Developing a Fit-for-Purpose Best Practice Knowledge Handbook using Generative AI

Gongfan Chen[1]*, Michael Pedraza[2], Roya Albaloul[2], Alex Albert[2], and Edward Jaselskis[2]

[1] Dept of Engineering Technology & Constr. Mgmt., University of North Carolina Charlotte, NC, USA
[2] Dept of Civil, Constr. & Environ. Engineering, North Carolina State University, Raleigh, NC, USA

ABSTRACT: The Construction Industry Institute (CII) has produced extensive research shown to deliver cost, schedule, and safety benefits within the construction industry. However, low visibility and limited accessibility to the research have hindered its widespread adoption. This paper introduces a generative artificial intelligence (GenAI) framework designed to enhance the accessibility and utilization of CII research for industry practitioners. The proposed framework employs a Retrieval-Augmented Generation (RAG) approach by integrating CII best practice (BP) research reports into a large language model (LLM) to identify relevant insights. First, a hybrid method combines qualitative analysis with GenAI-driven question-answering to extract critical findings ("golden nuggets") from each BP report. These nuggets are then used to generate detailed action items (DAIs) by LLM. The validation process involves using an LLM judge method complemented by subject matter experts (SMEs) review. Lastly, an executive summary and frequently asked questions (FAQs) were generated for each BP by feeding GNs and DAIs information to LLM. As a result, a comprehensive BP knowledge handbook was generated. This handbook showcases the potential of GenAI in construction knowledge extraction in a passive way. Additionally, this study will also present a prototype where users can interact with the GenAI chatbot in an active way for knowledge harvesting and communication. This study contributes to advancing human-AI interaction in construction knowledge management, offering a scalable and user-friendly solution for bridging the gap between research and practice.

1. INTRODUCTION

The construction industry is currently facing multiple challenges, including escalating material costs, disruptions in the supply chain, and growing demands for efficiency in scheduling and safety. One of the major challenges is directly linked to constrained resources in the 'professional' ranks in both project/construction management. To address these issues, extensive research has been conducted to assist owners and contractors in optimizing project outcomes. For over four decades, the Construction Industry Institute (CII) has been at the forefront of industry research and development, focusing on improving construction processes and methodologies. Through its research, CII has established 17 Best Practices (BPs) designed to enhance industry performance (CII 2025). When effectively implemented, these BPs have been shown to deliver substantial improvements in cost efficiency, scheduling, and safety for both owners and contractors.

However, a significant challenge that impedes CII's extensive research implementation lies in its low visibility and accessibility for industry members, leading to the underutilization of CII BPs (Malhotra 2017). The vast volume of research outputs in different formats complicates the efficient identification and implementation of relevant BPs by member companies. It is significant to filter the "right" information and extract the "key" insights to enhance CII knowledge accessibility and informed decision-making among the teams. This research aims to fill this gap by providing a new CII BP handbook that will increase the CII knowledge visibility for member companies. The main objectives of this research include: (1) design a fit-for-purpose handbook that allows industry members to access key contents in the CII research report for each BP, and (2) investigate a user-friendly interface that can navigate CII research more easily.

This research introduces a generative artificial intelligence (GenAI) framework to harvest CII knowledge. First, a hybrid method, which combines qualitative analysis with a Retrieval-Augmented Generation (RAG) driven GenAI question-answering, is used to extract key insights (we called it "golden nuggets" (GNs)) from each BP report. These GNs are then used to generate detailed action items (DAIs) by the large language model (LLM). Lastly, an executive summary and frequently asked questions (FAQs) were generated for each BP by feeding GN and DAI information to LLM. To ensure the framework captures the right information, the validation process involves using an LLM judge method complemented by subject matter experts (SMEs) review. As a result, a comprehensive BP knowledge handbook was generated. This study will also present a prototype where users can interact with the GenAI chatbot in an active manner for CII knowledge communication.

2. LITERATURE REVIEW

Previous research has shown that applying CII BPs can improve various phases of construction projects (Deshpande et al. 2012). However, a survey conducted by Olumide et al. (2012) revealed significant gaps in the integration and usage frequency of BP knowledge. These challenges are largely attributed to organizational implementation issues. Kim (2014) also emphasized that easy and timely access to CII BP knowledge is critical for successful implementation. To address this, the CII Research Team (RT)-166 developed the (CII 2022a), along with supplementary resources (CII 2009, 2012), to support detailed implementation and actions through summaries and checklists drawn from various CII publications related to specific BP. However, a major limitation of these checklists is that they do not offer a comprehensive and in-depth distillation of the original materials. As a result, additional research and investigation are needed to support practitioners responsible for implementing CII BPs, which often require manual review of relevant subjects and final applicable tools.

A review of the CII accessible database reveals several barriers that hinder the effective distillation and use of CII BP information. First, the CII website's search engine returns multiple layers of content when users input a query, including research publications, tools, presentations, and others, making it difficult to filter relevant documents. Second, these returned documents were typically stored in PDF format with lengthy textual information, requiring extensive manual review to extract useful insights and imposing a significant cognitive burden. Third, during the review process, understanding the content often necessitates reading supplementary documents across multiple file formats from CII (e.g., PDFs, URLs, Excel files). This fragmented experience can lead to navigational difficulties, with users frequently feeling lost in the process of locating critical information despite investing substantial time and effort. Therefore, an integrated tool is needed to enhance the overall applicability and accessibility of CII BP knowledge.

3. RESEARCH METHODOLOGY

3.1 Overview

Figure 1 provides an overview of the research framework. The study is structured into three key stages, with Stage 1 focusing on the generation of comprehensive content for each CII BP research report. Initially, the research team collected 54 BP reports and their associated tools, all in PDF format. The CII Best Practices Guide 166-3 (CII 2022b) served as the foundational reference for summarizing each report, supplemented by the executive summaries available within the reports. Following this, the collected research reports were processed using a RAG framework to extract key insights, termed "golden nuggets" (GNs), from each document. The same approach was then applied to generate a structured summary of the detailed action items (DAIs) associated with each GN. The extracted GNs and DAIs were meticulously mapped to the corresponding life cycle phases to maximize their practical applicability. Additionally, hyperlinks to the original reports and tool web pages were incorporated, ensuring users could access detailed information as needed. To further enhance usability, we use an LLM to produce executive summaries and the top five FAQs for each BP. This process utilized the same RAG framework, integrating all GNs, DAIs, and research abstracts under each BP category.

To validate and refine the extracted content, a two-step evaluation process was implemented. First, an LLM-based evaluation method was employed, where a higher-level LLM (compared to the one used for summarization) acted as a quality control judge to assess the generated content. Subsequently, subject matter experts (SMEs) manually reviewed the extracted insights at multiple levels, providing qualitative feedback to enhance accuracy and reliability.

Once the content was finalized, two deployment approaches were explored:

- Excel-Based Aggregation – All extracted information was compiled into an Excel spreadsheet, enabling structured storage, easy accessibility, and offline usability.
- RAG-Enabled Chatbot – A privacy-preserving chatbot was developed, leveraging the RAG framework to provide an interactive retrieval experience while ensuring local data protection.

These deployment strategies aim to enhance the accessibility and practical adoption of BP handbook, ensuring that practitioners can efficiently retrieve relevant information based on their needs.

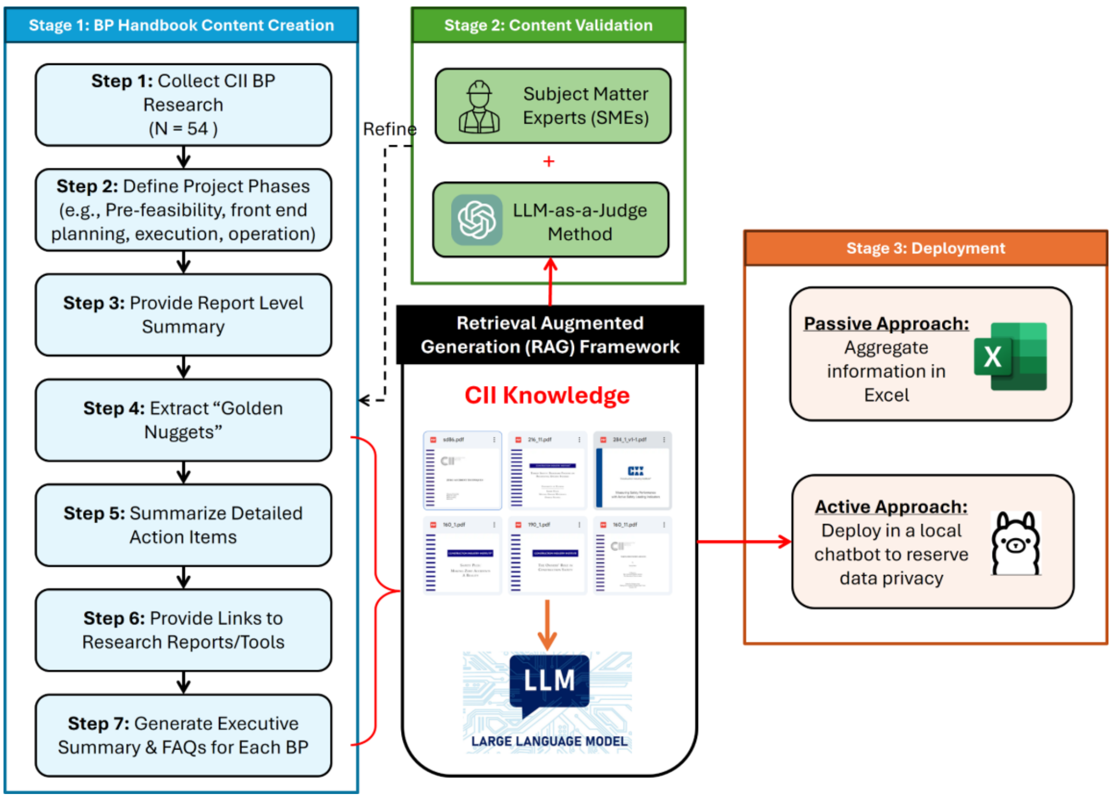

Figure 1. Research framework overview

3.2 Retrieval Augmented Generation (RAG)

This study employs a hybrid methodology that integrates qualitative analysis with GenAI-driven question-answering to systematically extract GNs from BP reports. To enhance the precision and relevance of the extracted insights, we implement retrieval-augmented generation (RAG) as a foundational component of the GenAI-driven question-answering framework. Prior research has demonstrated that RAG-enabled frameworks significantly improve domain-specific knowledge retrieval and communication (Li and Starly 2024, Chen et al. 2025), outperforming traditional fine-tuning approaches in terms of adaptability and performance (Ovadia et al. 2023).

RAG Dataflow: As shown in Figure 2, the RAG framework consists of several key stages. First, data from CII reports undergoes a series of preprocessing steps, including data loading, parsing, and cleaning to ensure a structured format. The processed text is then segmented into sub-chunks, which are subsequently converted into vector representations using an embedding model. These vectorized representations are stored in a vector database for efficient retrieval. Once a user query is received, it undergoes the same preprocessing and vectorization pipeline, transforming it into a query vector. This vector is then compared against the stored vectors using cosine similarity, enabling the system to retrieve the most relevant information from the database. Finally, the retrieved information serves as the context that is fed into the LLM, facilitating summarization and response generation tailored to the user's query.

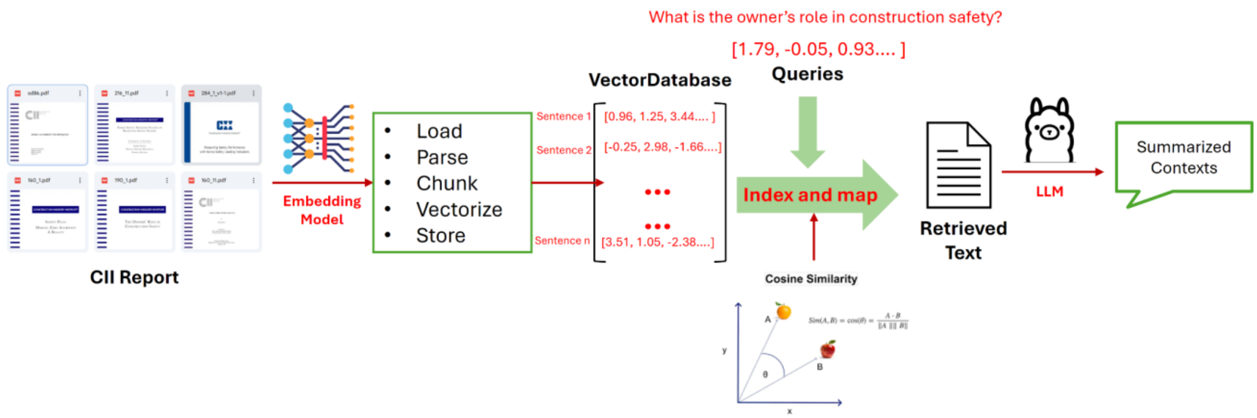

Figure 2. Retrieval augmented generation (RAG) data flow

Prompts Engineering: Prompts are mainly drafted for content summarization and generation. Elaborative prompts played an important role in generating reliable outcomes. This research meticulously crafts detailed prompts to guide LLM behaviors in generating expected outputs.

3.3 Stage 1: BP Handbook Content Creation

Step 1 – Collect CII BP Data: Data were primarily collected from the CII website, including research summary reports and research tool guidance, which are predominantly available in PDF format. The classification of reports as BPs was determined based on the CII Best Practices Guide 166-3 (CII 2022b), the CII website classification, and expert review. In total, 17 BPs encompassing 54 research studies were identified. The distribution of research studies across each BP is presented in Table 1.

Table 1. BP title and number of research for each BP

BP	# of research	BP	# of research
1. Advanced Work Packaging	3	10. Material Management	9
2. Alignment	3	11. Partnering	1
3. Benchmarking & Metrics	2	12. Planning for Modularization	2
4. Change Management	1	13. Planning for Startup	3
5. Constructability	1	14. Project Risk Assessment	4
6. Disputes Prevention & Resolution	1	15. Quality Management	4
7. Front End Planning	3	16. Team Building	3
8. Implementation of CII Research	3	17. Zero Accident Techniques	10
9. Lessons Learned	1	**SUM**	**54**

Step 2 - Define Project Lifecycle Phases: CII categorizes the project lifecycle into four primary phases: Business Case Development, Front-End Planning, Execution, and Operation. Each of these macro phases consists of more granular steps, as depicted in Figure 3 below. Project owners can utilize all GNs throughout the project lifecycle, from the pre-feasibility stage to operation. Additionally, this study identifies three primary contract delivery methods that contractors typically adopt. For instance, when using the design-build delivery method, contractors should focus on GNs beginning from the detailed scope phase through project turnover.

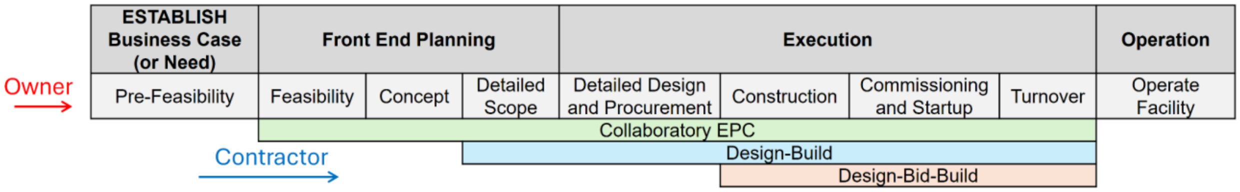

Figure 3. Project lifecycle

Step 3 - Provide Report Level Summary: The process of identifying report-level summaries involved an in-depth review of each document to extract its core themes, objectives, and key findings. During this analysis, the team observed that high-level summaries were commonly found in the executive summary, introduction, or conclusion section of the report. Consequently, the research team manually extracted the relevant summaries from each report.

Step 4 - Extract Golden Nuggets (GNs): Similar to the approach used for report-level summaries, most GNs were directly extracted from the original reports. Additionally, when well-summarized key points were not readily available, GenAI-generated summaries were utilized to supplement the extraction process.

- **Prompts for summarizing GNs:** *Based on the provided file, summarize the key information using no more than five bullet points. Each bullet point sentence should start with a verb, no more than 30 words, concise, complete, and representative. Each bullet sign should be "*".*

Step 5 - Summarize Detailed Action Items (DAIs): After identifying the GNs from each report, the next step was to translate these key insights into actionable tasks to drive implementation. To ensure clarity and effectiveness, it was essential to craft 3-5 specific tasks for each nugget, focusing on concrete actions that could be easily executed. To assist with this process, we once again utilized AI tools, providing clear instructions to generate bullet points that would start each task with a verb.

- **Prompts for summarizing DAIs for each GN:** *Based on the provided file, what would be the detailed action items for "GN Title"? Summarize from the construction project's perspective. You should summarize in no more than five bullet points, each bullet point sentence should start with a verb. Each bullet point should be concise, complete, and representative. Each bullet point should be less than 30 words. The bullet point sign should be "*".*

Step 6 - Provide Links to Research Reports/Tools: Once users identify valuable insights from GNs and DAIs, they may seek additional details for further exploration. To facilitate this, the research team provides direct links to the original research reports and implementation resources available on the CII website, ensuring easy access to comprehensive information as needed. Meanwhile, each tool guidance was also summarized using GenAI.

- **Prompts for summarizing key contents for research tools:** *Help me summarize this document in less than five bullet points. Each bullet point should be concise and representative, starting with a verb. Each bullet point should represent the key takeaways of this report and be less than 25 words.*

Step 7 - Generate Executive Summary & FAQs for Each BP: The research team gathered additional report abstracts by selecting the "Publication" section within each CII BP on the website. This effort aimed to enhance content representativeness for generating a comprehensive executive summary and FAQs. Consequently, the data utilized in the RAG-enabled framework for executive summary and FAQ generation includes research publication abstracts extracted from the website, along with summarized GNs and their corresponding DAIs from previous steps.

- **Prompts for generating an executive summary for each BP:** *Based on the provided file, write me an executive summary for "BP Title" using the "Abstract" textual information. The executive summary should be representative and comprehensive, highlighting the key takeaways we should know for knowledge of "BP Title." Write a cohesive and concise summary. When writing the summaries, also quote which "Research #" it is so I can track sources. Do not separate your writing by each "Research #".*

Prompts for generating FAQs for each BP: *Based on the provided file, help me generate five FAQs for "BP Title" using the "Abstract" textual information. The FAQs should be representative and comprehensive, highlighting the key takeaways we should know for knowledge of "BP Title." After generating the FAQs, provide suggested answers for each FAQ. The answers for each FAQ should be concise. Make sure the answers came from diverse "Research #" to enhance comprehensiveness. Please also list which "Research #" for each answer that you generated so readers can track sources. For example, an expected FAQ and answer format is:*
"FAQ1: xxxx
Answer: xxxx
Source: Research #"

3.4 Stage 2: Content Validation

LLM-as-a-Judge: This research first deployed the LLM-as-a-Judge method to evaluate the quality of summarized GNs and DAIs in previous steps using other LLMs. The LLM-as-a-Judge approach addresses the challenge of scalable and consistent evaluation in text generation, summarization, and reasoning, particularly when human evaluation is resource-intensive and time-consuming (Zheng et al. 2023). This method enables efficient assessment based on predefined rubrics for effective benchmarking. By the time the research was conducted, our study primarily utilized Llama 3 (Meta 2024) for summarization while employing the more advanced Phi-3 model as a judge, given its superior performance over Llama 3 (HuggingFace 2025). Since data privacy is a critical issue for CII, all models were implemented locally. The evaluation process adheres to the RAG framework, wherein each research report is uploaded to the Phi-3 model for assessment. To ensure a structured and rigorous evaluation, a set of carefully designed prompts is utilized:

- **Prompts for Evaluation:** *You are tasked with rating AI-generated summaries of construction research reports based on the given metric. You will be presented with an original research report and AI-generated golden nuggets (key research takeaways) and detailed action items for each golden nugget pulled out from the original research report as input. In the input, the original report is located in the file "original report" while the AI-generated summary is located in the file "summarized contents for the original report".*
  Metric:
    - *Check if the summarized golden nuggets and detail action items are true to the original report.*
    - *The golden nuggets should reflect the key research findings/takeaways majorly discussed in the original report.*
    - *The detailed action items should be a good summary for implementing each golden nugget.*
    - *The summary should be concise.*

  Evaluation criteria: The task is to judge the extent to which the metric is followed by the summary.
    - *1 - The metric is not followed at all*
    - *2 - The metric is followed only to a limited extent*
    - *3 - The metric is followed to a good extent*
    - *4 - The metric is followed mostly*
    - *5 - The metric is followed completely*

  Return the rating of your evaluation and explain the reasons.

Following internal team discussions, a score threshold of 3 was established as the passing criterion. Reports receiving a score above 3 are considered acceptable, while those scoring below 3 undergo manual review and refinement. The revised reports are then re-evaluated using the LLM-as-a-Judge method in an iterative process until they meet the required standard. This approach ensures a balance between automated evaluation efficiency and human oversight, enhancing the overall quality and reliability of the assessment.

Subject Matter Experts (SMEs) Review: After using the LLM to complete the initial assessment, the refined content is forwarded to SMEs for validation. The SMEs, primarily members of the CII RT-398 team, possess extensive industry experience, ensuring a rigorous evaluation process. Each SME was assigned to validate three Best Practices (BPs) by manually reviewing the original research reports and comparing them against the extracted insights generated by the proposed framework. The validation focused on four key aspects: (1) whether the extracted GNs accurately represent the most critical information from the report, (2) whether the summarized DAIs maintain accuracy and fidelity, (3) whether the GNs are assigned to the appropriate project lifecycle phases (4) whether all link, including research reports and tool links, remain valid. The authors systematically collected SME feedback and incorporated necessary refinements through manual revision, ensuring the final content meets the highest quality and reliability standards. The SME validation process involved multiple rounds of both virtual and in-person meetings. Additionally, students from a major public university in the U.S. who were enrolled in the CII BP course contributed to the validation of BP content through weekly homework assignments.

3.5 Stage 3: Deployment

Passive Approach: As an initial deliverable, the research team compiled all extracted information into an Excel spreadsheet, serving as a structured and static repository for seamless distribution among diverse stakeholders. The choice of Excel ensures broad accessibility, requiring minimal technical expertise while enabling efficient navigation and interaction with the dataset. This format enhances usability and information dissemination, allowing stakeholders to passively explore the validated content, search for relevant insights, and review findings with ease.

Active Approach: To enhance the usability and accessibility of the product, this research explored the local deployment of a chatbot, enabling active user engagement for rapid information retrieval. Unlike the static Excel-based repository, the chatbot provides an interactive interface, allowing users to query and extract relevant insights efficiently. Additionally, the local computer can function as a centralized host, providing a shareable access link that enables multiple users to interact with the chatbot remotely, ensuring seamless communication and broader usability across stakeholders.

The collected data is first stored locally and processed using a local embedding model (e.g., Ollama, which offers multiple model options) to convert textual information into high-dimensional vector representations. These embeddings are then stored in a vector database (e.g., ChromaDB) for efficient retrieval. For chatbot implementation, this study employs a LangChain pipeline as a prototype, enabling interaction with a locally hosted LLM (e.g., Llama 3, Mistral (Jiang et al. 2023)) and deploying the chatbot as a web-based interface (e.g., hosted on localhost). To facilitate remote accessibility, remote access tunneling software (e.g., ngrok) is utilized to establish a temporary, secure tunnel between the local machine and an externally accessible URL. This approach eliminates the need for complex network configurations while ensuring secure, public access to the chatbot from remote devices. As a result, remote users can seamlessly interact with the chatbot via a web-based application running on a local machine while maintaining data privacy and control. Furthermore, the proposed methodology is scalable and adaptable, allowing for future expansion if data traffic increases due to a growing number of users.

4. RESULTS AND ANALYSIS

LLM-as-a-Judge Results: Phi-3 assigns a score of 4 (indicating that the metric is mostly followed) to 46% of the 54 CII BP reports and a score of 5 (indicating complete adherence) to 42%. This suggests that Phi-3 found Llama 3 to be highly reliable in extracting key content. Only 12% of the reports received a score of 3 (0% for scores 1 and 2), where the authors refined their contents based on Phi-3's feedback. For instance, one comment for Target Safety Programs (RT-216) corresponding to a score of 3 is:

"The…rated as '3' because it follows the metric to a good extent, but there are areas where improvements can be made…The first golden nugget about…is accurately reflected...However, it lacks detail on how these methods contribute to achieving over 100 responses…The second golden nugget…but could be more comprehensive by including specifics of what kind of "in-depth" information was sought..."

Tool - BP Handbook (Excel version): As shown in Figure 4, multiple layers of CII (Construction Industry Institute) information have been systematically aggregated into an Excel format to enhance accessibility and usability.

- Horizontal Structure: The columns are organized according to project phases. Each phase contains its definition in the “Note” function. Contractors can identify and examine GNs associated with their respective project delivery methods.
- Vertical Structure: The rows are structured based on 17 Best Practices (BPs) (Layer 1), serving as the foundational layer of information. Each BP includes an executive summary, which is hyperlinked to a corresponding Word file, allowing seamless access and knowledge sharing among company members.
- Users can expand the content by clicking the left-side “+” button to access Layer 2, which contains detailed reports for each BP. Each report is also accompanied by a high-level summary available in the “Note” function. Each Report includes a hyperlink that directly connects users to the original source webpage for further reference.
- By continuing to expand Layer 3, users can access GNs, along with their corresponding project phases. At this stage, research tools related to the BP will also become available, with links to their original webpages. This layer of information represents the key takeaways for each CII research report.
- Further expansion into Layer 4 reveals the detailed action items associated with each GN, providing deeper insights and practical implementation steps.

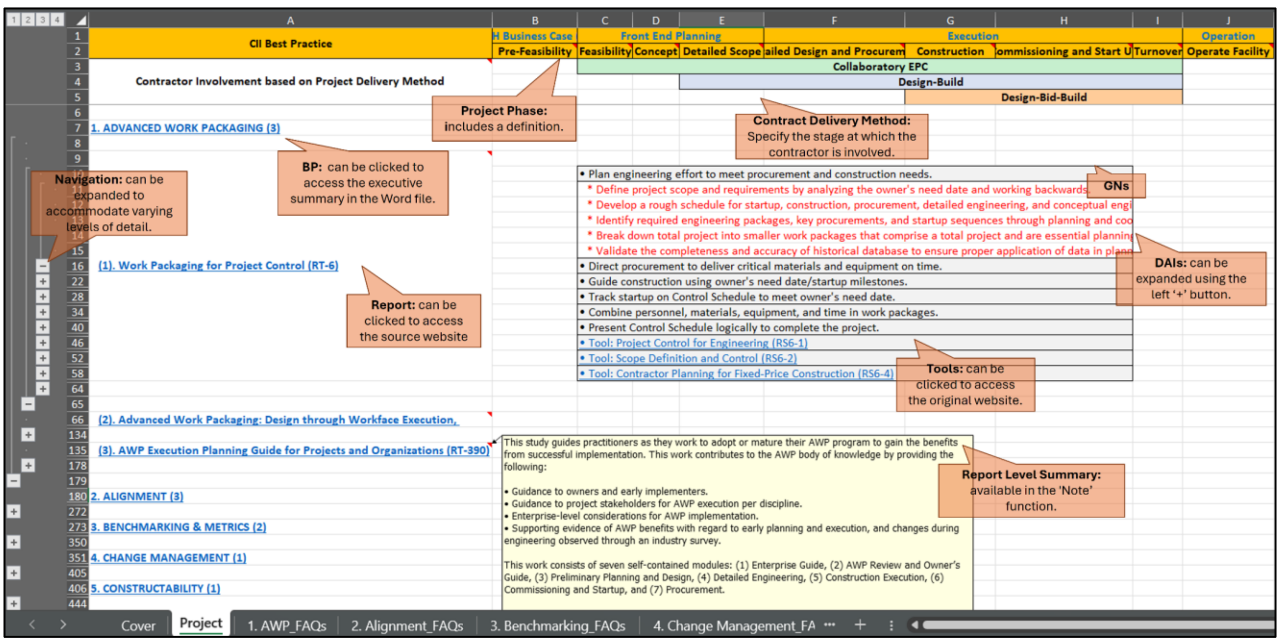

Figure 4. Screenshot of the BP handbook

This user-friendly interface allows users to navigate and explore different layers of information based on their specific needs. Additionally, they can always reference original sources through embedded hyperlinks, ensuring accuracy and comprehensiveness in their decision-making process.

SME Review Feedback: Each SME will compare what has been extracted in the BP handbook with the original report and provide detailed comments and feedback. The authors will integrate their feedback to refine the content extraction. For example, one SME comment for RT-307 is:

> “***Overall Research Summary:*** *While this information is good to have, it does not help the user with best practices on how to mitigate threats of counterfeit materials… p.32 has a great summary…includes the following: …**Bullets/Sub-bullets/Nuggets:** …Our summaries/golden nuggets should focus on p.19-26 where they explain the risk management process to mitigate the threats of counterfeit materials…”*

RAG-Enabled Local Chatbot: Figure 5 presents a screenshot of the implementation of a domain knowledge-supported local chatbot within a localhost environment. To evaluate its functionality, we tested an example query: “*What is the owner’s role in construction safety?*” The chatbot successfully retrieved relevant safety BPs from the CII and generated a structured response. This response included key action items in bullet points along with corresponding references, demonstrating the chatbot’s ability to facilitate domain-specific communication and support informed decision-making. Through multiple rounds of testing, the chatbot consistently produced meaningful outputs, ensuring that users could extract valuable insights for internal information sharing.

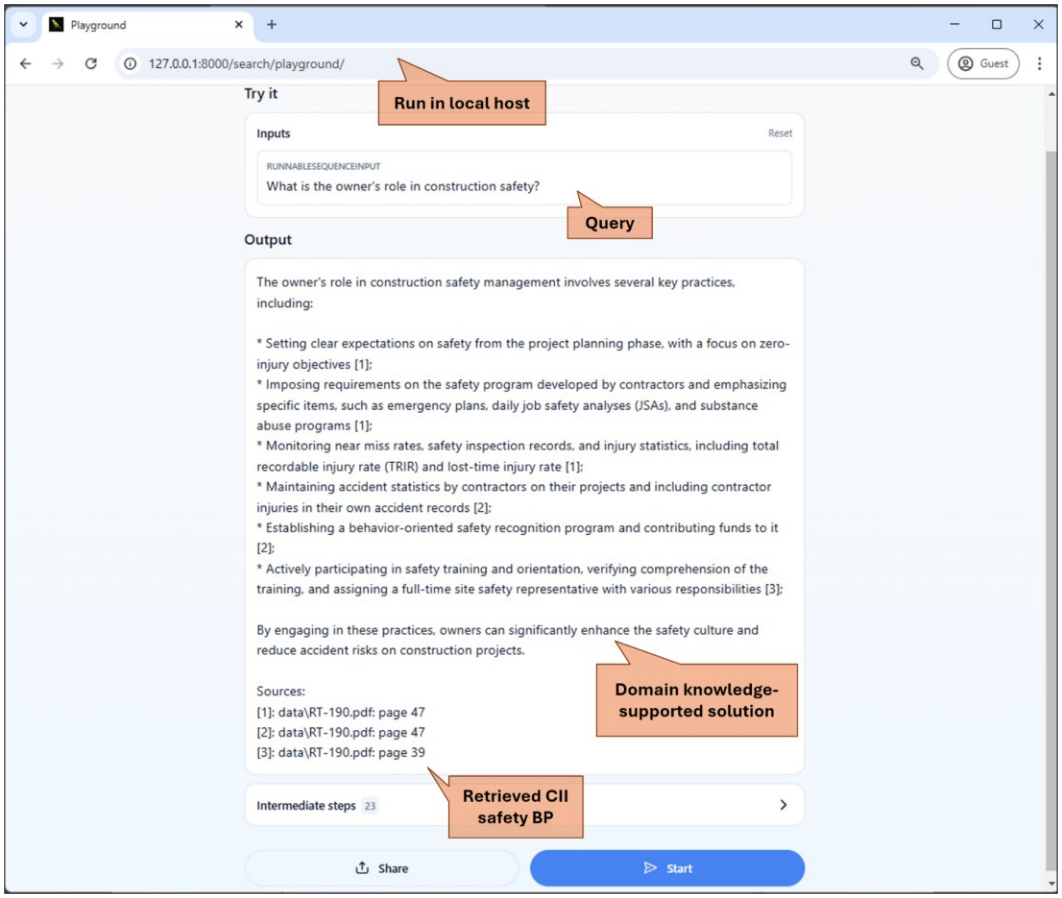

Figure 5. Screenshot of implementing a domain knowledge-supported local chatbot

In addition, this research will validate the finalized handbook by collaborating with end users to develop practical use cases for the tool. For instance, using it to design targeted education and training modules or to facilitate the implementation of BPs within their organizations; streamlining existing operational procedures by embedding the handbook’s frameworks into daily activities, ensuring that best practices are sustainably woven into organizational routines.

5. CONCLUSION

This study developed a CII BP handbook using a GenAI-enabled framework. The framework is built on RAG-enhanced GenAI for question-answering and prompt engineering, ensuring effective knowledge retrieval and structured responses. As a result, CII BP knowledge was systematically integrated into a user-friendly Excel interface, allowing users to efficiently explore key research insights. Additionally, this study led to the development of a domain knowledge-supported local chatbot, enabling users to seamlessly query and retrieve critical information from the CII BP knowledge base.

This study makes two key contributions. First, it enhances the accessibility and applicability of CII research by transforming raw “data” into structured domain “knowledge”. Traditionally, CII research exists as extensive datasets, requiring users to manually extract relevant “information” from research reports. This study addresses this limitation by systematically distilling critical knowledge from BP modules, ensuring that construction professionals can readily access actionable insights to support informed decision-making. This is achieved through multi-layered GenAI information filtering that integrates advanced textual understanding with expert validation from SMEs. Second, this research advances the field of human-AI interaction in construction project management. The proposed GenAI framework facilitates the development of an on-site construction AI assistant specialized in domain-specific knowledge retrieval and decision support. By introducing an intuitive, user-friendly interface, this study promotes the adoption of GenAI among construction professionals, particularly those with limited expertise in AI technology.

ACKNOWLEDGMENTS

The authors would like to gratefully acknowledge the funding support provided by the CII for the study. The authors would also like to thank the members of the CII research team RT-398 for their support and guidance throughout this study.

REFERENCES

Chen, G., Alsharef, A., Ovid, A., Albert, A., and Jaselskis, E. 2025. Meet2Mitigate: An LLM-powered framework for real-time issue identification and mitigation from construction meeting discourse. Advanced Engineering Informatics, **64**: 103068. doi:10.1016/j.aei.2024.103068.

CII. 2009. The Implementation Planning Model: Steps to Success, Version 1.2.

CII. 2012. Support for the Implementation Champion: The Experience Reference Index.

CII. 2022a. CII Best Practices Guide: Improving Project Performance, Version 5.0.

CII. 2022b. Improving Project Performance CII BEST PRACTICES GUIDE.

CII. 2025. CII Best Practices. Available from https://www.construction-institute.org/cii-best-practices. [accessed 4 February 2025].

Deshpande, A.S., Salem, O.M., and Miller, R.A. 2012. Analysis of the Higher-Order Partial Correlation between CII Best Practices and Performance of the Design Phase in Fast-Track Industrial Projects. Journal of Construction Engineering and Management, **138**(6): 716–724. doi:10.1061/(ASCE)CO.1943-7862.0000487.

HuggingFace. 2025. Artificial Analysis LLM Performance Leaderboard. Available from https://huggingface.co/spaces/ArtificialAnalysis/LLM-Performance-Leaderboard. [accessed 6 February 2025].

Jiang, A.Q., Sablayrolles, A., Mensch, A., Bamford, C., Chaplot, D.S., Casas, D. de las, Bressand, F., Lengyel, G., Lample, G., Saulnier, L., Lavaud, L.R., Lachaux, M.-A., Stock, P., Scao, T. Le, Lavril, T., Wang, T., Lacroix, T., and Sayed, W. El. 2023. Mistral 7B. arXiv,: 1–9.

Kim, S.B. 2014. Assessment of CII Best Practices usage in the construction industry. KSCE Journal of Civil Engineering, **18**(5): 1228–1238. doi:10.1007/s12205-014-0221-z.

Li, Y., and Starly, B. 2024. Building a knowledge graph to enrich ChatGPT responses in manufacturing service discovery. Journal of Industrial Information Integration, **40**: 100612. doi:10.1016/j.jii.2024.100612.

Malhotra, P.M. 2017. Promoting the use of Construction Industry Institute (CII) Best Practices in the Upstream, Midstream, and Mining (UMM) sector. North Carolina State University.

Meta. 2024. Introducing Llama 3.1: Our most capable models to date. Available from https://ai.meta.com/blog/meta-llama-3-1/. [accessed 25 October 2024].

Olumide, A., Walewski, J., and Anderson, S. 2012. Assessing CII Best Practices Applicability. *In* Construction Research Congress 2012. American Society of Civil Engineers, Reston, VA. pp. 2439–2448.

Ovadia, O., Brief, M., Mishaeli, M., and Elisha, O. 2023. Fine-Tuning or Retrieval? Comparing Knowledge Injection in LLMs. arXiv,. doi:https://doi.org/10.48550/arXiv.2312.05934.

Zheng, L., Chiang, W.-L., Sheng, Y., Zhuang, S., Wu, Z., Zhuang, Y., Lin, Z., Li, Z., Li, D., Xing, E.P., Zhang, H., Gonzalez, J.E., and Stoica, I. 2023. Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena. *In* 37th Conference on Neural Information Processing Systems (NeurIPS 2023).