Large Language Model-based data-driven framework for digital transformation in construction industry

Ruiyang Ma,[1,2] Geoffrey Qiping Shen,[3] Peiliang Lou,[4] and Ming Wang,[5]

[1] Department of Building and Real Estate, Hong Kong, Polytechnic University, Kowloon 999077, Hong Kong; e-mail: ruiyang0508@126.com
[2] School of Management Science and Real Estate, Chongqing University, No. 174, Shazheng St., Shapingba District, Chongqing 400044, PR China
[3] Department of Building and Real Estate, Hong Kong, Polytechnic University, Kowloon 999077, Hong Kong; e-mail: geoffrey.shen@polyu.edu.hk
[4] School of Computer Science and Technology, Xi'an Jiaotong University, No. 28, Xianning West St., Beilin District, Xi'an, Shaanxi 710049, PR China; e-mail: loupeiliang312@stu.xjtu.edu.cn
[5] Department of Building and Real Estate, Hong Kong, Polytechnic University, Kowloon 999077, Hong Kong; e-mail: cd.melvin.wang@gmail.com

ABSTRACT

With the wide and fragmented use of digital technology in construction, a systematic digital transformation (DT) of the industry is needed. The industry's 'synergy development' context, marked by diverse data resources and significant investment, complicates collaboration and burdens the DT process. Notably, the transformation knowledge of DT is often 'buried' within the vast data produced by daily management processes, making it challenging to discern the rules of DT without labor-intensive and time-consuming manual methods. Hence, a well-established data-driven framework for enhancing the DT process to promote whole-life-cycle industry transformation is essential. The large language model (LLM) supercharges the data-driven framework, enabling automated reasoning and precise insights to be derived from extensive datasets, thus fostering a smarter DT framework to manage the DT process. Therefore, this study uses a question-answering system based on an LLM and a localized knowledge base to guide decision-makers in developing engagement strategies that improve DT performance and foster collaboration. This study presents a practical application of LLMs in the DT of construction enterprises, anticipates future applications, and explores their potential use throughout a construction project's transformation lifecycle.

1. INTRODUCTION

As the global economy embarks on a period of digital transformation (DT), digital technology has emerged as a crucial catalyst for economic growth worldwide and serves as a pivotal engine for the advancement of high-quality economic development. Concurrently, industries across the spectrum are engaging in the process of DT (Ezeokoli et al., 2016; Kutnjak et al., 2019). Contrary to other sectors, construction, a cornerstone industry of the nation, is characterized by fragmentation, non-replicability, transience, and decentralization, which makes DT particularly challenging(Koeleman et al., 2019). Digital transformation (DT) has evolved as an essential phenomenon in strategic researches (Vial, 2021). It is defined as the process of an organisation or industry's adoption and implementation of digital technologies, which contributes to further fast provision of strategies meeting the managerial needs for different fields, especially in construction industries (Samuelson & Stehn, 2023; Warner & Wäger, 2019). However, the construction industry has been slow to adopt new technologies and has yet to undergo a significant transformation (Adekunle et al., 2021). While numerous enterprises have initiated their DT process, some still fail to comprehend its complexity, overlooking the importance of extracting and synthesizing daily construction and management data, experience, and knowledge, leading to a "black box" scenario. Furthermore, with many enterprises involved in the construction industry's upstream and downstream chains, promoting digitization across the construction lifecycle necessitates enhanced data flow among multiple participants. However, the knowledge of DT in the construction industry is difficult to see directly, encapsulated, or 'buried' within the vast amount of data generated during routine daily management processes, which requires labor-intensive and time-consuming manual methods. A comprehensive, systematic analysis of these data possesses significant potential for the effective interpretation of DT knowledge, which enhances. Further, sophisticated AI models, particularly the Large Language model (LLM), supercharge the data-driven framework, enabling automated reasoning and precise insights to be derived from extensive datasets.

LLM excels in various natural language processing tasks, showcasing its potential to offer a unified solution for accurately addressing complex problems, such as facilitating problem identification and solutions and extracting key information from big data (Thirunavukarasu et al., 2023). However, in the construction DT scenario, two primary limitations persist. **Firstly**, a comprehensive data-driven framework for DT in the construction industry is still lacking (F. Li et al., 2023). A data-driven framework would significantly contribute to our understanding of how to enhance DT, shape enterprise governance and engineer management in the construction industry and improve the industry's overall performance. **Secondly**, the potential of advanced generative AI models, specifically LLM, for application in construction vertical scenarios has been overlooked. In the context of the construction domain, the capabilities of LLM remain limited, which may suffer from deficiencies in veracity and precision, potentially yielding what can be described as ' hallucinatory fact' (Maynez et al., 2020). Therefore, a well-established, large-language model-based framework for enhancing DT in the construction industry is essential. This study employs a question-answering system grounded in an LLM, a localised knowledge base, and a knowledge graph as an exemplar to address the issue of catastrophic forgetting of data within the construction professional realm. It constructs a practical application of LLMs in the DT of construction enterprises, envisages future application scenarios, and explores the potential use of such models throughout the entire life cycle of a construction project's transformation. This research thus offers innovative concepts for the intelligent enhancement of the construction industry.

2. RESEARCH BACKGROUND

2.1 *Data-driven approaches in construction management*

The use of data-driven approaches in construction enterprise management has been a topic of increasing interest in recent years. Several studies have explored the area, highlighting the potential benefits and challenges associated with integrating these techniques into construction enterprise management practices . The examination and manipulation of extensive datasets facilitate the detection of problems, identification of potential hazards, and recognition of opportunities for improvement (Bilal et al., 2016; F. Li et al., 2023; Lu et al., 2015). Employing data analytics for the refinement of the construction supply chain enables enterprise to reduce waste, streamline logistics, diminish operational expenditures, and advance sustainability (Atuahene et al., 2020; F. Li et al., 2023; Yu et al., 2020). The data-driven framework has also been developed to recognize key information from big data, facilitate communication, improve project outcomes, reduce costs, and increase efficiency (F. Li et al., 2023; Walvoord et al., 2022).

However, the DT revolution in construction companies is still in its early stages. The construction industry, especially in the realm of intelligent construction, generates vast amounts of data daily. These data encompass a wide range of information, from building models, design, cost predictions, and management, which are generally unstructured, heterogeneous, and therefore extremely complex to deal with. In addition, the data is generated by a wide range of isolated and networked machines and devices, cloud-based solutions, business planning and management (F. Li et al., 2023). Consequently, it is essential to investigate and develop comprehensive data-driven approaches specific to the construction industry (Bilal et al., 2016; Ismail et al., 2018). For example, data-driven algorithms can promote construction project quality (Wang et al., 2018); the capabilities of construction enterprises can be evaluated through the use of predictive and prescriptive analytics within a big data platform (Ngo et al., 2020); and the vast data gathered on the progress of construction projects can be analyzed by the AI models, yielding insightful and comprehensive perspectives on construction practices (Bilal et al., 2016). Leveraging data from construction projects can optimize everything from design to execution, propelling the next evolution of industrial innovation. This enables the development of improved efficient techniques to address potential issues, facilitates decision-making, and facilitates strategy development for construction management (F. Li et al., 2023).

2.2 *LLM in construction management*

The advent of LLMs has precipitated a transformative shift in knowledge dissemination and big data analytics. Models such as OpenAI's GPT-3/GPT-4, Meta's LLaMA, Google's LaMDA, and Stanford's Alpaca have demonstrated exceptional aptitude in processing and producing text with a high degree of linguistic fidelity to human expression (Yin et al., 2023). Trained on extensive datasets, LLMs exhibit a broad spectrum of linguistic capabilities, including natural language understanding, generation, translation, and sentiment analysis. These models employ neural networks to process and analyze textual data (Xu et al., 2023). By absorbing statistical patterns and semantic representations from abundant text and data, these models acquire the capacity to grasp the complexities of language and generate coherent, contextually relevant responses. This positions LLMs as an optimal AI approach for managing extensive volumes of text from varied sources, such as academic publications, industry reports, and regulatory documents. LLMs have revolutionized knowledge sharing and collaboration by their adeptness in understanding and generating human-like text.

Large language models have found applications in finance, medicine, and other professional domains (Y. Li et al., 2023; Singhal et al., 2023; Thirunavukarasu et al., 2023; C. Wu et al., 2023; S. Wu et al., 2023); however, their utilization in the construction industry remains limited. Prieto et al. proposed a ChatGPT-based construction plan generation approach that can generate construction schedules for construction projects (Prieto et al., 2023). Uddin et al. used ChatGPT to assist construction site hazard identification and conducted experiments to verify that ChatGPT-assisted construction site hazard identification had a certain improvement effect (Uddin et al., 2023). However, the difficulty of using LLMs to build a framework for DT in the construction field lies in how to enhance the problem-solving ability of artificial intelligence in the professional field. Training data of LLMs is not verified for domain-specific accuracy, which leads to an issue of 'garbage in, garbage out' (Thirunavukarasu et al., 2023). Consequently, this study advances a paradigm for intelligent information framework by integrating 'LLMs + construction knowledge databases', referencing the question-answering system format. It employs domain-specific knowledge to refine the output generated by LLMs.

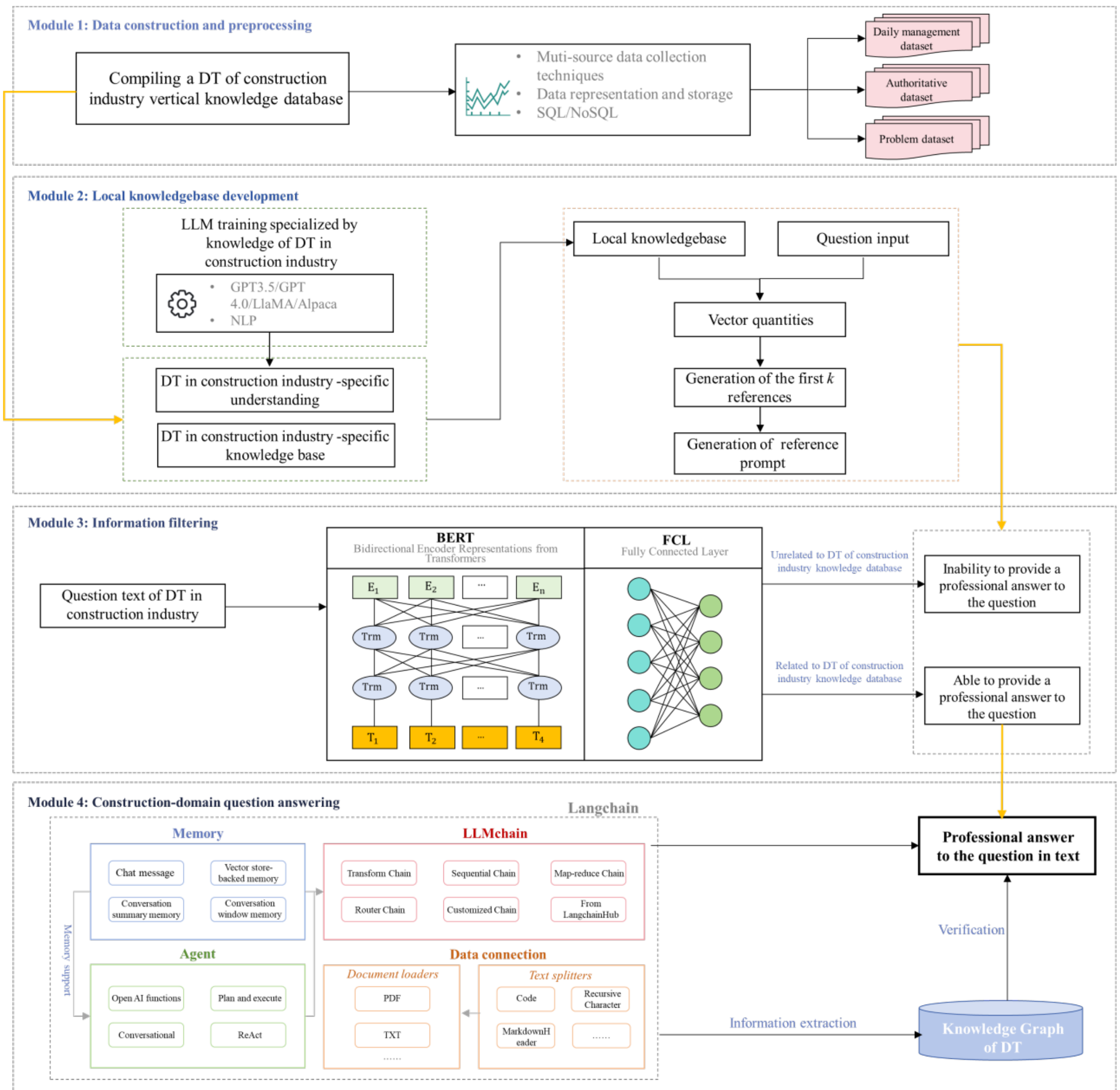

Figure 1. LLM-based data-driven framework for DT in construction industry

3. FRAMEWORK CONSTRUCTION

This paper introduces a question-and-answer framework to facilitate the DT of the construction industry. The objective of the framework is to investigate the integration of an LLM with a construction-specialized knowledge base, delving into the synergistic fusion of the language model and knowledge graph to enhance the domain-specific effectiveness of Q&A systems in the DT of the construction industry. This paper proposes a framework that aims to accomplish the following functions (as shown in Figure 1): **data construction and preprocessing, local knowledgebase development, information filtering, and construction-domain question answering** (Zhang et al., 2023). To realize these modules, the system capitalizes on construction professional knowledge and LLM (GPT Turbo 4.0), employing LangChain, a language model integration framework, to synergize both (Topsakal & Akinci, 2023). It introduces a framework that merges the LLM with a knowledge graph (Sun et al., 2023). The local knowledge base modules are engineered to integrate the LLM with a localized knowledge repository, adapting prompts to align with task-specific demands, thereby furnishing users with more precise and task-relevant prompt cues. The information filtering module is designed to mitigate the risk of misinformation generated by LLMs, thereby enhancing answer accuracy. The domain-specific Q&A module delivers expert responses by integrating a professional knowledge base focused on DT in the construction industry with an LLM.

As the flow chart in Figure 2 shows, firstly, the text pertaining to DT in the construction industry is processed for classification to ascertain its relevance to the field. Secondly, a local knowledge base is constructed, within which text-related knowledge is sought via LangChain. The knowledge base is an output of LLM which includes specific knowledge base and understanding within the construction industry. This information, along with the pertinent question, is inputted into an LLM as a prompt. The model then generates a response, utilising professional knowledge through logical reasoning. Following this, the answers undergo knowledge extraction, and the extracted triples are matched with the knowledge bases in the knowledge graph to verify the professional accuracy of the responses.

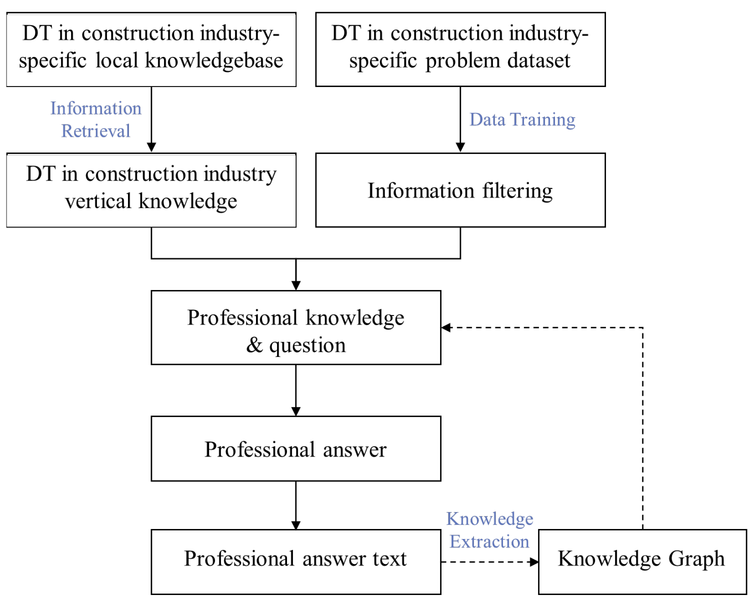

Figure 2. Flow chart of Q&A system

4. SPECIFIC MODULE CONSTRUCTION

4.1 *Data construction and preprocessing*

A vast amount of data related to DT exists, and so the module will be to compile a comprehensive, data-driven database specifically focused on DT in the construction industry. The implementation of this framework necessitates the aggregation and organisation of construction-specific datasets to underpin its execution. The database will serve as a valuable resource for obtaining a DT-specific LLM to enable the data-driven framework. Data will be sourced from the following functional domains, including structured and unstructured data. The dataset includes:
(1) Daily management dataset: The sources can be summarized as construction project monitoring dataset, experimental dataset, simulation dataset, organization management dataset, company information dataset, project field survey, and unstructured sources such as social networks and crawler data from websites.

(2) Authoritative dataset: This dataset comprises authoritative normative data sourced from official channels within the construction industry, encompassing industry norms, construction standards, safety regulations, etc. These datasets provide the large model with the necessary professional and regulatory guidance for generating responses.

(3) Problem dataset: This dataset is utilized to train the information filtering model, aiming to mitigate the inaccuracy of responses due to data omission. Initially, a data piece is selected from the relevant dataset to generate a prompt. This prompt is then input into the LLM to produce data, and this process is iterated until all pertinent data is selected (Zhang et al., 2023).

Algorithm 1 generates data based on LLM
Input: Associated text A and the LLM's API Key
Output: Problem data B, generated based on the LLM

1. llm_connection←creat(API_KEY) # Connections to the LLM interface are facilitated through the use of the API_KEY
2. for *i* : =1 to N do
3. *a* ←select(A) # Select a piece of relevant data
4. prompt ← P(*a*) # Generate a prompt based on the selected data
5. *b* ←llm result (*llm_connection, prompt*) # Utilize the established connection to interface with the LLM, input a prompt, and subsequently generate a response
6. B ←abstract(*b*) #Problem data is extracted from LLM generated results and summarized into B
7. End for

In this context, 'A' denotes the complete set of related data, while 'a' signifies a single piece of related data. 'B' represents the entire set of problem data generated, and 'b' stands for a single piece of generated data. The function 'create' establishes a connection to the LLM, for example, ChatGPT, using the user-provided API_KEY. The function 'select' chooses a specific piece of data. 'P' generates a relevant prompt based on the selected data, and 'llmresult' retrieves the response produced by the LLM. The term 'abstract' implies the extraction of problem data from the generated responses and its subsequent summarization.

4.2 *Local knowledgebase development*

To optimize indexing and retrieval, sentence transformers and the FAISS database are pre-emptively employed to construct an index, enhancing efficiency. The sentence-transformers library generates vectors for user queries and the related local knowledge base entries. To avoid computational waste in pairwise similarity comparisons, FAISS—a database by Facebook AI Research that leverages clustering and inverted indexes—facilitates rapid similarity searches. Data vectors are clustered into groups of similar vectors to expedite similarity calculations, with inter-group comparisons preceding intra-group calculations. This approach swiftly pinpoints local knowledge base segments closely matching user queries. Incorporating the top-*k* most similar segments into the prompt further refines the accuracy of the generated responses.

4.3 *Information filtering*

For specialized queries, the LLM's responses need be constrained to the relevant field by incorporating a BERT (bidirectional coder representations from transformers) -based text filter into the system (Devlin et al., 2019). This allows for effective question filtering, narrowing down the scope of the model's responses. Consider the set of all possible inputs to the LLM as C, the subset of questions within a specific professional field it can answer as D, and the subset of questions for which it can generate expert responses as E, with C > D > E. Utilizing a fine-tuning approach could constrain D to E, potentially reducing the model's response capabilities. However, implementing a filter allows C to be narrowed to D, ensuring the inquiries fall within D's scope as much as possible. This information filtering safeguards the system's capacity to address the maximum number of queries within its competency, thereby minimizing the risk of fact distortion. The training data is fed into BERT, followed by the output from BERT being passed through a fully connected layer (FCL) to yield the classification result.

4.4 *Construction-domain question answering*

To enhance the professionalism of responses from the LLM's knowledge question-answering system, this study incorporates expert knowledge into the knowledgebase via prompts in the “local knowledgebase development” module. In this module, LangChain is applied to load files in the knowledge database, embed and vectorise the file contents, construct prompts for questions, send them to LLMs and get responses. Linked by LangChain, the framework produces more specialised answers. The module identifies question-related expert knowledge within the knowledgebase, which, combined with the question text, forms the fused knowledge base (as shown in Figure 3). This is then input into LLM to generate the final response.

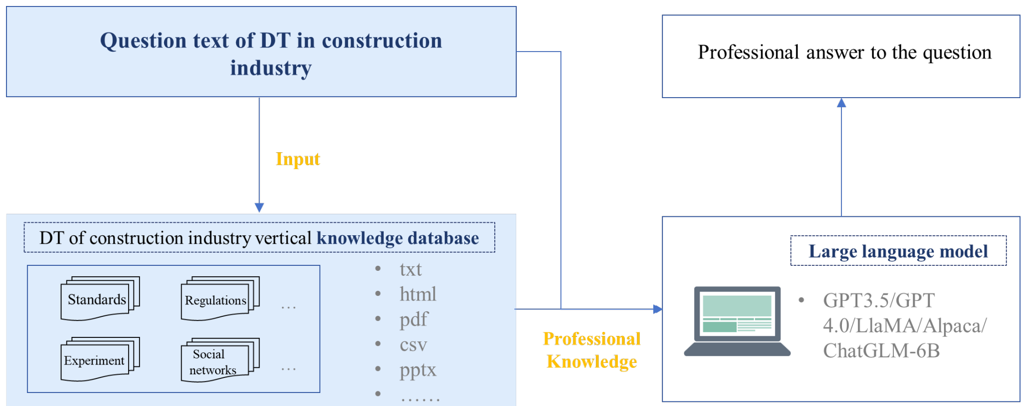

Figure 3. Flow chart of the LangChain-driven Q&A module

5. DISCUSSIONS

In the modern implementation of DT, in addition to summarizing professional knowledge, it is crucial to understand problems more effectively and present professional knowledge in a comprehensive way. The novel approach this framework applies is the integration of LLMs in constructing the question-answering system.

LLMs, existing as advanced computation models, have demonstrated good performances in Natural Language Processing (NLP) tasks with less effort. According to an existing study which compared several sentiment analysis tools and prompt-based ChatGPT, which shows ChatGPT's leading performances with no tuning and training (Lossio-Ventura et al., 2024). Other studies have proven that LLMs, along with prompt engineering, can be considered an essential emerging skill that helps leverage the full potential of AI in summarizing and answering professional knowledge via text generation (Meskó, 2023; Venerito et al., 2024).

Regarding data-driven frameworks which assist the construction industry, previously proposed methods have been focused on data mining with methods such as K-means clustering, support vector machines (SVM) and regression analysis (Majeed et al., 2021; Yang et al., 2020; Zhao et al., 2022). The output of these frameworks is only statistics and the summary of big data that do not enable the generation of texts with automated reasoning and precise insights from text enquiries. Applying LLMs in a data-driven framework can better leverage the professional knowledge of the construction industry's Local Knowledgebase to build a system that combines professionalism with ease of operation and flexible methods for querying professional information, thereby improving the question-answering strategies lacking in the field of DT.

6. APPLICATION SCENARIO PROMOTION

Big data analytics significantly advances the construction industry by enhancing building design, cost management, energy use forecasting, material performance, and decision-making based on LangChain+LLMsystems (Yan et al., 2020). The extensive data from a construction project's lifecycle—covering weather conditions, workforce activities, materials, vehicular movements, energy usage, etc.—can refine AI models, offering deeper, actionable insights for the construction industry(Bilal et al., 2016). Such big data analytics facilitate improved scheduling, planning, and the formulation of effective construction methodologies. Moreover, streamlined data flow aids in the management of construction projects by supporting strategic decision-making. The prior article preliminarily explored the potential for applying LLM of DT within the construction industry. However, further deliberation is required to identify specific application scenarios. This study outlines the implementation strategies for this method in select scenarios, serving as a reference for future LLM applications in construction management.

(1) **Specification Retrieval:** Engineers frequently consult specifications, a process that can be streamlined by LLM. These models enable direct access to specific information and relevant standards through vast conversational queries, simplifying the search for and verification of specifications in a more effective way.
(2) **Engineering Document Generation:** During the engineering design phase, a substantial amount of textual content is required. Engineers typically adjust and modify existing templates to meet these needs. However, an LLM can be employed to undertake this task, facilitating the generation of documents such as tenders and design descriptions, thereby replacing the need for manual intervention by engineers.
(3) **Knowledge Sharing Platform:** Construction management is a decentralized process with ad hoc organizational structures and non-linear workflows. Given the diverse projects and information levels of subcontractors, obtaining accurate data can be challenging for general contractors and project decision-makers. Consequently, an information and data sharing platform has been established to offer reliable recommendations for various collaboration issues.
(4) **Construction Engineering Education:** The deployment of LLMs can facilitate a comprehensive question-answering system in an educational scenario. By incorporating prevalent industry textbooks and establishing a localised knowledge base, the model's capacity to address professional queries can be significantly enhanced, thereby advancing construction engineering education.

7. CONCLUSION

In the LLMs era, harnessing the 'emergent' abilities of LLMs for construction management applications is crucial for future competitive advantage. This research proposes a data-driven framework paradigm, amalgamating 'Large Language Models (LLMs) + construction knowledge databases', modelled on a question-answering system. It leverages domain-specific knowledge to enhance the output produced by LLMs. This framework focuses solely on the application of DT in the construction industry, but its design principles can be extended to other sectors such as law, finance, and education. However, a standard benchmark to evaluate the system's expertise is currently lacking.

Therefore, future work could involve developing a system specifically to assess the accuracy and professionalism of these verticals.

REFERENCES

Adekunle, S. A., Aigbavboa, C. O., Ejohwomu, O., Adekunle, E. A., & Thwala, W. D. (2021). Digital transformation in the construction industry: A bibliometric review. *Journal of Engineering, Design and Technology*.

Atuahene, B. T., Kanjanabootra, S., & Gajendran, T. (2020). Benefits of big data application experienced in the construction industry: A case of an Australian construction company. *Proceedings of the 36th Annual Association of Researchers in Construction Management (ARCOM) Conference, Virtual Conference, Leeds, UK*, 7–8

Bilal, M., Oyedele, L. O., Qadir, J., Munir, K., Ajayi, S. O., Akinade, O. O., Owolabi, H. A., Alaka, H. A., & Pasha, M. (2016). Big Data in the construction industry: A review of present status, opportunities, and future trends. *Advanced Engineering Informatics*, *30*(3), 500–521.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In J. Burstein, C. Doran, & T. Solorio (Eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (pp. 4171–4186). Association for Computational Linguistics.

Ezeokoli, F. O., Okolie, K. C., Okoye, P. U., & Belonwu, C. C. (2016). Digital transformation in the Nigeria construction industry: The professionals' view. *World Journal of Computer Application and Technology*, *4*(3), 23–30.

Ismail, S. A., Bandi, S., & Maaz, Z. N. (2018). An appraisal into the potential application of big data in the construction industry. *International Journal of Built Environment and Sustainability*, *5*(2).

Koeleman, J., Ribeirinho, M. J., Rockhill, D., Sjödin, E., & Strube, G. (2019). Decoding digital transformation in construction. *Capital Projects & Infrastructure Practice*.

Kutnjak, A., Pihiri, I., & Furjan, M. T. (2019). Digital Transformation Case Studies Across Industries – Literature Review. *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 1293–1298.

Li, F., Laili, Y., Chen, X., Lou, Y., Wang, C., Yang, H., Gao, X., & Han, H. (2023). Towards big data driven construction industry. *Journal of Industrial Information Integration*, 100483.

Li, Y., Wang, S., Ding, H., & Chen, H. (2023). Large Language Models in Finance: A Survey. *4th ACM International Conference on AI in Finance*, 374–382.

Lossio-Ventura, J. A., Weger, R., Lee, A. Y., Guinee, E. P., Chung, J., Atlas, L., Linos, E., & Pereira, F. (2024). A Comparison of ChatGPT and Fine-Tuned Open Pre-Trained Transformers (OPT) Against Widely Used Sentiment Analysis Tools: Sentiment Analysis of COVID-19 Survey Data. *JMIR Mental Health*, *11*(1), e50150.

Lu, W., Chen, X., Peng, Y., & Shen, L. (2015). Benchmarking construction waste management performance using big data. *Resources, Conservation and Recycling*, *105*, 49–58.

Majeed, A., Zhang, Y., Ren, S., Lv, J., Peng, T., Waqar, S., & Yin, E. (2021). A big data-driven framework for sustainable and smart additive manufacturing. *Robotics and Computer-Integrated Manufacturing*, *67*, 102026.

Maynez, J., Narayan, S., Bohnet, B., & McDonald, R. (2020). *On Faithfulness and Factuality in Abstractive Summarization* (arXiv:2005.00661). arXiv.

Meskó, B. (2023). Prompt Engineering as an Important Emerging Skill for Medical Professionals: Tutorial. *Journal of Medical Internet Research*, *25*(1), e50638.

Ngo, J., Hwang, B.-G., & Zhang, C. (2020). Factor-based big data and predictive analytics capability assessment tool for the construction industry. *Automation in Construction*, *110*, 103042.

Prieto, S. A., Mengiste, E. T., & García de Soto, B. (2023). Investigating the use of ChatGPT for the scheduling of construction projects. *Buildings*, *13*(4), 857.

Samuelson, O., & Stehn, L. (2023). Digital transformation in construction – a review. *Journal of Information Technology in Construction*, *28*, 385–404.

Singhal, K., Azizi, S., Tu, T., Mahdavi, S. S., Wei, J., Chung, H. W., Scales, N., Tanwani, A., Cole-Lewis, H., & Pfohl, S. (2023). Large language models encode clinical knowledge. *Nature*, *620*(7972), 172–180.

Sun, J., Xu, C., Tang, L., Wang, S., Lin, C., Gong, Y., Ni, L. M., Shum, H.-Y., & Guo, J. (2023). *Think-on-Graph: Deep and Responsible Reasoning of Large Language Model on Knowledge Graph* (arXiv:2307.07697). arXiv.

Thirunavukarasu, A. J., Ting, D. S. J., Elangovan, K., Gutierrez, L., Tan, T. F., & Ting, D. S. W. (2023). Large language models in medicine. *Nature Medicine*, 1–11.

Topsakal, O., & Akinci, T. C. (2023). Creating large language model applications utilizing langchain: A primer on developing llm apps fast. *Proceedings of the International Conference on Applied Engineering and Natural Sciences, Konya, Turkey*, 10–12.

Uddin, S. J., Albert, A., Ovid, A., & Alsharef, A. (2023). Leveraging ChatGPT to Aid Construction Hazard Recognition and Support Safety Education and Training. *Sustainability*, *15*(9), 7121.

Venerito, V., Lalwani, D., Del Vescovo, S., Iannone, F., & Gupta, L. (2024). Prompt engineering: The next big skill in rheumatology research. *International Journal of Rheumatic Diseases*, *27*(5), e15157.

Vial, G. (2021). Understanding digital transformation: A review and a research agenda. In *Managing Digital Transformation*. Routledge.

Walvoord, E. C., Howenstine, M. S., Allen, B. L., Ribera, A. K., Nabhan, Z. M., Tori, A. J., Eichholtz, R. D., & Dankoski, M. E. (2022). Engaging All Stakeholders to Create a Trusted, Data-Driven, Process Improvement Approach to Addressing Learner Mistreatment. *Teaching and Learning in Medicine*, 1–11.

Wang, D., Fan, J., Fu, H., & Zhang, B. (2018). Research on optimization of big data construction engineering quality management based on RNN-LSTM. *Complexity*, *2018*.

Warner, K. S. R., & Wäger, M. (2019). Building dynamic capabilities for digital transformation: An ongoing process of strategic renewal. *Long Range Planning*, *52*(3), 326–349.

Wu, C., Lin, W., Zhang, X., Zhang, Y., Wang, Y., & Xie, W. (2023). *PMC-LLaMA: Towards Building Open-source Language Models for Medicine* (arXiv:2304.14454). arXiv.

Wu, S., Irsoy, O., Lu, S., Dabravolski, V., Dredze, M., Gehrmann, S., Kambadur, P., Rosenberg, D., & Mann, G. (2023). *BloombergGPT: A Large Language Model for Finance* (arXiv:2303.17564). arXiv.

Xu, R., Wang, X., Wang, T., Chen, Y., Pang, J., & Lin, D. (2023). *PointLLM: Empowering Large Language Models to Understand Point Clouds* (arXiv:2308.16911). arXiv.

Yan, H., Yang, N., Peng, Y., & Ren, Y. (2020). Data mining in the construction industry: Present status, opportunities, and future trends. *Automation in Construction*, *119*, 103331.

Yang, Y., Lu, X., & Wu, L. (2020). Integrated data-driven framework for fast SCUC calculation. *IET Generation, Transmission & Distribution*, *14*(24), 5728–5738.

Yin, S., Fu, C., Zhao, S., Li, K., Sun, X., Xu, T., & Chen, E. (2023). A Survey on Multimodal Large Language Models. *arXiv Preprint arXiv:2306.13549*.

Yu, T., Liang, X., & Wang, Y. (2020). Factors Affecting the Utilization of Big Data in Construction Projects. *Journal of Construction Engineering and Management*, *146*(5), 04020032.

Zhao, C., Dinar, M., & Melkote, S. N. (2022). A data-driven framework for learning the capability of manufacturing process sequences. *Journal of Manufacturing Systems*, *64*, 68–80.

Zhang, H., Wang, X., Han, L., Li, Z., & Chen, Z. (2023). Research on Question Answering System on Joint of Knowledge Graph and Large Language Models (in Chinese). *Journal of Frontiers of Computer Science and Technology*, 1.