recommendation.

Figure 5 presents the prototype interface for a representative target case. The interface displays the imported accident report, ontology-constrained E/B extraction results, rule- and case-based reasoning statistics, fused recommendations, and the LLM-generated explanation. The interface demonstrates the complete workflow from accident-report input to structured feature extraction, dual-source reasoning, response fusion, and explanation output.

The extracted E/B features and fused recommendations for the three target cases are summarized in Table 9.

The three target cases shared three responses: on-site organized rescue (R1_4), emergency calling (R2_1), and reporting to authorities (R2_2). These common outputs reflect the personnel-entrapment conditions identified in all three cases. Differences appeared in the additional rule- and case-supported responses, including initial reinforcement, rescue resources, and coordination measures.

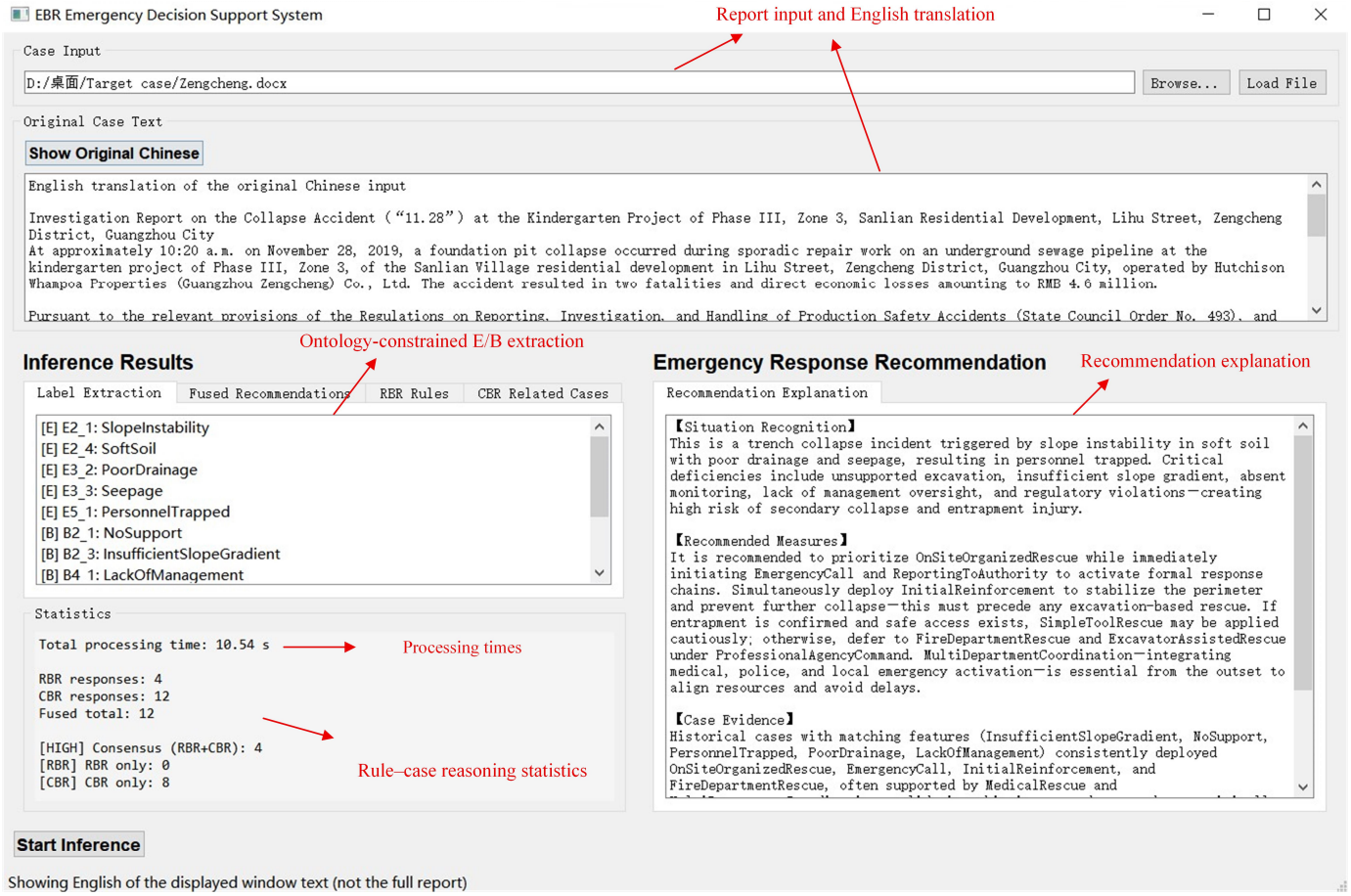


Figure 5. Prototype interface and recommendation output for a representative target case.

Table 9. Application results of the proposed framework for the target cases.

Case	Extracted Key E/B Features	Main Fused Recommendation
Target case 1: Liwan	SlopeInstability; PersonnelTrapped; SoftSoil; InsufficientSlopeGradient; ImproperExcavationMethod; LackOfManagement; RegulatoryViolations	OnSiteOrganizedRescue; EmergencyCall; ReportingToAuthority; SimpleToolRescue; PoliceAssistance
Target case 2: Zengcheng	SlopeInstability; SoftSoil; PoorDrainage; Seepage; PersonnelTrapped; NoSupport; InsufficientSlopeGradient; ImproperExcavationMethod; LackOfManagement; MonitoringNotImplemented; WarningNotProcessed; RegulatoryViolations	InitialReinforcement; OnSiteOrganizedRescue; EmergencyCall; ReportingToAuthority; SimpleToolRescue; LocalEmergencyActivation; FireDepartmentRescue; MedicalRescue; PoliceAssistance; ProfessionalAgencyCommand; MultiDepartmentCoordination; ExcavatorAssistedRescue
Target case 3: Baiyun	SlopeInstability; SoftSoil; PersonnelTrapped; NoSupport; LackOfManagement; RegulatoryViolations	InitialReinforcement; OnSiteOrganizedRescue; EmergencyCall; ReportingToAuthority; SimpleToolRescue; PoliceAssistance; CompanionRescue; MedicalRescue; MultiDepartmentCoordination

For Target case 1, the fused output contained five responses: the three common measures, simple-tool rescue, and police assistance. Target case 2 included drainage, seepage, and support deficiencies; its fused output contained initial reinforcement and additional historical options involving professional rescue, medical assistance, emergency activation,

coordination, and excavator-assisted rescue. Target case 3 also included absent support, and RBR contributed initial reinforcement alongside the rescue and coordination responses obtained through CBR. These outputs represent response categories for decision support; their implementation requires assessment of site conditions and available resources.

To examine the contribution of each reasoning component, the RBR, CBR, and fused outputs were compared for the three target cases. Table 10 summarizes the number of distinct response labels and their source composition.

Table 10. Comparison of response outputs and source composition across the three reasoning configurations.

Target Case	RBR Responses	CBR Responses	Fused Responses	Shared Responses	RBR-Only Responses	CBR-Only Responses
Liwan	3	5	5	3	0	2
Zengcheng	4	12	12	4	0	8
Baiyun	4	8	9	3	1	5

The three configurations were compared using the same validated E/B inputs. RBR generated responses from explicit rules, CBR collected historical responses from the two highest-ranked cases, and fusion combined the distinct response labels while retaining their sources. The comparison examined which measures each component contributed and how those contributions related to the target scenario. A larger response set was not treated as evidence of better recommendation quality.

For the Liwan case, RBR generated on-site organized rescue (R1_4), emergency calling (R2_1), and reporting to authorities (R2_2). All three responses were also present in the retrieved historical cases. CBR additionally contributed simple-tool rescue (R1_2) and police assistance (R3_3), producing five fused responses. These additions provide historical options beyond the rule-derived set, but their application remains conditional: tool-assisted rescue requires assessment of ground stability and victim accessibility, while police assistance depends on requirements for perimeter control, traffic management, or public order. Because the RBR responses were a subset of the CBR responses, fusion did not expand the CBR label set; its contribution was to distinguish the three responses with dual support from the two case-derived options.

For the Zengcheng case, RBR generated initial reinforcement (R1_3), on-site organized rescue (R1_4), emergency calling (R2_1), and reporting to authorities (R2_2). CBR included these four responses and eight additional labels covering simple-tool rescue, local emergency activation, fire and medical rescue, police assistance, professional command, multi-department coordination, and excavator-assisted rescue. The fused set therefore contained 12 responses. The rule-derived reinforcement response addresses the reported support deficiency, while the case-derived measures broaden the available rescue and coordination options. However, these additional labels are historical references rather than an instruction to implement every measure. In particular, excavator-assisted rescue requires professional assessment of site stability, equipment access, and the risk of further injury. As in the Liwan case, fusion added source differentiation rather than additional labels beyond CBR.

For the Baiyun case, RBR generated four responses, whereas CBR supplied eight. The two components shared on-site organized rescue (R1_4), emergency calling (R2_1), and reporting to authorities (R2_2). Initial reinforcement (R1_3) was provided only by RBR, while CBR contributed companion rescue (R1_1), simple-tool rescue (R1_2), medical rescue (R3_2), police assistance (R3_3), and multi-department coordination (R4_2). Fusion retained

nine distinct responses. This case illustrates explicit complementarity: RBR supplied a stabilization-related response absent from the retrieved historical response sets, while CBR supplied additional rescue and coordination options. Companion and tool-assisted rescue nevertheless require safety assessment and should not be interpreted as permission to enter an unstable, unsupported excavation.

The end-to-end processing times recorded in one run for each target case were 13.29 s, 10.54 s, and 12.48 s for the Liwan, Zengcheng, and Baiyun cases, respectively. Timing covered E/B extraction, rule-based reasoning, case retrieval, response fusion, and recommendation explanation generation, excluding file loading and the translation used for English display. These measurements describe the observed processing times under the tested configuration; processing time may vary with input length, network conditions, and LLM service response time.

Overall, the comparison identifies two contributions of hybrid reasoning. First, rules can supply responses that are absent from the retrieved historical records, as illustrated by initial reinforcement in the Baiyun case. Second, fusion makes the distinction between rule-supported, case-supported, and jointly supported responses explicit. Historical cases broaden the available response options, while rules provide an explicit connection between encoded scenario conditions and response categories. These results demonstrate output complementarity and traceability in the examined cases; they do not establish that all retained measures are necessary or that a larger fused set improves operational outcomes.

4.4. Comparative Evaluation of Case Retrieval

To examine the value of the E/B-based knowledge representation for case retrieval, two text-based methods were introduced for comparison: BM25 and BGE-small-zh-v1.5. BM25 ranks historical reports according to lexical matches, whereas BGE-small-zh-v1.5 ranks them according to semantic similarity between report texts. In contrast, the proposed method compares structured E/B labels representing accident conditions and deficiencies. The comparison therefore examines whether explicitly representing these features helps identify historical cases with response-relevant scenario similarities, rather than relying on overall textual similarity.

All three methods searched the same 69 historical cases using the three target accidents presented in Section 4.3. Actual emergency response descriptions were excluded from the text retrieval inputs, and R labels were not used in E/B similarity calculation. The text-based implementations returned the five highest-ranked cases, of which the first two were selected for comparison with the two cases returned by E/B-based retrieval. Their original reports were then compared with the target accidents in terms of ground conditions, water-related conditions, support deficiencies, excavation activities, and personnel entrapment. Table 11 presents the retrieved case identifiers in ranking order. Raw similarity scores were not compared because the three methods use different scoring scales.

Table 11. Comparison of the two highest-ranked historical cases retrieved by E/B-based and text-based methods.

Target Accident	E/B-Based Retrieval	BM25	BGE-Small-zh-v1.5
Target 1: Liwan	50;17	4;11	6;4
Target 2: Zengcheng	44;32	6;4	6;4
Target 3: Baiyun	33;11	4;12	4;6

Notes: Numbers denote historical case identifiers; within each cell, the first and second identifiers indicate the first- and second-ranked results, respectively.

Both text-based methods scored the filtered target-report text against historical excerpts covering the accident process through the accident-nature section. Explicit emergency-action sentences and sections describing rescue or post-accident arrangements were excluded, and ambiguous sentences were omitted from scoring. BM25 used a custom implementation with jieba tokenization, k1 = 1.5, and b = 0.75; tokens shorter than two characters and numeric-only tokens were removed, without a stopword list. BGE used BAAI/bge-small-zh-v1.5 through SentenceTransformer. Texts were divided into 400-character chunks with an 80-character overlap. Normalized chunk embeddings were averaged and normalized again to obtain a case vector, and cases were ranked by cosine similarity.

For the Liwan accident, the E/B-based method retrieved Case 17, which shares soft-ground conditions, an excessively steep excavation slope, improper excavation sequencing, and personnel entrapment. This case provides a closer correspondence to the target's slope and excavation conditions than the shallower pipeline excavations in Cases 4 and 11 or the artificial bored-pile excavation in Case 6 retrieved by the text-based methods. This illustrates the value of representing accident conditions and deficiencies explicitly: relevant features can be matched despite differences in report wording and project descriptions. However, the first-ranked E/B-based result, Case 50, provides only a partial match, showing that the structured representation does not eliminate retrieval limitations.

For the Zengcheng accident, the first-ranked E/B-based result, Case 44, combines a pipeline excavation setting, soft ground, water seepage, absence of support, and personnel burial. This combination corresponds closely to the target's water-affected, unsupported excavation and provides a relevant historical reference for considering scene stabilization and rescue coordination. Both text-based methods retrieved Cases 6 and 4. Although Case 4 shares relevant unsupported pipeline-excavation conditions, Case 6 concerns an artificial bored-pile excavation with different working-space and support conditions. The value of E/B-based retrieval therefore lies in making the combination of response-relevant conditions explicit, rather than treating general similarity between collapse reports as sufficient evidence of scenario correspondence. The second-ranked Case 32 nevertheless remains a partial match because its engineered support conditions differ from those of the target.

For the Baiyun accident, both E/B-based results, Cases 33 and 11, share the combination of pipeline excavation, unstable soil, absent support, and personnel entrapment. These results provide a more consistent scenario reference than the text-based result sets, which also include either an engineered retaining-wall failure involving soil and water inrush (Case 12) or an artificial bored-pile excavation (Case 6). Case 4, retrieved by both text-based methods, is also relevant. Thus, the advantage observed in this example is the concentration of the E/B-based results on the target's specific combination of conditions, rather than an exclusive ability to retrieve relevant accidents.

Cases 4 and 6 recur across the text-based results for different target accidents. This pattern indicates that the text-based rankings do not always change substantially with the target scenario. One possible explanation is that common collapse-related vocabulary and broader narrative similarities contribute strongly to lexical and semantic matching, even when excavation settings or failure conditions differ. However, the repeated retrieval of these cases alone does not establish this explanation, and Case 4 is genuinely relevant to several targets. The practical limitation is therefore that textual similarity does not directly reveal which response-relevant conditions support the match; these conditions must be checked in the retrieved narratives.

In contrast, the E/B-based method represents reported conditions and deficiencies using a common vocabulary and exposes the matched features as retrieval evidence. Within the E/B/R case structure, the retrieved cases are also linked to historical response labels,

supporting subsequent response fusion without requiring users to locate these measures manually in each report. These capabilities illustrate the framework's value in organizing scenario information, explaining case selection, and connecting historical scenarios with response knowledge. These findings illustrate the framework's retrieval and traceability value in the three examined scenarios. The complementary contributions of RBR and CBR to the response outputs are examined separately in Section 4.3.

4.5. Questionnaire-Based Evaluation of Recommendation Results

The questionnaire evaluated the prototype outputs available at the time of the survey. Subsequent revisions corrected CBR retrieval from the full Neo4j case base and introduced a hierarchy-aware Dice similarity measure to account for unmatched historical features and reduce the tendency to favor cases containing many labels. These revisions changed some retrieved cases and recommendation outputs. The questionnaire is therefore retained as an initial assessment of the earlier prototype's perceived recommendation quality and explanation clarity.

A total of 19 valid responses were collected: nine respondents were engineering or construction-management personnel, one was a university teaching or research professional, and nine held master's degrees with a focus on construction project management. Fifteen respondents had also participated in the preceding SWRL rule questionnaire, although the two questionnaires addressed different assessment tasks. Given the small sample, the results are interpreted as an exploratory assessment of participant perceptions.

For each target case, respondents reviewed the accident scenario, extracted E/B features, recommendations, and accompanying explanation from the earlier prototype. Four criteria were evaluated: measure relevance, referring to correspondence with the scenario's main risk characteristics; content completeness, referring to coverage of its principal emergency response needs; field applicability, referring to the perceived value of the recommendations as a reference for implementation; and explanation clarity, referring to whether the explanation helped respondents understand the recommendation logic. Each criterion was rated on a seven-point Likert scale, from 1 ("very unsuitable") to 7 ("very suitable"). The results are summarized in Table 12.

Table 12. Questionnaire-based evaluation results for the emergency response recommendations.

Case	Measure Relevance	Content Completeness	Field Applicability	Explanation Clarity	Overall Mean
Target case 1	6.00	6.00	5.84	5.89	5.93
Target case 2	6.11	6.05	5.89	5.95	6.00
Target case 3	6.05	6.05	6.00	5.84	5.99
Average	6.05	6.03	5.91	5.89	5.97

All three target cases received overall mean scores close to 6.0 on the seven-point scale. Across the 228 individual ratings obtained from the 19 respondents, three cases, and four evaluation criteria, 220 ratings were equal to or higher than 5, accounting for 96.49% of the total. Target case 2 received the highest overall mean score of 6.00, followed by Target case 3 at 5.99 and Target case 1 at 5.93.

Measure relevance and content completeness received the highest average scores, at 6.05 and 6.03, respectively. These results indicate that the respondents generally considered the recommended measures to be consistent with the identified accident features and sufficiently comprehensive for the presented scenarios. Field applicability and explanation clarity received slightly lower average scores of 5.91 and 5.89, respectively.

The relatively lower scores for field applicability and explanation clarity suggest that further improvement is needed in the presentation of implementation details. Such details may include the sequence of rescue actions, availability of equipment and personnel, surrounding environmental constraints, and measures for preventing secondary collapse.

Overall, the questionnaire results provide supporting evidence that the recommendations were generally perceived as relevant, complete, applicable, and understandable for the three target cases. However, the evaluation reflects respondent perceptions under case-based assessment conditions and should not be interpreted as evidence of effectiveness in real-time emergency operations. Broader evaluation involving additional cases and experienced emergency response practitioners is required before field application. The ratings reflect participants' perceptions of the evaluated outputs and do not independently verify the factual accuracy of the LLM explanations or the operational effectiveness of the updated recommendations.

5. Discussion

5.1. Implications for Construction Safety and Emergency Decision Support

The main contribution of this study lies in integrating an LLM, the E/B/R ontology, RBR, and CBR into a traceable framework for organizing and reusing knowledge from foundation pit collapse accident reports. Previous studies have examined ontology or knowledge graph construction [16], domain-knowledge-enhanced LLM applications [17], ontology-based LLM extraction [18], and hybrid reasoning in engineering tasks [19]. Building on these studies, the proposed framework connects unstructured accident reports with ontology-constrained feature extraction, knowledge-driven reasoning, response fusion, and explanation generation. This integration provides a structured means of transforming historical accident information into reusable knowledge for construction emergency decision support.

The E/B/R ontology provides a shared representation of accident events, deficiencies in the bearing system, and corresponding emergency responses. Ontologies support the explicit specification, sharing, and reuse of domain concepts and relationships [7]. In construction safety research, ontology-based approaches have been applied to automated safety checking and construction risk knowledge management [8,9]. The present study extends this application to post-accident emergency response by linking Event (E) and Bearing (B) features extracted from accident reports with Response (R) measures. The ontology therefore serves not only as a classification structure but also as the knowledge layer supporting feature extraction, rule construction, case retrieval, and response recommendation.

The framework also provides a controlled approach to incorporating LLMs into construction emergency decision support. Direct use of LLMs in safety-critical decision-making may produce unreliable results because of hallucination and insufficient grounding in domain knowledge [13,14]. Accordingly, the LLM in this study does not directly determine emergency responses. Its role is limited to ontology-constrained E/B feature extraction and the subsequent explanation of the reasoning results. Candidate labels are restricted by the ontology, returned in a structured format, and subjected to code validation and filtering. The core response recommendations are generated by the knowledge-driven RBR and CBR modules. This division of responsibilities improves the traceability of the recommendation process and reduces reliance on unconstrained LLM generation.

RBR and CBR provide two complementary sources of response knowledge. CBR reuses the experience contained in previous cases to address a new problem [10], and its applicability to construction safety risk management has been demonstrated in previous research [12]. SWRL rules, meanwhile, provide an explicit mechanism for ontology-based inference [33]. In the proposed framework, RBR generates predefined baseline measures

from validated E/B features, whereas CBR supplements these measures with responses retrieved from related historical accidents. The fused result retains the source of each response, allowing users to identify whether a recommendation is supported by rules, historical cases, or both.

The results for the three target cases illustrate different contributions of the two reasoning components. In the Liwan and Zengcheng cases, all RBR responses were also included in the CBR outputs. Fusion therefore preserved the CBR response sets while distinguishing jointly supported measures from case-supported options. In the Baiyun case, RBR additionally supplied initial reinforcement, which was absent from the retrieved historical response sets, while CBR contributed further rescue and coordination options. These findings illustrate that fusion can preserve rule-derived measures missing from historical records and make the support for each response explicit. Its value lies in response complementarity and traceability, rather than in increasing the number of recommendations alone.

Overall, the proposed framework offers an explainable approach to reusing accident-report knowledge in foundation pit emergency response. The ontology, rules, retrieved cases, and response-source labels make the recommendation process inspectable, while the LLM-generated explanation improves the accessibility of the structured reasoning results. The current validation supports the preliminary applicability of this approach as a construction safety decision-support prototype. Broader validation is still required before its effectiveness can be generalized to other projects, accident types, or real-time emergency environments.

5.2. Practical Implications, Limitations, and Transferability

From a practical perspective, the proposed framework should be regarded as a knowledge transformation and decision-support prototype rather than a real-time emergency command system. The current implementation uses accident investigation reports as input because these reports contain relatively reliable descriptions of accident processes, risk features, causes, and response measures. By transforming these textual records into structured E/B features, rule-driven responses, case-supported responses, fused recommendations, and corresponding explanations, the framework supports the systematic reuse of historical emergency response knowledge.

The framework can support several construction safety management activities. First, it can assist safety managers and emergency management personnel in reviewing previous foundation pit collapse accidents and identifying relationships between accident features and response measures. Second, it can support emergency training, case-based learning, and emergency plan improvement by converting fragmented accident narratives into structured and traceable knowledge. Third, the prototype provides a reasoning structure that could be extended to process preliminary site descriptions or other on-site information in future applications. However, such real-time use would require additional data interfaces, continuous information updating, and further system validation.

The ontology, rule base, and case base can be maintained as new accident information and emergency management knowledge become available. Structured knowledge bases and semantic representations support the accumulation and reuse of construction safety knowledge [7–9]. In the current framework, however, these updates are performed manually. New ontology concepts must be reviewed before being incorporated, SWRL rules must be encoded and validated using relevant standards and expert knowledge, and new cases must satisfy the predefined inclusion and structuring requirements. This controlled maintenance process supports knowledge quality and traceability, although it does not provide automatic ontology evolution or rule learning.

Several limitations should be acknowledged. First, the case base was constructed from publicly available accident investigation reports. Of the 1110 collected reports, only 69 foundation pit collapse reports satisfied the inclusion criteria and were retained. Although these reports provide useful factual information, their level of detail varies, and some reports contain incomplete descriptions of emergency responses. Consequently, the current case base may not cover all possible foundation pit collapse conditions.

Second, the validation scale remains limited. The availability of suitable public foundation pit collapse reports constrains the domain-specific source material. The LLM comparison used 39 test samples obtained through the random sampling, re-annotation, and exclusion procedure described in Section 4.2. The micro-F1 score of 0.821 therefore characterizes extraction performance on this test set, while the repeated-run evaluation provides complementary evidence of E/B extraction consistency under fixed settings. Neither evaluation establishes general performance across all foundation pit collapse conditions. The framework was also applied to only three target cases, and the questionnaire evaluations involved 20 participants in the rule assessment and 19 valid responses in the earlier prototype recommendation assessment. The recommendation questionnaire assessed the earlier prototype outputs and provided an exploratory assessment of participant perceptions; it did not directly validate the updated recommendation outputs. Evaluation using a larger and more diverse independently annotated dataset, additional target cases, and emergency response practitioners remains future work. Third, the ontology is currently static, and the SWRL rules are primarily derived from standards, accident knowledge, and expert judgment. The CBR module also relies on the current ontology-feature-based similarity calculation, while the fusion strategy preserves responses generated by the two reasoning sources according to predefined criteria. These designs improve transparency and implementation feasibility but may limit adaptation to previously unseen or rapidly changing accident conditions. Future studies could investigate ontology updating, rule learning, alternative case-similarity methods, and adaptive fusion strategies as more validated data become available.

Fourth, the model comparison was limited to Qwen3.6-Plus and DeepSeek-V3 for the Chinese accident-report extraction task. The observed results should therefore not be interpreted as a general comparison of LLM performance. Additional models, report types, prompt settings, and deployment environments should be examined in future studies. Moreover, the generated responses are decision-support recommendations rather than complete emergency plans. Actual emergency actions must still be determined by qualified personnel according to site conditions, rescue resources, weather, surrounding structures, personnel status, and the risk of secondary collapse.

The framework has potential adaptability to other construction collapse and emergency management scenarios. Its general process—E/B/R ontology representation, LLM-constrained extraction, RBR and CBR, response fusion, and explanation generation—could be applied to scenarios such as trench collapse, rainfall-induced slope failure, retaining-wall instability, or temporary excavation collapse. However, this transfer cannot be achieved by directly applying the current knowledge base. The domain-specific ontology, rules, case base, and validation criteria would need to be reconstructed and evaluated for each new application. Therefore, the current study provides a domain-specific prototype for emergency response to foundation pit collapse accidents and a potentially adaptable framework for future knowledge-driven construction emergency decision-support applications.

6. Conclusions

This study developed an LLM-enabled E/B/R ontology-driven hybrid reasoning framework for emergency response recommendation in foundation pit collapse accidents.

The framework integrates E/B/R ontology representation, LLM-constrained extraction, RBR, CBR, response fusion, and LLM-based explanation. The LLM is limited to extracting candidate E/B features within the ontology label space and explaining the final results, while the emergency response recommendations are generated by the knowledge-driven RBR and CBR modules. This design provides a traceable approach to transforming accident investigation reports into structured knowledge for construction emergency decision support.

The E/B/R ontology was constructed through top-down design and bottom-up refinement and contains 51 subclasses, including 16 Event subclasses, 18 Bearing subclasses, and 17 Response subclasses. From 1110 initially collected construction collapse accident reports, 69 foundation pit collapse reports were retained through automatic classification, keyword-assisted screening, and manual verification. The calibrated E/B/R labels were organized into a structured case–feature matrix and represented in Neo4j, providing a semantic basis for feature extraction, rule construction, case retrieval, and historical response reuse.

The validation results showed that ontology constraints and structured post-processing enabled the LLM module to produce usable E/B extraction results. On the 39-sample comparison set, Qwen3.6-Plus achieved a micro-F1 score of 0.821, compared with 0.723 for DeepSeek-V3, and was therefore adopted in the prototype. Repeated-run testing under fixed extraction settings yielded an E/B label-set consistency of 86.7%, with no invalid ontology labels or output-formatting errors observed across 30 runs. The SWRL rules were also evaluated by domain participants, with all 21 rules receiving mean scores above 5 on the seven-point scale.

Application to three target cases illustrated the contributions of RBR and CBR. On-site organized rescue, emergency calling, and reporting to authorities were supported by both components in all three cases. CBR supplied additional historical response options, while RBR contributed initial reinforcement absent from the retrieved response sets in the Baiyun case. In the Liwan and Zengcheng cases, fusion retained the same response labels as CBR while distinguishing jointly supported measures from case-only options. The comparison with BM25 and BGE-small-zh-v1.5 further illustrated how E/B-based retrieval exposes matched accident features and links retrieved scenarios to historical responses. These findings support the framework's retrieval, complementarity, and traceability value in the examined cases, without establishing general superiority or operational effectiveness.

The questionnaire-based evaluation assessed the earlier prototype outputs and accompanying explanations. The three target cases received overall mean scores of 5.93, 6.00, and 5.99 on a seven-point scale, with an overall average of 5.97. These ratings indicate favorable participant perceptions of the evaluated outputs in terms of measure relevance, content completeness, field applicability, and explanation clarity. Because the CBR implementation and retrieval method were updated during revision, these scores do not directly validate the updated recommendations. The revised outputs were examined separately through the comparison of reasoning configurations.

Taken together, the results constitute a preliminary validation of an ontology-based emergency response knowledge recommendation framework. The report-based case analyses and comparisons provide evidence of its ability to organize and retrieve response-relevant knowledge and make recommendation sources traceable in the examined scenarios. They do not demonstrate effectiveness in actual emergency response operations, which requires further evaluation under real-world operational conditions.

The present study remains subject to several limitations. The case base was derived from publicly available reports, the framework was applied to only three target cases, and the current ontology and SWRL rules require manual maintenance. The case-similarity and fusion methods also require further evaluation using a larger and more diverse set

of accidents. Moreover, the prototype processes accident reports and should be regarded as a decision-support tool rather than a real-time emergency command system. Future research should expand the case base and target-case validation, refine the similarity and fusion strategies, examine ontology and rule updating, and evaluate the framework using additional LLMs and site-specific information. These developments would support broader application of knowledge-driven and explainable decision support in construction emergency management.

Author Contributions: Conceptualization, T.W.; methodology, T.W.; software, T.W.; validation, T.W. and J.L.; formal analysis, T.W.; investigation, T.W.; resources, T.W.; data curation, T.W. and J.L.; writing—original draft preparation, T.W.; writing—review and editing, T.W. and J.L.; visualization, T.W.; supervision, J.L.; project administration, J.L.; funding acquisition, J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of Hunan Province, China (Grant No. 2021JJ30744), and supported by the Postgraduate Scientific Research Innovation Project of Changsha University of Science & Technology (Grant No. CLKYCX25010).

Institutional Review Board Statement: Ethical review and approval were not required because the voluntary, anonymous questionnaire collected no sensitive or identifiable information and posed minimal risk.

Informed Consent Statement: Informed consent was obtained from all participants.

Data Availability Statement: The data presented in this study are available on reasonable request from the corresponding author.

Acknowledgments: During manuscript preparation, the authors used ChatGPT (OpenAI; the latest model available at the time of use) solely for language polishing and translation. All research content and conclusions were developed by the authors, who reviewed the output and take full responsibility for this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. E/B/R Ontology Classes for Foundation Pit Collapse Accidents

Table A1. Event-trigger ontology classes.

Dimension	Parent Code	Parent Category	Subclass Code	Subclass Label
Event triggers	E1	ExternalLoading	E1_1	TemporaryLoading
			E1_2	MechanicalVibration
			E1_3	AdjacentStructureLoad
	E2	SoilSupportResponse	E2_1	SlopeInstability
			E2_2	ToeUndermining
			E2_3	SupportStructureDeformation
			E2_4	SoftSoil
	E3	HydrologicalTrigger	E3_1	ContinuousRainfall
			E3_2	PoorDrainage
			E3_3	Seepage
	E4	ConstructionEnvironmentCondition	E4_1	NightConstruction
			E4_2	AdverseWeather
			E4_3	ConstrainedSite
	E5	PostCollapseHumanCondition	E5_1	PersonnelTrapped
			E5_2	ContinuedCollapseRisk
			E5_3	HazardousRescueEnvironment

Table A2. Bearing-deficiency ontology classes.

Dimension	Parent Code	Parent Category	Subclass Code	Subclass Label
Bearing deficiencies	B1	DesignDefect	B1_1	MissingSpecialDesign
			B1_2	NonCompliantDesign
			B1_3	DesignError
	B2	StructuralConditionDefect	B2_1	NoSupport
			B2_2	InsufficientSupport
			B2_3	InsufficientSlopeGradient
Bearing deficiencies	B3	TechnicalProcessDesignError	B3_1	ImproperExcavationMethod
			B3_2	ImproperSlopeMethod
			B3_3	InappropriateSupportMethod
	B4	ManagementExecutionFailure	B4_1	LackOfManagement
			B4_2	MonitoringNotImplemented
			B4_3	WarningNotProcessed
	B5	HumanBehaviorFailure	B5_1	EnteringHazardousZone
			B5_2	WorkingWithoutProtection
			B5_3	RegulatoryViolations
	B6	ExternalGovernanceFailure	B6_1	PermitViolation
			B6_2	InadequateSupervision
			B6_3	IllegalDefyingConstruction

Table A3. Emergency response ontology classes.

Dimension	Parent Code	Parent Category	Subclass Code	Subclass Label
Emergency responses	R1	ImmediateOnsiteResponse	R1_1	CompanionRescue
			R1_2	SimpleToolRescue
			R1_3	InitialReinforcement
			R1_4	OnSiteOrganizedRescue
			R1_5	EmergencyEvacuation
	R2	ReportingActivation	R2_1	EmergencyCall
			R2_2	ReportingToAuthority
			R2_3	LocalEmergencyActivation
	R3	ProfessionalRescue	R3_1	FireDepartmentRescue
			R3_2	MedicalRescue
			R3_3	PoliceAssistance
	R4	CommandCoordination	R4_1	ProfessionalAgencyCommand
			R4_2	MultiDepartmentCoordination
			R4_3	ResourceAllocation
	R5	MechanicalAssistRescue	R5_1	ExcavatorAssistedRescue
			R5_2	HeavyMachinerySupport
			R5_3	CraneAssistance

Appendix B. SWRL Rule Set for Emergency Response Recommendation

Table A4. SWRL-based RBR rule set.

Rule ID	Rule Group	Triggering E/B Features	Inferred Response	Basis
RBR-01	Life-safety baseline	E5_1 PersonnelTrapped	R2_1 EmergencyCall	A
RBR-02	Life-safety baseline	E5_1 PersonnelTrapped	R1_4 OnSiteOrganizedRescue	A
RBR-03	Life-safety baseline	E5_1 PersonnelTrapped	R2_2 ReportingToAuthority	A
RBR-04	Life-safety baseline	E5_2 ContinuedCollapseRisk	R1_5 EmergencyEvacuation	A
RBR-05	Life-safety baseline	E5_1 PersonnelTrapped; E5_3 HazardousRescueEnvironment	R4_2 MultiDepartmentCoordination	C

Table A4. *Cont.*

Rule ID	Rule Group	Triggering E/B Features	Inferred Response	Basis
RBR-06	Life-safety baseline	E5_1 PersonnelTrapped; E5_3 HazardousRescueEnvironment	R4_1 ProfessionalAgencyCommand	C
RBR-07	Life-safety baseline	E5_1 PersonnelTrapped; E5_3 HazardousRescueEnvironment; E4_3 ConstrainedSite	R4_3 ResourceAllocation	C
RBR-08	Life-safety baseline	E5_1 PersonnelTrapped; E5_3 HazardousRescueEnvironment	R2_3 LocalEmergencyActivation	C
RBR-09	Risk-control baseline	E3_1 ContinuousRainfall; B2_1 NoSupport	R1_3 InitialReinforcement	B
RBR-10	Risk-control baseline	E3_1 ContinuousRainfall; B2_2 InsufficientSupport	R1_3 InitialReinforcement	B
RBR-11	Risk-control baseline	E3_3 Seepage; B2_1 NoSupport	R1_3 InitialReinforcement	B
RBR-12	Risk-control baseline	E2_1 SlopeInstability; B2_1 NoSupport	R1_3 InitialReinforcement	A
RBR-13	Risk-control baseline	E2_3 SupportStructureDeformation; B2_2 InsufficientSupport	R1_3 InitialReinforcement	A
RBR-14	Upgrade-triggered response	E5_2 ContinuedCollapseRisk; E5_3 HazardousRescueEnvironment	R2_3 LocalEmergencyActivation	A
RBR-15	Upgrade-triggered response	E3_1 ContinuousRainfall; E2_1 SlopeInstability; B4_3 WarningNotProcessed	R2_3 LocalEmergencyActivation	A
RBR-16	Upgrade-triggered response	E3_3 Seepage; E2_3 SupportStructureDeformation; B4_2 MonitoringNotImplemented	R2_3 LocalEmergencyActivation	B + C
RBR-17	Extended candidate rule	E5_1 PersonnelTrapped; E5_3 HazardousRescueEnvironment	R3_1 FireDepartmentRescue	C + case commonality
RBR-18	Extended candidate rule	E5_1 PersonnelTrapped; E5_3 HazardousRescueEnvironment	R3_2 MedicalRescue	A + case commonality
RBR-19	Extended candidate rule	E2_1 SlopeInstability; E5_1 PersonnelTrapped; E5_3 HazardousRescueEnvironment	R5_1 ExcavatorAssistedRescue	C + case commonality
RBR-20	Extended candidate rule	E2_3 SupportStructureDeformation; E5_1 PersonnelTrapped; E5_3 HazardousRescueEnvironment	R5_3 CraneAssistance	C + case commonality
RBR-21	Extended candidate rule	E5_1 PersonnelTrapped; E4_3 ConstrainedSite; E5_3 HazardousRescueEnvironment	R5_2 HeavyMachinerySupport	C + case commonality

Notes: A = general emergency response requirements, including emergency reporting, on-site rescue, evacuation, response activation, and information reporting. B = foundation-pit risk-control requirements, including temporary stabilization, monitoring, warning response, and emergency control measures. C = escalation and coordination requirements, including local emergency activation, professional command, multi-department coordination, and resource allocation.

Appendix C. Prompt Template for Ontology-Constrained E/B Extraction

The original prompt used in the prototype system was implemented in Chinese because the accident investigation reports analyzed in this study were written in Chinese. For accessibility to international readers, a complete English translation of the final prompt is provided below. The ontology codes, prompt structure, extraction constraints, and output format are consistent with those used in the prototype system.

Prompt Template Used in the Prototype System:

You are an "E/B label extraction assistant for construction foundation pit collapse accidents."

Task: Based on the input text describing a construction foundation pit collapse accident, extract the most appropriate E and B labels from the provided E/B ontology labels.

Definitions:

E represents event triggers, including accident scenarios, triggering conditions, and on-site states.

B represents bearing deficiencies, including deficiencies in system bearing, control, management, or response capacity.

R represents emergency responses. However, R labels must not be extracted in this task. They are generated by the subsequent RBR and CBR modules.

Important requirements:

1. Select labels only from the provided E/B label set.
2. Extract only E and B labels; do not extract R labels.
3. Do not create new labels.
4. Do not output label names.
5. Do not provide explanations, reasoning processes, or additional text.
6. Output only strict JSON.
7. If there is no explicit semantic evidence for a label category, return an empty array for that category.
8. If the same meaning appears repeatedly in the text, retain the corresponding label only once.
9. Extract a label only when explicit semantic evidence is present in the text. Do not extract labels based solely on inference.
10. Priority should be given to identifying key labels concerning trapped personnel, slope instability, soft soil, absence of support, insufficient support, unsafe operations, and management deficiencies.

The required output format is:

{“E”: [“E2_1”, “E5_1”], “B”: [“B2_1”, “B5_3”]}

If no label is identified for a category, return an empty array:

{“E”: [], “B”: []}

The E labels are as follows:

E1_1 Temporary surcharge
E1_2 Mechanical vibration
E1_3 Adjacent structural load
E2_1 Slope instability
E2_2 Slope-toe weakening
E2_3 Support structure deformation
E2_4 Soft soil
E3_1 Continuous rainfall
E3_2 Poor drainage
E3_3 Seepage or water inrush
E4_1 Nighttime construction
E4_2 Adverse weather
E4_3 Restricted site space
E5_1 Personnel trapped
E5_2 Risk of secondary collapse
E5_3 Rescue-constrained environment

The B labels are as follows:

B1_1 Lack of specialized design
B1_2 Design noncompliance
B1_3 Design error

B2_1 Absence of support
B2_2 Insufficient support
B2_3 Inadequate slope grading
B3_1 Improper excavation method
B3_2 Incorrect slope-grading method
B3_3 Improper support method
B4_1 Management deficiency
B4_2 Monitoring not implemented
B4_3 Warning not addressed
B5_1 Entry into a dangerous area
B5_2 Unprotected work
B5_3 Unsafe operation
B6_1 Improper permitting
B6_2 Inadequate supervision
B6_3 Illegal or noncompliant construction

Label identification guidelines:

1. Extract E5_1 when the text indicates that personnel were buried, trapped, pinned beneath collapsed material, unable to escape independently, or otherwise confined within the collapsed mass.

2. Extract E5_2 when the text indicates a secondary collapse, repeated collapse, continued collapse, ongoing failure, or a slope that remains unstable.

3. Extract E5_3 when the text indicates rescue difficulty, restricted rescue operations, confined rescue space, obstructed rescue access, or an inability to approach the accident location.

4. Extract E2_1 when the text indicates slope collapse, slope instability, soil collapse, soil sliding, or a cave-in.

5. Extract E2_4 when the text indicates soft soil, weak soil, a weak soil layer, poor foundation soil, low soil stability, loose soil after excavation, or an unstable soil layer.

6. Extract E2_3 when the text indicates deformation of a support structure, inclination or deformation of a retaining wall, abnormal support axial force, support deformation, or failure of a support structure.

7. Extract E3_1 when the text indicates continuous rainfall, prolonged rainfall, heavy rain, or rainwater erosion.

8. Extract E3_2 when the text indicates poor drainage, accumulated water inside the foundation pit, or inadequate drainage.

9. Extract E3_3 when the text indicates seepage, water inrush, water leakage, piping, or sudden water inflow.

10. Extract E4_1 when the text indicates nighttime construction, nighttime operations, or emergency repair work conducted at night.

11. Extract E4_2 when the text indicates adverse weather, heavy rainfall, strong winds, or rainstorms.

12. Extract E4_3 when the text indicates a restricted site, limited working space, restricted rescue space, or difficulty in bringing machinery into the site.

13. Extract B2_1 when the text indicates that no support was installed or that no supporting measures were adopted.

14. Extract B2_2 when the text indicates insufficient support, improperly implemented support, or support that failed to satisfy applicable requirements.

15. Extract B2_3 when the text indicates an excessively steep slope, inadequate slope grading, or failure to grade the slope as required.

16. Extract B1_1 when the text indicates that no specialized plan was prepared, no specialized construction plan was available, or a specialized design was absent.

17. Extract B1_2 when the text indicates that a design or construction plan did not comply with applicable standards or specifications.

18. Extract B1_3 when the text indicates a design error, design-calculation error, or design defect.

19. Extract B3_1 when the text indicates an improper excavation sequence, over-excavation, unauthorized excavation, or an improper excavation method.

20. Extract B3_2 when the text indicates incorrect slope grading, an improper slope-grading method, or failure to grade the slope as required.

21. Extract B3_3 when the text indicates an improper support method, incorrect support construction, or unreasonable support measures.

22. Extract B4_1 when the text indicates inadequate or absent management, failure to implement safety responsibilities, failure to conduct safety and technical briefings, failure to appoint safety-management personnel, failure to conduct safety inspections, failure to stop unsafe behavior promptly, or disorganized on-site management.

23. Extract B4_2 when the text indicates that monitoring was not conducted, monitoring arrangements were not implemented, monitoring instruments were not installed, or deformation monitoring was not performed.

24. Extract B4_3 when the text indicates that a warning was not addressed, identified hazards were not rectified, detected problems were not handled, or monitoring abnormalities were not addressed.

25. Extract B5_1 when the text indicates entry into a dangerous area, unauthorized or risky entry into the foundation pit, or entry into a hazardous slope area.

26. Extract B5_2 when the text indicates unprotected work, failure to wear personal protective equipment, or the absence of safety-protection measures.

27. Extract B5_3 when the text indicates unsafe operations, noncompliant operations, improper instructions, unlicensed operations, or unauthorized work.

28. Extract B6_1 when the text indicates improper permitting or approval, failure to obtain the required permit, improper issuance of a permit, or inadequate approval that allowed an unqualified organization to undertake construction.

29. Extract B6_2 when the text indicates inadequate supervision, failure of supervision personnel to perform their duties, insufficient supervisory inspection, or inadequate governmental oversight.

30. Extract B6_3 when the text indicates illegal construction, unlicensed construction, construction without the required qualifications, unauthorized construction, resistance to supervision, refusal to rectify identified problems, violation of a stop-work order, or continued construction despite a stop-work requirement.

Note:

The preceding label identification guidelines are used only to support E/B information extraction and must not be used to generate emergency response measures. Emergency response R labels are generated by the subsequent rule-based and case-based reasoning modules.

Accident text:

[User-input case text]

Output strict JSON only. Do not provide any additional text.

References

1. Wu, B.; Zeng, J.; Zhu, R.; Yang, F.; Liu, C.; Xie, Y. A Collapse Risk Assessment Method for Subway Foundation Pit Based on Cloud Model and Improved Dempster–Shafer Evidence Theory. *Sci. Rep.* **2024**, *14*, 2653. [CrossRef]

2. Chin, Y.-T.; Shen, S.-L.; Zhou, A.-N.; Chen, J. Foundation Pit Collapse on 8 June 2019 in Nanning, China: A Brief Report. *Safety* **2019**, *5*, 68. [CrossRef]
3. Zhou, Z.; Liu, S.; Qi, H. Mitigating Subway Construction Collapse Risk Using Bayesian Network Modeling. *Autom. Constr.* **2022**, *143*, 104541. [CrossRef]
4. Xu, X.B.; Hu, Q.; Huang, T.M.; Chen, Y.; Shen, W.M.; Hu, M.Y. Seepage Failure of a Foundation Pit with Confined Aquifer Layers and Its Reconstruction. *Eng. Fail. Anal.* **2022**, *138*, 106366. [CrossRef]
5. Xu, N.; Ma, L.; Liu, Q.; Wang, L.; Deng, Y. An Improved Text Mining Approach to Extract Safety Risk Factors from Construction Accident Reports. *Saf. Sci.* **2021**, *138*, 105216. [CrossRef]
6. Zhang, F.; Fleyeh, H.; Wang, X.; Lu, M. Construction Site Accident Analysis Using Text Mining and Natural Language Processing Techniques. *Autom. Constr.* **2019**, *99*, 238–248. [CrossRef]
7. Gruber, T.R. Toward Principles for the Design of Ontologies Used for Knowledge Sharing? *Int. J. Hum.-Comput. Stud.* **1995**, *43*, 907–928. [CrossRef]
8. Lu, Y.; Li, Q.; Zhou, Z.; Deng, Y. Ontology-Based Knowledge Modeling for Automated Construction Safety Checking. *Saf. Sci.* **2015**, *79*, 11–18. [CrossRef]
9. Ding, L.Y.; Zhong, B.T.; Wu, S.; Luo, H.B. Construction Risk Knowledge Management in BIM Using Ontology and Semantic Web Technology. *Saf. Sci.* **2016**, *87*, 202–213. [CrossRef]
10. Aamodt, A.; Plaza, E. Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. *AI Commun.* **1994**, *7*, 39–59. [CrossRef]
11. Su, Y.; Yang, S.; Liu, K.; Hua, K.; Yao, Q. Developing A Case-Based Reasoning Model for Safety Accident Pre-Control and Decision Making in the Construction Industry. *Int. J. Environ. Res. Public Health* **2019**, *16*, 1511. [CrossRef]
12. Lu, Y.; Yin, L.; Deng, Y.; Wu, G.; Li, C. Using Cased Based Reasoning for Automated Safety Risk Management in Construction Industry. *Saf. Sci.* **2023**, *163*, 106113. [CrossRef]
13. Chen, M.; Tao, Z.; Tang, W.; Qin, T.; Yang, R.; Zhu, C. Enhancing Emergency Decision-Making with Knowledge Graphs and Large Language Models. *Int. J. Disaster Risk Reduct.* **2024**, *113*, 104804. [CrossRef]
14. Agrawal, G.; Kumarage, T.; Alghamdi, Z.; Liu, H. Can Knowledge Graphs Reduce Hallucinations in LLMs?: A Survey. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*; Duh, K., Gomez, H., Bethard, S., Eds.; Association for Computational Linguistics: Mexico City, Mexico, 2024; pp. 3947–3960. [CrossRef]
15. Ji, Z.; Lee, N.; Frieske, R.; Yu, T.; Su, D.; Xu, Y.; Ishii, E.; Bang, Y.J.; Madotto, A.; Fung, P. Survey of Hallucination in Natural Language Generation. *ACM Comput. Surv.* **2023**, *55*, 248. [CrossRef]
16. Johansen, K.W.; Schultz, C.; Teizer, J. Knowledge Graph Exploitation to Enhance the Usability of Risk Assessment in Construction Safety Planning. *Adv. Eng. Inform.* **2025**, *65*, 103305. [CrossRef]
17. Wang, Y.; Luo, H.; Fang, W. An Integrated Approach for Automatic Safety Inspection in Construction: Domain Knowledge with Multimodal Large Language Model. *Adv. Eng. Inform.* **2025**, *65*, 103246. [CrossRef]
18. Shim, M.; Choi, H.; Koo, H.; Um, K.; Lee, K.-H.; Lee, S. OmEGa(Ω): Ontology-Based Information Extraction Framework for Constructing Task-Centric Knowledge Graph from Manufacturing Documents with Large Language Model. *Adv. Eng. Inform.* **2025**, *64*, 103001. [CrossRef]
19. Meng, W.; Zhang, H.; Ai, Q.; Bao, T.; Yan, J. CBR-RBR Fusion Based Parametric Rapid Construction Method of Bridge BIM Model. *Adv. Eng. Inform.* **2023**, *57*, 102086. [CrossRef]
20. Huo, X.; Yin, Y.; Jiao, L.; Zhang, Y. A Data-Driven and Knowledge Graph-Based Analysis of the Risk Hazard Coupling Mechanism in Subway Construction Accidents. *Reliab. Eng. Syst. Saf.* **2024**, *250*, 110254. [CrossRef]
21. Goh, Y.M.; Ubeynarayana, C.U. Construction Accident Narrative Classification: An Evaluation of Text Mining Techniques. *Accid. Anal. Prev.* **2017**, *108*, 122–130. [CrossRef]
22. Zhong, B.; Pan, X.; Love, P.E.D.; Ding, L.; Fang, W. Deep Learning and Network Analysis: Classifying and Visualizing Accident Narratives in Construction. *Autom. Constr.* **2020**, *113*, 103089. [CrossRef]
23. Qiao, J.; Wang, C.; Guan, S.; Shuran, L. Construction-Accident Narrative Classification Using Shallow and Deep Learning. *J. Constr. Eng. Manag.* **2022**, *148*, 04022088. [CrossRef]
24. Baker, H.; Hallowell, M.R.; Tixier, A.J.-P. Automatically Learning Construction Injury Precursors from Text. *Autom. Constr.* **2020**, *118*, 103145. [CrossRef]
25. Tixier, A.J.-P.; Hallowell, M.R.; Rajagopalan, B.; Bowman, D. Automated Content Analysis for Construction Safety: A Natural Language Processing System to Extract Precursors and Outcomes from Unstructured Injury Reports. *Autom. Constr.* **2016**, *62*, 45–56. [CrossRef]
26. Chen, W.T.; Bria, T.A. A Review of Ontology-Based Safety Management in Construction. *Sustainability* **2023**, *15*, 413. [CrossRef]
27. Doukari, O.; Wakefield, J.; Martinez, P.; Kassem, M. An Ontology-Based Tool for Safety Management in Building Renovation Projects. *J. Build. Eng.* **2024**, *84*, 108609. [CrossRef]

28. Chi, N.-W.; Lin, K.-Y.; Hsieh, S.-H. Using Ontology-Based Text Classification to Assist Job Hazard Analysis. *Adv. Eng. Inform.* **2014**, *28*, 381–394. [CrossRef]
29. Wang, H.-H.; Boukamp, F. Ontology-Based Representation and Reasoning Framework for Supporting Job Hazard Analysis. *J. Comput. Civ. Eng.* **2011**, *25*, 442–456. [CrossRef]
30. Pedro, A.; Pham-Hang, A.-T.; Nguyen, P.T.; Pham, H.C. Data-Driven Construction Safety Information Sharing System Based on Linked Data, Ontologies, and Knowledge Graph Technologies. *Int. J. Environ. Res. Public. Health* **2022**, *19*, 794. [CrossRef]
31. Li, W.; Wu, P.; Huang, J.; Xu, Y. A New Paradigm for Construction Safety Management in China: Introducing Knowledge Graph and Accident Database into the Early-Stage of BIM. *J. Clean. Prod.* **2024**, *470*, 143367. [CrossRef]
32. Kong, F.; Ahn, S. Use of Knowledge Graphs for Construction Safety Management: A Systematic Literature Review. *Information* **2024**, *15*, 390. [CrossRef]
33. Horrocks, I.; Peter, F.P.; Boley, H.; Tabet, S. SWRL: A Semantic Web Rule Language Combining OWL and RuleML. Available online: https://www.w3.org/submissions/SWRL/ (accessed on 17 May 2026).
34. Ma, L.; Zhao, X.; Jiang, R.; Wu, C.; Liao, L.; Yang, Z.; Tan, J. Adopting Large Language Models in the Construction Industry: Drivers, Barriers, and Strategic Implications from China. *Buildings* **2025**, *15*, 4296. [CrossRef]
35. Kampelopoulos, D.; Tsanousa, A.; Vrochidis, S.; Kompatsiaris, I. A Review of LLMs and Their Applications in the Architecture, Engineering and Construction Industry. *Artif. Intell. Rev.* **2025**, *58*, 250. [CrossRef]
36. Sun, X.; Chen, T.; Huang, Q.; Yang, X.; Lv, Y. Study on Comprehensive Risk Assessment of Marine Environment Safety Based on Public Safety Triangle Theory. *E3S Web Conf.* **2018**, *53*, 03028. [CrossRef]
37. Reason, J. Human Error: Models and Management. *BMJ* **2000**, *320*, 768–770. [CrossRef]
38. U.S. Department of Energy. *Root Cause Analysis Guidance Document*; U.S. Department of Energy: Washington, DC, USA, 1992.
39. *ISO 22320:2018*; Security and Resilience—Emergency Management—Guidelines for Incident Management. ISO: Vernier, Switzerland, 2018. Available online: https://www.iso.org/standard/67851.html (accessed on 13 May 2026).
40. *JGJ 120-2012*; JGJ 120-2012 Technical Specification for Retaining and Protection of Building Foundation Excavations (English)—Code of China. China Architecture & Building Press: Beijing, China, 2012. Available online: https://www.codeofchina.com/standard/JGJ120-2012.html (accessed on 13 May 2026).

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