Collaborative Safety Intelligence in Construction Sites through Agentic RAG Systems

Alessandra Corneli[1], Tommaso Pieroni[1], Alessandro Carbonari[1] and Berardo Naticchia[1]

[1] Polytechnic University of Marche, Department of Civil and Building Engineering and Architecture, Italy
a.corneli@staff.univpm.it, t.pieroni@pm.univpm.it, alessandro.carbonari@staff.univpm.it, b.naticchia@staff.univpm.it

Abstract –

Construction site safety management is challenged by the difficulty of accessing and applying safety procedures during on-site activities, particularly in dynamic and multicultural environments. While Large Language Models offer new opportunities for information retrieval, their use in safety-critical construction contexts raises concerns related to the risk of hallucination, source reliability and contextual relevance. Retrieval-Augmented Generation (RAG) addresses these limitations by grounding responses in controlled, domain-specific sources. In this context, this research proposes an agentic RAG system integrated into a wearable smart-helmet to provide contextualized, real-time safety support to workers on site. Experimental evaluation demonstrated strong retrieval accuracy for safety-critical numerical parameters and effective proactive clarification, with 95% precision and recall in ambiguity detection and correct responses in 88% of cases after clarification.

Keywords –

Artificial Intelligence, Retrieval-Augmented Generation, Construction Safety, Agentic AI, Construction Management, Human–AI collaboration, On-site decision support, Smart wearable device.

1 Introduction

A relevant aspect of construction site safety management concerns the correct execution of operational procedures [1], which is closely related to workers' level of knowledge and their ability to access and interpret safety-related information during on-site activities. This remains challenging in dynamic construction environments characterized by real-time operations and time constraints. In this context, conversational interface–based systems, capable of providing rapid and contextualized access to reference documentation, may represent an effective support tool for construction workers [2].

Artificial intelligence plays a key role in enabling such systems. While numerous AI applications have been developed for construction safety, most existing approaches rely on Deep Learning and Computer Vision, particularly CNN-based models and YOLO frameworks, widely used for PPE detection, proximity monitoring, and unsafe behavior recognition. In parallel, traditional Machine Learning techniques have been applied to structured data, such as accident records and sensor measurements. More recently, Natural Language Processing (NLP) and Large Language Models (LLMs) have been applied to incident analysis, regulatory documentation, and textual safety logs; however, these approaches remain less prevalent than vision-based solutions in construction safety applications [3]. Despite these advances, existing systems operate passively without supporting workers in actively querying and interpreting safety documentation on demand. Furthermore, direct use of general-purpose LLMs introduces hallucination risk, which is unacceptable in safety-critical contexts.

Building on these premises, this research proposes an agentic Retrieval-Augmented Generation (RAG) system designed to support workers in safety-related matters during construction activities. The proposed conversational interface proactively requests clarifications when queries are insufficiently detailed, progressively reconstructing operational context before generating a response. The system is activated exclusively at the worker's initiative, ensuring interaction that does not interfere with task execution. The main contributions of this paper are: (i) an agentic RAG pipeline with proactive clarification for safety-critical queries; (ii) a dual ingestion strategy for heterogeneous document types; and (iii) an evaluation protocol assessing retrieval traceability and clarification effectiveness.

2 Literature Review

2.1 AI-Based Systems for Construction Site Safety Management

The adoption of AI-based systems for construction site safety management has increased significantly in recent years, aiming to enhance traditional safety practices through proactive risk identification, real-time monitoring, and data-driven decision-making. A systematic review by Akhanova et al. highlights the widespread use of AI for safety training, hazard recognition, and behavioral analysis, contributing to improved safety awareness and compliance among construction workers [4].

Most existing applications rely on Machine Learning and Deep Learning integrated with Computer Vision, enabling automated detection of unsafe behaviors, PPE compliance monitoring, and hazard identification through video surveillance, drones, and sensors [5]. AI has also been applied to predictive safety management, where historical accident data and operational parameters are analyzed to anticipate risks and support preventive actions. The integration of AI with Building Information Modelling (BIM) further enhances safety planning and compliance checking by enabling contextualized risk analysis within digital site representations [6]. At a broader scale, AI-based safety solutions are increasingly framed within the paradigm of intelligent construction, which combines AI, IoT, and data analytics to improve safety, efficiency, and automation throughout the construction lifecycle [7]. Despite these advances, conversational AI systems grounded in domain-specific documentation remain underexplored in construction safety management, leaving a critical gap in on-site, real-time decision support.

2.2 LLM and RAG based Approaches in Construction

Addressing this gap requires moving beyond detection-oriented approaches toward systems capable of interpreting regulatory language and procedural documentation in context. Recent advances in LLMs have enabled more natural interaction with construction data, supporting document analysis, knowledge retrieval, and decision support across the building lifecycle. In the AEC domain, LLMs have been applied to facilitate access to unstructured information such as safety regulations, technical documentation, and project reports, reducing cognitive load and improving information accessibility [8]. However, general-purpose LLMs present limitations in safety-critical contexts, particularly due to hallucinations and limited contextual grounding.

To address these challenges, RAG has been adopted to ground model responses in authoritative, domain-specific sources, improving reliability and contextual relevance. RAG-based approaches have demonstrated superior performance in safety knowledge retrieval and regulatory compliance compared to standalone LLMs [8]. More recently, agentic RAG architectures have been proposed, enabling autonomous management of retrieval strategies and multi-step reasoning. Studies have shown that agentic RAG systems integrated with BIM and operational data can deliver context-aware insights through natural language interfaces [9], while supporting adaptability and human-in-the-loop interaction in complex engineering workflows [10]. Additionally, RAG-enhanced LLMs have been applied to natural language querying of semantic building models, significantly lowering technical barriers and improving access to BIM-based information during building operations and management[11].

3 Methodology

3.1 System Conceptualization

The research step reported in this paper is a work package of a wider project, which is targeted to the delivery of a fully-fledged safety helmet assisting construction workers on-site (Figure 1). Basically, workers will be equipped with safety helmets embedding sensors (e.g. cameras and environmental monitoring sensors) and a multilingual voice interface allowing workers to send queries and get responses in the operator's native language without interrupting task execution. Helmets will incorporate not only the audio module (i.e. speakers and microphone) but also a communication module connected with either an embedded PC or the worker's smartphone (e.g. via Bluetooth LE) to transfer audio data and other real-time data (e.g. positional information). This will support vocal interaction and data exchange with a cloud system based on a micro-service architecture, hosting AI Reasoning tools to provide workers with assistance. The cloud system will integrate a common data environment (CDE) to manage information, too. The CDE in the cloud platform must

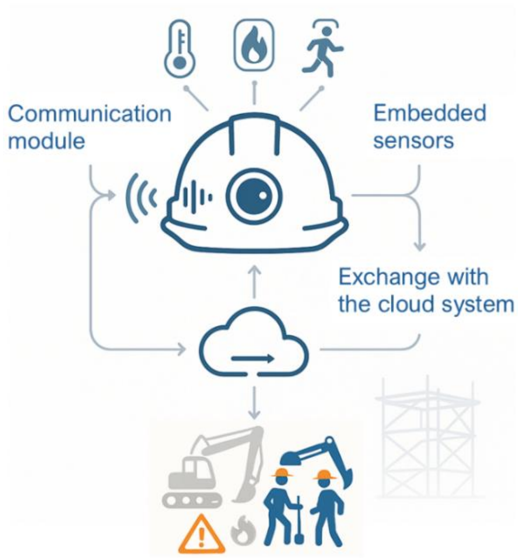

Figure 1. Conceptual scheme of the safety management system enabled by the safety helmet.

manage and visualize heterogeneous data, combining BIM and other information models, as well as related documents concerning safety management, such as health and safety plans and site-specific plans contextualized to types of activities and projects. This entails the identification of BIM data relevant to construction operations, pictures and point clouds depicting the actual progress of construction works.

A wireless sensor network based on Low-Power Wireless Protocols (LPWP) will monitor operators, machinery and protective equipment, with data collected and filtered through wired gateways and stored in the CDE. As a result, the overall system will establish a human-machine interface between workers and AI-based agents, augmenting workers' capabilities concerning the detection and mitigation of hazards, by delivering contextualized, accessible information through LLMs embedded in an automatic workflow.

The present paper focuses specifically on the development of the AI agent and its agentic RAG capabilities, while hardware integration and field deployment are the subject of ongoing work. The development environment is described in Section 3.2. Section 4 tackles the agent's core challenges, such as retrieving accurate, contextually appropriate information from a heterogeneous document corpus spanning from the health and safety plans to specific safety protocols, as well as technical data about the building under construction.

3.2 Visual Programming Tools

The system was implemented using n8n, an open-source workflow automation platform that provides a graphical interface where workflows are constructed by connecting modular nodes. Each node represent discrete operations (data transformations, API calls, conditional logic, database queries), as shown in Figure 2. This visual programming approach facilitated rapid prototyping and iteration, allowing modifications to retrieval parameters, chunking strategies, and data source integration without code refactoring. It also enabled seamless low-code integration of LangChain components (vector stores, embeddings, agents, memory), external APIs (OpenAI, Cohere), and databases (PostgreSQL with pgvector), allowing development efforts to focus primarily on system architecture and construction safety logic rather than integration details.

In addition, n8n provides transparent execution logs that expose data flow, response times, and errors, which were instrumental in diagnosing retrieval failures and tuning system parameters while developing the workflow. In this research, n8n was used for rapid prototyping; however, it natively supports production-grade features (e.g., queue-based execution, horizontal scaling, fault tolerance), making it a viable platform in preparation of future operational deployment.

4 Prototype Development

4.1 Document Ingestion Pipeline

The file ingestion pipeline is automatically triggered when a document is uploaded to the Google Drive folder (connected via API keys). The ingestion pipeline adopts a dual strategy based on document type. For textual documents (PDF, DOCX, TXT), the system uses an LLM-based semantic chunking approach with configurable minimum and maximum chunk sizes. Smaller documents are chunked between 400 and 1,000 characters, while large documents (e.g., more than 100 pages) use larger chunks ranging from 1,500 to 3,000 characters. For each segment, GPT-4o-mini (selected for its favourable cost-performance trade-off) identifies natural topic boundaries (section headers, paragraph breaks, etc.) and outputs a coherent split point, ensuring complete sentences and semantically consistent chunks. Undersized chunks are automatically merged with adjacent ones when within the defined limits, preserving contextual integrity for safety-critical queries. While semantic chunking requires additional computational resources compared to fixed-size splitting, this trade-off is essential in safety-critical applications where fragmenting procedural steps or separating numerical parameters from their context could lead to dangerous misinterpretation during retrieval. For tabular data (CSV, XLSX), the system adopts a different strategy: rows are stored as JSONB objects in a PostgreSQL relational table (document_rows), enabling precise SQL queries for calculations and structured data analysis.

4.2 Vector Database and Embedding Strategy

Document chunks are embedded using OpenAI's text-embedding-3-small model and stored in PostgreSQL with pgvector. Text-embedding-3-small was selected due to its strong cost-performance trade-off ($0.02/1M tokens), offering reliable semantic embeddings for retrieval tasks at substantially lower cost than larger models. The model produces embeddings with 1536 dimensions. A metadata table (document_metadata) maintains references to source documents, including file IDs, titles, URLs, and the complete schema definition for tabular files. This metadata layer enables the agent to understand available data structures before formulating queries.

The vector store implements a two-stage retrieval process with reranking. Initial retrieval uses cosine similarity to fetch the top 25 most relevant chunks (k=25), which are then reranked using Cohere's rerank-v3.5 model to select the 4 most contextually appropriate results. Cohere rerank-v3.5 was selected for its cross-encoder architecture, which evaluates chunk relevance against the query in a contextually integrated manner.

4.3 Agentic Reasoning and Tool Orchestration

The prototype implements an agentic architecture capable of adaptive tool selection based on query characteristics. The agent is powered by OpenAI's GPT-4o-mini and has access to four distinct tools: (i) “Postgres PGVector Store” for semantic search across chunked documents, (ii) “List Documents” to query metadata and schemas, (iii) “Query Document Rows” for SQL-based numerical retrieval, and (iv) “Postgres Chat Memory” to store conversation history. Tool selection follows a query-driven policy: semantic queries trigger the PGVector Store, schema or provenance questions trigger List Documents, and numerical queries trigger Query Document Rows. For example "What is the maximum load capacity for scaffolding on the east facade?" triggers a metadata lookup followed by semantic retrieval, while "What is the average number of safety incidents per month?" prompts a direct SQL query on tabular data.

This multi-tool approach enables the system to handle complex queries requiring both structured and unstructured data and cross-referencing information across multiple documents, a capability crucial given the nature of construction site documentation. When relevant information spans multiple documents or a query is complex, the agent iteratively invokes retrieval tools to aggregate content across sources. The LLM-based semantic chunking strategy (Section 4.1) supports this by preserving thematic coherence at ingestion time, reducing fragmentation of related information across distant chunks.

4.4 Automatic Update Management

A critical operational requirement for construction sites is ensuring documentation currency. When a file is updated or deleted in the monitored Google Drive folder, the automated workflow: (i) delete obsolete vector embeddings associated with the old file ID from “documents_pg” table, (ii) remove outdated tabular rows from “document_rows”, (iii) update or remove metadata entries in “document_metadata” table and (iv) re-process the updated file through the ingestion pipeline.

This synchronization runs every 15 minutes, preventing the agent from citing outdated safety procedures or obsolete equipment specifications.

5 Evaluation Methodology

Evaluation focused on operational accuracy for safety-critical applications, where factual correctness takes precedence over linguistic fluency. The agentic RAG architecture was selected because it enables dynamic tool selection based on query characteristics, and because construction site operations require access to project-specific documentation, structurally absent from any general-purpose model's training data. For general

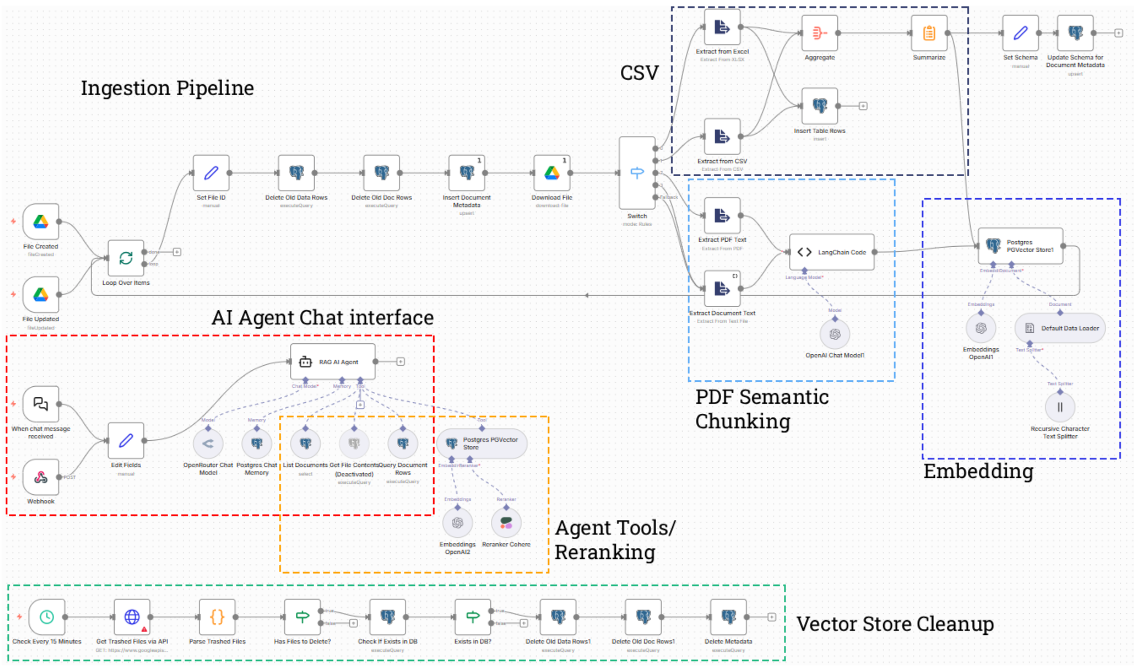

Figure 2. Functional Clusters of the Agentic RAG Model on n8n.

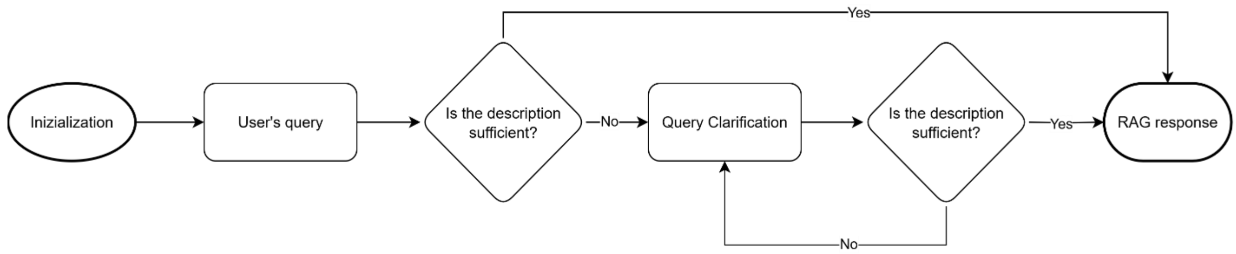

Figure 3. Conversation pipeline.

regulatory queries, Lee et al. [8] already provides external evidence of RAG superiority over standalone LLMs in the construction safety domain. We assessed: (i) retrieval traceability to expected source documents, (ii) exact-match accuracy for numerical safety parameters and (iii) proactive clarification effectiveness. The query clarification part (Figure 3) is essential to prevent hallucinations and to enable the model to reconstruct the operator's context at a sufficient level to provide accurate answers grounded in site documentation, since incomplete information could lead to dangerous guidance. Accordingly, we designed two types of experiments to evaluate the RAG system: (i) Retrieval Performance Assessment and (ii) Proactive Clarification Effectiveness Evaluation.

The RAG corpus comprised 7 documents (5-140 pages, 318 total) including operational safety guidelines (Trento Safety Notebook), electrical hazard manual, cement Material Safety Data Sheets (SDS), scaffolding assembly, use, and dismantling plan (PiMUS), and dense regulatory text (Italian D.Lgs. 81/08 – Annex XVIII). The dataset contained PDF files with formatted text, images, and tables, plus CSV spreadsheets. No preprocessing or refinement was performed in order to test extraction and elaboration of project-specific numerical values under a realistic scenario, where existing site documentation is used without modification.

5.1 System Prompt Refinement

The model is instructed to behave as a construction safety management expert, providing actionable guidance grounded exclusively in retrieved site documentation.

To ensure reliability, the prompt enforces: mandatory document retrieval prior to response generation, SQL restriction to numerical/tabular data only, internet access disabled by default, explicit prohibition of information fabrication and mandatory clarification requests when essential contextual information is missing. In cases where conflicting values may appear in different regulatory documents, the system is designed to surface the retrieved information without autonomous arbitration, leaving conflict resolution to the human operator. This is a conservative design choice appropriate for safety-critical contexts. Responses are constrained to short professional outputs and the model is told to respond in the operator's language.

5.2 Retrieval Performance Assessment

This evaluation assessed whether the system consistently locates information in expected source documents and extracts quantitative safety parameters with exact precision, since numerical errors in distances, time limits, or load capacities are critical requirements that could have severe consequences.

We developed 30 operational safety queries distributed across four categories: Equipment procedures (n=8), Exposure limits (n=7), PPE selection (n=4), and Scaffolding/temporary works (n=11). Queries were formulated in natural language as workers would pose them, without technical optimization. All queries were in English despite primarily Italian source documents, testing cross-lingual retrieval capability. For each query, we annotated: (i) expected source document, (ii) expected page/section, and (iii) expected numerical values (if applicable). Performance was assessed using two metrics:

1. Source Traceability Accuracy (STA): Binary assessment of whether the system retrieved information from the expected document. A query was marked correct if the response cited or demonstrably used content from the expected source (document and page/section).
2. Numerical Precision (NP): For queries requiring quantitative answers, exact-match accuracy for numerical values and units.

$$STA = \frac{Correct\ source\ retrievals}{Total\ queries} \quad (1)$$

$$NP = \frac{Exact\ numerical\ matches}{Queries\ requiring\ numerical\ data} \quad (2)$$

5.3 Proactive Clarification Effectiveness Evaluation

We evaluated clarification behaviour using 40 queries structured as controlled pairs: 20 intentionally ambiguous queries missing critical safety parameters (equipment specifications, product identifiers, operational contexts, or location details) and 20

Table 1. Structure of the spreadsheet used to gather retrieval accuracy assessment's data.

Query text	Expected doc (page)	Retrieved doc (page)	Traceability	Expected values	Extracted values	Numerical Precision	Response text
1	PiMUS(9)	PiMUS(9)	✓	20 cm	20 cm	✓	2
3	Electrical_ Hazard(14)	Safety_ Trento(73)	×	IP44, IP21	IP44	×	4
5	PiMUS(5)	PiMUS(-)	×	-	-	-	6

[1] How high should the base plates of the uprights of the 13th scaffolding section on the south side be?
[2] The bases of the uprights of the 13th column on the south side must be adjusted to 20 cm.
[3] What is the minimum degree of protection for an electrical panel?
[4] The minimum degree of protection required for construction site electrical panels (ASC) is IP44 with the door closed, as indicated by CEI EN 61439-4 and the 'QuadernoDiCantiere2023Trento' (Trento Construction Site Notebook 2023).1
[5] Can you tell me the name of the site manager for scaffolding assembling?
[6] The PiMUS document does not indicate the name of the site manager. Would you like me to search other files?

corresponding fully-specified clear queries.

Each "ambiguous-clear" pair differed only in the presence of the missing parameters, enabling direct assessment of the system's ability to distinguish between adequate and insufficient query specifications. For each query, the following were recorded: whether a clarification request was issued, whether clarification was objectively required, the content of the clarification question, and its relevance, rated by human using a 0-3 "Relevance Score" (RS) (0 = No request; 1 = Off topic; 2 = Generic; 3 = Highly specific).

A simulated operator response was then provided, and the final answer was assessed for accuracy. Performance was assessed using the following metrics:

$$Precision = TP/(TP + FP) \quad (3)$$

$$Recall = TP/(TP + FN) \quad (4)$$

$$CRS\_mean = \sum RS/N\ clarification \quad (5)$$

In this context, a true positive (TP) corresponds to an ambiguous query that correctly triggered a clarification request, while a false positive (FP) indicates an unnecessary clarification for a clear query. A false negative (FN) occurs when the system responds directly to an ambiguous query without requesting clarification, and a true negative corresponds to a clear query answered without clarification.

5.4 Evaluation Protocol

To ensure query independence, the PostgreSQL chat history was truncated before each test query, simulating a worst-case deployment scenario in which workers pose isolated questions without prior conversational context. Response latency was recorded for each query across both experiments.

6 Results

6.1 Results of Retrieval Performance Assessment

Table 1 reports three representative tests out of thirty. Failure cases shows that, in the second reported test, the information was retrieved from an unexpected source. The answer is partially correct but lacks an additional detail that was present in the intended reference document. In the third test, the model retrieved the correct document but failed to locate the required information. The system reported the failure instead of hallucinating.

Table 2 reports the aggregated results for the full set of 30 executed queries.

Table 2. Results of Retrieval Performance Assessment

Category	n	Retr. Acc.	Num. Prec.	Avg Latency
Equipment	8	7/8	5/6	18,82s
Exposure	7	7/7	3/3	18,94s
PPE	4	4/4	1/1	21,13s
Scaffolding	11	10/11	8/8	22,76s
OVERALL	30	28/30	17/18	20,60s

The overall STA for the evaluated sample was 93%, while numerical precision reached 94%. The total cost of executing this test was €0.30, calculated directly from the API key usage records.

6.2 Results of Proactive Clarification Evaluation

Table 3 presents three representative queries out of

Table 3. Structure of the spreadsheet used to gather clarification requests data

Query text	Clar. necessary	Clar. requested	Clar. question	Relevance Score	User Clar.	Final Response	Response Correctness
[1]	✓	✓	[2]	3	[3]	[4]	✓
[5]	✓	✗	-	0	-	[6]	✗
[7]	✗	✓	[8]	3	[9]	[10]	✓

[1] Can I use this cable reel?

[2] To assess suitability, indicate: CEI EN 61316 marking, presence of thermal protector, cable type (H07RN-F or equivalent) and cross-section, expected capacity (16/32/63 A) and whether the cable will be completely or partially wound on the drum.

[3] The cable is type H07RN-F, the cross-section is 2 mm², the current is 16 A, the marking is CEI EN 61316, and there is a thermal protector.

[4] No; for a 16 A cable reel, the standard requires a minimum cross-section of 2.5 mm², so a 2 mm² cable is not suitable.

[5] May I store materials on a scaffold?

[6] It depends on the load capacity specified in the project: typically, construction scaffolding is designed for 300 kg/m² and maintenance scaffolding for 150 kg/m², so always check the calculation report/PiMUS and the manufacturer's instructions before storing materials.

[7] Minimum width of scaffolding decking?

[8] Do you mean the minimum width of the individual boards of the decking or the overall clear width of the deck/walkway?

[9] Overall width.

[10] The overall width of the deck must not be less than 90 cm.

the 40 executed. It is evident that the clarification requests issued by the model are highly specific, enabling the reconstruction of sufficient context to provide an answer as accurate as possible based solely on the operator's responses. The first example demonstrates high-quality clarification for equipment specification deficiency. When presented with the ambiguous query "Can I use this cable reel?", the system correctly identified the absence of critical safety parameters and requested specific technical details.

It then proceeded to accurately determine non-compliance, stating that 16A reels require minimum 2.5mm² cross-section according to standards. This interaction pattern exemplifies the system's ability to reconstruct sufficient operational context by addressing information gaps directly. The second query is considered a failure because the model posed a too-generic question instead of requesting missing details in an ambiguous query, even though the reported values were correct. Ideally, it would have been able to infer the load class of the scaffolding from the PiMUS and respond accordingly. This highlights a limitation in the system's ability to recognize when generic regulatory values require project-specific contextualization. The third query represents the only case of over-clarification. While the model doubt is technically valid, domain experts considered the query sufficiently clear given construction site vernacular typically refers to overall deck width when using the term "decking." This case underscores the challenge of calibrating clarification thresholds to match domain-specific linguistic conventions even though the query was considered sufficiently clear.

Table 4 reports the aggregated results for the full set of 40 executed queries. A 2.8/3.0 relevance score indicates highly specific clarification questions. Analysis of the 20 clarification requests shows that 16 queries (80%) received maximum RS, explicitly naming the missing parameters. After clarification, the system provided correct answers in 88% of cases. The two post-clarification failures involved: (i) a vibration exposure query where the system retrieved noise exposure limits instead of vibration data, indicating a retrieval error rather than clarification failure; and (ii) a scaffolding configuration query where the system provided partial information. These failures highlight that clarification effectiveness depends on both successful clarification and subsequent accurate retrieval, two coupled but distinct system capabilities.

Table 4. Results of Proactive Clarification Evaluation

Metric	Value
CRR_ambiguous	95% (19/20)
Over-clarification rate	5%(1/20)
CRR_mean	2.8

Precision	95%
Recall	95%
Correct Response	88%

Performance varied by ambiguity type: detection was perfect for missing equipment specifications and product identifiers, while lower accuracy was observed for missing operational context and location/configuration, suggesting greater difficulty in handling spatial and contextual ambiguities that require domain-specific inference about what constitutes "sufficient" location specification.

7 Conclusions and Discussion

This study proposed a RAG system to support construction workers in safety-related tasks through a conversational interface. Unlike conventional RAG approaches, the system proactively requests targeted clarifications when user queries present information gaps, preventing unsafe assumptions. Experimental results showed strong performance, with 95% precision and recall in clarification detection, a low over-clarification rate (5%), and correct responses in 88% of cases after clarification. The system demonstrated non-hallucination behaviour by explicitly reporting retrieval failures rather than generating unsafe guidance. Nevertheless, some limitations remain. The average response latency of approximately 20 seconds is currently unsuitable for seamless voice-based interaction; addressing this in operational deployment might require migration to a lighter conversational stack, with asynchronous processing and response caching as candidate strategies.

Future work includes comparative testing with non-agentic RAG systems, field validation, ablation analysis of individual system components, scalability testing on larger and more heterogeneous document corpora, Optical Character Recognition integration for technical drawings, photos and documents, and knowledge graph augmentation to improve cross-document reasoning. Multi-turn and multimodal interaction capabilities will also be explored to further align the system with realistic on-site usage patterns, alongside dedicated evaluation of its multilingual capabilities.

Acknowledgements

This research was partially funded by the PRIN 2022 project DiTHiCe – Digital Twin Development for the Resilient Management of Minor Historical Centres (CUP: E53C24002740006).

References

[1] A. Guglielmi, A. Leva, G. Campo, V. Meloni, P. Conte, and A. Leonardi, INFORTUNI IN EDILIZIA: CARATTERISTICHE, FATTORI CAUSALI, MISURE PREVENTIVE. 2022. [Online]. Available: https://www.inail.it/cs/internet/attivita/dati-e-statistiche/banca--dati-statistica.html

[2] F. Yang, ; Yuan Tian, J. Zhang, and A. M. Asce, “Supporting Construction Worker Well-Being with a Multi-Agent Conversational AI System.” doi: 10.48550/arXiv.2506.07997

[3] S. J. Badhan and R. Samsami, “Artificial Intelligence (AI) in Construction Safety: A Systematic Literature Review,” Nov. 01, 2025, Multidisciplinary Digital Publishing Institute (MDPI). doi: 10.3390/buildings15224084.

[4] G. Akhanova, A. Ashmawi, C. H. Ko, and P. Nguyen, “Enhancing construction safety training using artificial intelligence: existing applications and future directions,” Safety and Reliability, 2025, doi: 10.1080/09617353.2025.2575706.

[5] M. Tamoor, H. M. Imran, D. Imran, and G. Chaudhry, “Revolutionizing Construction Site Safety through Artificial Intelligence,” Journal of Development and Social, vol. 4, no. 3, pp. 2709–6254, doi: 10.47205/jdss.2023(4-III)103.

[6] N. Rane, “Integrating Building Information Modelling (BIM) and Artificial Intelligence (AI) for Smart Construction Schedule, Cost, Quality, and Safety Management: Challenges and Opportunities,” SSRN Electronic Journal, 2023, doi: 10.2139/ssrn.4616055.

[7] L. Zhang, Y. Li, Y. Pan, and L. Ding, “Advanced informatic technologies for intelligent construction: A review,” Eng. Appl. Artif. Intell., vol. 137, Nov. 2024, doi: 10.1016/j.engappai.2024.109104.

[8] J. Lee, S. Ahn, D. Kim, and D. Kim, “Performance comparison of retrieval-augmented generation and fine-tuned large language models for construction safety management knowledge retrieval,” Autom. Constr., vol. 168, Dec. 2024, doi: 10.1016/j.autcon.2024.105846.

[9] M. Arslan, S. Munawar, L. Mahdjoubi, and P. Manu, “Monitoring indoor environmental conditions in office buildings using a sustainable Agentic RAG-LLM system,” Energy Build., vol. 347, Nov. 2025, doi: 10.1016/j.enbuild.2025.116276.

[10] C. Ávila and E. Rivera, “Conceptualising RAG-Driven Agentic AI with Multi-Layer MCP for Seismic Structural Systems,” Dec. 05, 2025. doi: 10.20944/preprints202512.0534.v1.

[11] M. Li, Z. Hu, P. Mohebi, S. Li, and Z. Wang, “Enhancing LLM-based building data query with chain-of-thought, retrieval-augmented generation, and fine-tuning,” Autom. Constr., vol. 182, Feb. 2026, doi: 10.1016/j.autcon.2025.106738.