2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

POTENTIAL OF LARGE LANGUAGE MODELS FOR CONSTRUCTION QUOTATION WORKFLOWS

Anja Brelih, Žiga Krvina, and Robert Klinc
University of Ljubljana, Faculty of Civil and Geodetic Engineering

Abstract

Large language models (LLMs) can automate text-intensive tasks in construction. However, preparing investigation quotations remains manual and time-consuming, with outputs varying between engineers due to non-deterministic steps like item selection and text drafting. This study proposes an LLM-assisted human-in-the-loop (HITL) workflow and evaluates several models (GPT 4.1, GPT 5, Claude Sonnet 4.5, and a fine-tuned open-source model, QwQ 32B) on 596 inquiry-quotation pairs. GPT 5 achieved the highest item-selection F1 score, while Claude 4.5 delivered comparable accuracy with the fastest processing time. This approach improves quotation efficiency and consistency, demonstrating practical benefits for construction workflows.

Introduction

Artificial intelligence (AI) is increasingly adopted in project management to support decision-making, automate repetitive tasks, and improve overall project performance (Taboada et al., 2023; Brelih et al., 2025). Recent advances in large language models (LLMs) have enabled systems that can understand and generate technical text, creating new opportunities for automating construction workflows.

Construction projects are characterised by unique products, processes, and project-specific contexts, which limit the applicability of rigid, fully predefined automation approaches (Turk, 2006). Quotation preparation is a critical but time-consuming activity for engineering companies. Preparing a quotation typically involves analysing client inquiries, selecting appropriate investigation or service items, estimating quantities, and composing a structured technical and commercial response. While parts of this workflow can be automated using conventional software, key steps remain non-deterministic and rely heavily on expert judgement. These include (1) selecting relevant investigation or service items and (2) formulating a coherent quotation text adapted to the specific inquiry.

This paper examines whether LLMs can effectively support and partially automate quotation preparation. Generic LLMs do not have access to internal company knowledge such as service catalogues, pricing structures, and domain-specific standards. To address this limitation, an open-source LLM was fine-tuned with domain-specific data and evaluated against selected commercial LLM solutions. For the use case, data from a company operating in the field of geotechnical investigations was used.

The contributions of this paper are threefold: (i) the design of an LLM-supported quotation workflow for (geotechnical) engineering practice, (ii) quantitative evaluation of multiple LLMs using accuracy- and time-based performance metrics, and (iii) assessment of the practical impact of LLM integration on quotation preparation efficiency and consistency.

The remainder of the paper is structured as follows. Section 2 presents related work, section 3 describes the quotation workflow and LLM integration concept, section 4 outlines the methodology, section 5 reports the results, section 6 discusses implications, and section 7 provides conclusions.

Related work

This section reviews relevant research on the application of AI in construction workflows, the use of LLMs for technical and domain-specific tasks, and existing AI-based approaches for cost estimation and quotation processes.

AI in construction workflows

Within construction informatics, research focuses on methods for creating, processing, and exchanging construction-specific information in both human- and machine-readable forms. AI has therefore been increasingly explored as a decision-support technology in construction management, particularly for scheduling, resource allocation, risk forecasting, and monitoring (Taboada et al., 2023).

Recent studies indicate that AI-powered systems can assist project teams by synthesising heterogeneous data sources, generating alternative scenarios, and providing early warnings of potential deviations (Mikhridinova et al., 2024). However, many existing approaches rely on predefined workflows and limited interaction capabilities, which restrict their adaptability in highly dynamic project environments.

LLMs for technical and domain-specific tasks

Large language models have demonstrated strong capabilities in understanding natural language and generating coherent technical text (Devlin et al., 2019;

Lewis et al., 2020). Their integration into AI agents enables advanced reasoning, task planning, and interaction with external tools through application programming interfaces (Ruan et al., 2023; Liu et al., 2024).

Recent surveys highlight the growing role of LLM-based agents as intelligent collaborators that can process diverse information, coordinate tasks, and support complex decision-making (Xi et al., 2025). Nevertheless, their effectiveness is constrained by the quality of training data, limited domain adaptation, and the stochastic nature of model outputs (Brelih et al, 2025). Consequently, human-in-the-loop (HITL) designs are widely recommended, particularly in safety-critical domains such as construction.

AI support for cost estimation and quotation processes

AI techniques have been explored for various cost-related tasks, including cost estimation, resource optimisation, and project staffing (Taboada et al., 2023; Mikhridinova et al., 2024). These studies demonstrate the potential of LLM-based systems to assist engineers in analysing project data and generating recommendations.

More recently, Gatto et al. (2025) propose an LLM-based methodology to automate the classification and extraction of cost-related descriptions in building projects, mapping textual specifications to a hierarchical taxonomy and populating a structured database to support downstream estimation tasks. Their work highlights how LLMs can reduce manual effort in interpreting and organising cost-relevant information, while also exposing the need for careful prompt design and validation in real-world workflows.

However, existing research primarily addresses cost prediction or high-level decision support rather than the end-to-end preparation of quotations. Little attention has been given to automating the selection of investigation or service items and the generation of quotation texts tailored to specific client inquiries.

Research gap

While prior research demonstrates the potential of AI and LLM-based agents to support construction workflows, there is a lack of empirical studies. This paper addresses this gap by evaluating multiple LLMs and proposing an LLM-supported quotation workflow tailored to domain-specific requirements for construction practice.

Quotation preparation workflow and LLM integration

In the company analysed, quotation preparation is a multi-step process initiated by a client inquiry received via email, which includes a brief project description, location, and requested services. An engineer reviews the inquiry, identifies the relevant type of geotechnical investigation, and manually selects appropriate service items from the internal service catalogue. Based on the selected items, quantities are estimated and prices are retrieved from internal pricing lists.

The engineer then prepares a quotation document that combines structured information (selected items, quantities, prices) with unstructured text describing the scope of work, assumptions, and conditions. The final quotation is reviewed and sent to the client. This workflow is largely manual and depends on individual experience, which can result in variability in both content and structure.

Although parts of the workflow follow predefined rules, several steps require expert judgement. Selecting of investigation or service items based on the inquiry content, estimating of quantities for individual items, and formulating of a coherent quotation text tailored to the specific project are non-deterministic steps and similar inquiries may result in different valid solutions depending on interpretation and experience.

LLM-supported workflow

To address these limitations, an LLM-supported workflow with human-in-the-loop (HITL) was designed. Upon receipt of a client inquiry, the inquiry text is provided to the LLM along with contextual information about available services, pricing structures, and domain-specific guidelines.

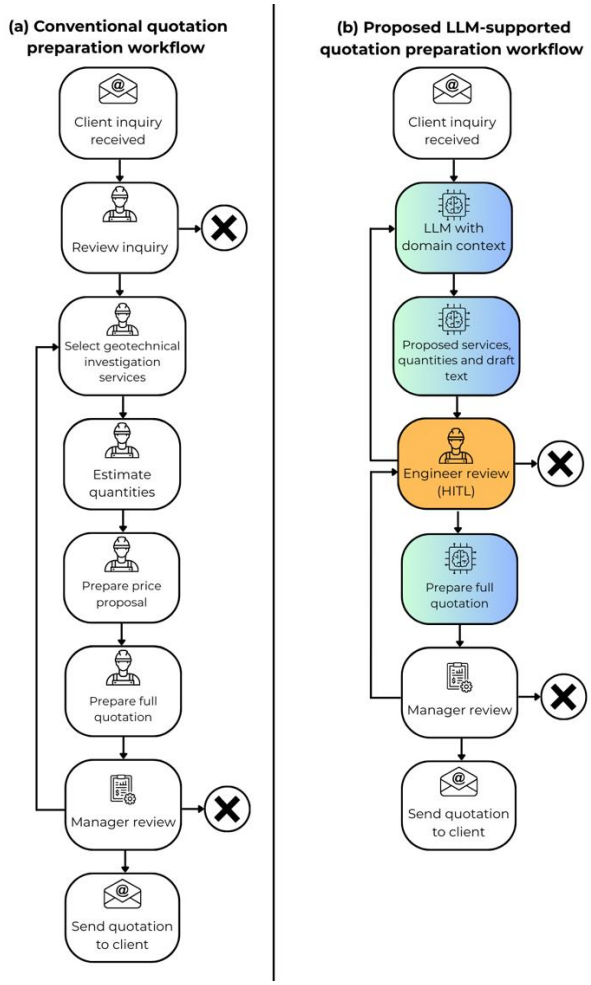

Figure 1: (a) Conventional quotation preparation workflow and (b) Proposed LLM-supported quotation preparation workflow.

The LLM generates a proposed set of service items, suggested quantities, and a draft quotation text. The engineer reviews the proposed output, makes corrections where necessary, and approves the final quotation. This HITL approach ensures that responsibility remains with the engineer while reducing the time required for routine tasks.

Quotation preparation combines structured data (service items, quantities, prices) with semi-structured and unstructured textual information (project descriptions, assumptions, and conditions). These characteristics make the process well suited for LLM, as LLMs are particularly effective at processing natural language and generating coherent technical text.

The repetitive nature of quotation preparation and the availability of historical quotations enable the creation of domain-specific training and evaluation datasets, supporting model adaptation and performance assessment.

The proposed workflow, shown in Figure 1, shifts the engineer's role from manual creation (a) to supervisory validation and refinement (b).

In this context, the HITL design is not just a safeguard but a critical architectural choice for deploying AI systems in construction workflows. The LLM serves as a cognitive assistant, while engineers retain full authority over validation, correction, and approval. This ensures that accountability and compliance remain firmly anchored in human expertise, which is essential in a highly regulated and safety-sensitive industry. HITL enables explainability, auditability, and trust, which are key requirements when AI-generated outputs may affect technical or financial decisions. Recent studies emphasized that in many industrial domains, fully autonomous AI agents may not be justified, and AI is more effective when deployed as a supervised contributor, particularly in contexts demanding trust, reliability, and regulatory compliance (He et al., 2024; Kotsiopoulos et al., 2024).

Methodology

We evaluated both commercial and open-source LLMs for supporting quotation preparation. The commercial models included GPT 4.1 and GPT 5 (OpenAI), and Claude Sonnet 4.5 (Anthropic). As an open-source baseline, we used QwQ 32B in (i) its base form and (ii) a fine-tuned version adapted to the company domain using anonymized examples from past project inquiries.

To assess the impact of supplementary project information, GPT 4.1 and GPT 5 were evaluated in two input configurations: (a) inquiry text only and (b) inquiry text plus attachments (when provided with the client inquiry). All models were prompted using a unified prompt template requiring a structured output (JSON) to ensure comparability across model families. This schema was designed to reflect real quotation workflows, and included fields such as project metadata, client inquiry content, and the desired response structure (e.g., service item list, estimated quantities, response text). Prompts were iteratively refined across models to ensure consistency. Domain knowledge was embedded into the prompts through a combination of static context injection (e.g., glossaries, service catalog excerpts, company-specific terminology) and heuristic templates derived from past quotation examples. Due to contractual and confidentiality constraints, external tool integration (e.g., retrieval-augmented generation) was not used, and no real-time application programming interface (API) calls were permitted.

Listing 1: Example prompt structure (abbreviated)

SYSTEM PROMPT:

Task: *Based on the client inquiry, generate a structured JSON listing of required geotechnical services, quantities, and a brief rationale.*

Context: *The company is based in Melbourne, Australia.*

Instructions:

1. *Review the client inquiry.*
2. *Select required work items from a predefined list.*
3. *Estimate quantities and justify selections in one sentence each.*
4. *Output a JSON array including item name, quantity, and justification.*

CLIENT INQUIRY *(excerpt):*

"…4 boreholes to 10 m or refusal. Provide retaining wall parameters and bearing capacity. Northern wall = gravity, Southern = MSE…"

EXPECTED OUTPUT FORMAT (JSON):

```
[
{"work_item": "Soil Drilling (<10m)", "quantity": 4},
{"work_item": "CBR Testing", "quantity": 3},
{"work_item": "Geotechnical Report", "quantity": 1}
]
```

The evaluated models were selected to represent both state-of-the-art commercial systems and high-capacity open-source alternatives. GPT 4.1 and GPT 5 (OpenAI) were included due to their status as leading commercial LLMs at the time of our research, with broad adoption and strong performance across technical NLP tasks. Claude Sonnet 4.5 (Anthropic) was selected based on reports of fast inference and strong reasoning in applied scenarios. QwQ 32B was chosen as a high-capacity open-source model suitable for domain fine-tuning. Its fine-tuned variant (QwQ 32B FT) was trained on domain-specific data to explore the feasibility of internal model adaptation without commercial licensing. Simpler baselines (e.g., BERT variants, rule-based systems) were not included, as the application scenario required generative capabilities and flexible outputs.

We compiled a dataset of 596 quotation inquiries collected over several years from a geotechnical

engineering firm. These inquiries vary in format, language quality, and complexity, reflecting realistic working conditions. From this dataset, we extracted a representative evaluation set of 120 inquiries, which were manually annotated with expected service items and quantities for comparison.

Fine-tuning of the open-source model

The open-source model was adapted using a two-stage, parameter-efficient fine-tuning approach:

1. Domain adaptation (DA). The model was adapted on domain texts describing geotechnical investigation service items (including professional item descriptions) to incorporate company terminology and service scope knowledge.
2. Supervised fine-tuning on input-output (inquiry-quotation) pairs (I/O). The base model, augmented with the DA adaptation, was further fine-tuned using supervised learning on curated inquiry-quotation examples. Each training example consisted of (i) an instruction plus the anonymised inquiry text as input and (ii) a strict JSON output containing the selected service items and their quantities as the learning target. This stage was designed to improve both domain-appropriate item selection and adherence to a consistent, machine-readable output structure.

The resulting fine-tuned model is referred to as QwQ 32B-FT in the remainder of the paper.

Dataset

The dataset consists of 596 real inquiry-quotation pairs collected from a geotechnical engineering company. Each pair contains (i) the anonymised client inquiry (email text and associated attachments, where applicable) and (ii) the corresponding engineer approved quotation, from which the reference service items and quantities were extracted.

The dataset was divided into 476 training pairs (80 %) and 120 test pairs (20 %). The training set was used exclusively for fine-tuning QwQ 32B-FT. The 120 pair test set was used for evaluation of all models.

Due to the signed NDA and client confidentiality, original data cannot be disclosed. However, a manually constructed mock example is provided below to illustrate the structure of a typical case:

Listing 2: Example inquiry-quotation pair

Inquiry (excerpt):

We are preparing a residential development in Werribee. Please include standard fieldwork, Atterberg limits, PSD tests, and recommendations for slab-on-ground foundations.

Output (excerpt):

[{"work_item": "Soil Drilling (<3m)", "quantity": 2},
{"work_item": "Atterberg Limits", "quantity": 2},
{"work_item": "PSD Testing", "quantity": 2},
{"work_item": "Geotechnical Report", "quantity": 1}]

Evaluation metrics

Performance was quantified using both quality and operational metrics. Service item selection is treated as a multi-label selection problem, evaluated using precision, recall, and F1 score. In addition, we measured quantity accuracy and processing time.

Quantity accuracy is the ratio of correctly determined quantities to all quantities proposed by the model for the selected items.

Processing time is the end-to-end time required to generate a complete structured output for the inquiry set. The paper reports the total time for 120 inquiries and average time per inquiry.

Results

This section reports model performance on the 120 inquiry test pairs using item selection metrics (precision, recall, F1), quantity accuracy, and processing time. Table 1 summarises the quantitative comparison across GPT 4.1, GPT 5, Claude Sonnet 4.5, QwQ 32B base, and QwQ 32B-FT. Figure 2 visualises the trade-off between effectiveness (F1) and speed (processing time).

GPT 5 with attachments achieved the highest F1 score (50%), driven by the highest recall (65 %), while Claude Sonnet 4.5 achieved a similar F1 score (48 %) with the fastest processing time (6.84 s/inquiry). The open-source baseline QwQ 32B performed substantially worse than commercial models on all quality metrics. QwQ 32B-FT improved precision (from 11 % to 21 %) and F1 score (from 8 % to 18 %) but remained far below the commercial LLMs. Quantity accuracy is low for all evaluated models (≤ 15 %), indicating that quantity determination remains a major bottleneck even when item selection improves.

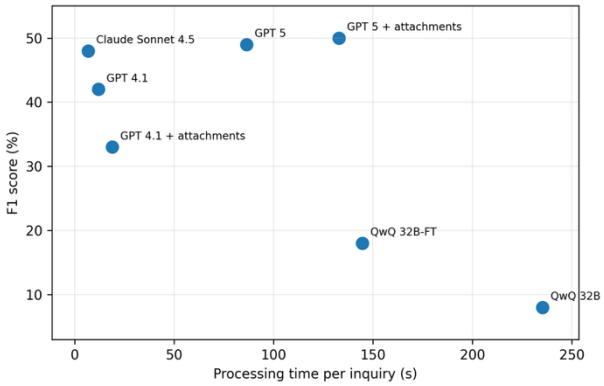

Figure 2: Scatter plot of F1 score relative to processing time.

Figure 2 shows that Claude Sonnet 4.5 offers the best balance between speed and quality (high F1 score at low latency), while GPT 5 achieves a slightly higher F1 score but with significantly longer runtime, especially when attachments are included. Figure 2 also shows that the fine-tuned version of the open-source model (QwQ 32B-FT) improved significantly in processing time as well as in F1 score, compared to the base model QwQ 32B.

Model	Precision (%)	Recall (%)	F1 score (%)	Quantity accuracy (%)	Total processing time (h)	Processing time per inquiry (s)
GPT 4.1	41	52	42	10	0,4	12,0
GPT 4.1 + attachments	38	38	33	8	0,63	18,9
GPT 5	43	63	49	13	2,88	86,4
GPT 5 + attachments	43	**65**	**50**	**15**	4,43	132,9
Claude Sonnet 4.5	**45**	55	48	11	**0,228**	**6,84**
QwQ 32B	11	7	8	1	7,84	235,2
QwQ 32B-FT	21	17	18	5	4,82	144,6

Table 1: Model performance on 120 inquiries.

Figure 3 shows that quantity accuracy remains consistently low across all evaluated models, with the best-performing configuration (GPT 5 with attachments) achieving only 15 %. Even models that perform well in service item selection and text generation do not reliably determine numerical quantities. This confirms that quantity estimation is the main bottleneck for higher levels of automation in quotation preparation and supports the use of hybrid workflows, where LLMs propose candidate items and draft text, while quantities are computed or verified using deterministic or rule-based tools.

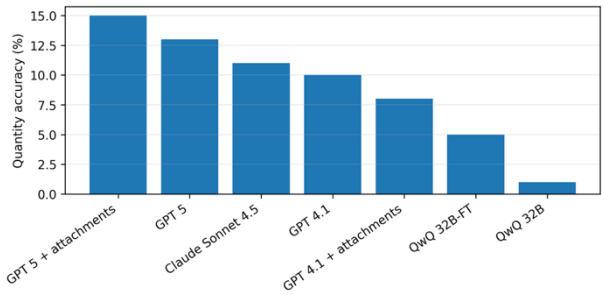

Figure 3: Quantity accuracy across evaluated model configurations.

Although all the accuracies were low, there is a notable increase in accuracy for the fine-tuned open-source model QwQ 32B-FT (4%) compared to the base model QwQ 32B (1%).

Discussion

The results indicate that commercial LLMs can meaningfully support quotation preparation by improving the accuracy of selecting geotechnical investigation service items, but they do not yet provide reliable quantity estimation without additional safeguards. Across models, the highest F1 score for item selection is approximately 50 %, meaning that even the best-performing configuration on average still misses or misidentifies a substantial portion of required items. This level of performance supports the proposed human-in-the-loop workflow: the models can accelerate drafting and reduce manual effort, but responsibility and final validation must remain with engineers.

A key practical finding is the quality-latency trade-off. GPT 5 with attachments achieves the best overall selection performance (highest F1 score and recall), but at a markedly higher processing time than other models. In contrast, Claude Sonnet 4.5 achieves an F1 score close to the best-performing GPT 5 configuration while being the fastest model in this study. For deployment in a real quotation preparation workflow, where the primary goal is to quickly provide a strong draft for engineer review, Claude Sonnet 4.5 appears to offer the most favourable operational profile.

The effect of attachments is not uniform. For GPT 5, attachments provide small improvements in recall and F1 score, suggesting that additional project context can help when the model can reliably integrate it. For GPT 4.1, however, adding attachments decreases performance sharply. This suggests that naively appending additional documents may introduce noise, cause relevant content to be diluted, or trigger truncation effects in long contexts.

The persistently low quantity accuracy across all models (≤15 %) is the most important limitation for practical adoption. This aligns with broader observations in construction-focused LLM research: generative models can produce fluent text and plausible suggestions, but numerical correctness and strict domain constraints remain challenging, especially without structured inputs and verification mechanisms. For quotation preparation, this motivates a hybrid architecture. LLMs can propose candidate items and draft explanatory text, while quantities should be computed or checked by deterministic tools (e.g. rule-based calculators) and engineer before finalisation.

The open-source model results clarify the current limits of local deployment. Although QwQ 32B-FT improved over the base QwQ 32B model, it did not reach competitive performance against commercial models. Moreover,

processing time remained high, undermining the operational benefits of automation. This does not preclude open-source adoption, but it indicates that achieving comparable performance likely requires larger models, higher-quality domain datasets, stronger retrieval augmentation, or more extensive fine-tuning, each requiring non-trivial engineering and infrastructure effort.

Finally, several limitations affect generalisability. The dataset originates from a single (geotechnical) company and a specific service catalogue, so model behaviour may differ in other contexts. The evaluation focuses on structured outputs (items and quantities), and the quality of the generated quotation text was not scored with a dedicated metric, although it is operationally important. Although the use case focuses on inquiries for geotechnical investigations, the results, and especially the workflow, are applicable across the entire construction domain.

While this study focuses on generative LLMs for their flexibility and alignment with comapny expectations, we acknowledge the absence of classical baselines such as BERT classifiers or rule-based matching systems. These approaches have been successfully applied to other construction-related tasks but were not included here due to the specific need for coherent text generation and adaptive item selection. Additionally, time and resource constraints limited the scope of model benchmarking. Future work could extend the evaluation to include retrieval-augmented generation or hybrid architectures that incorporate rule-based precision.

A qualitative review of model outputs suggests that accuracy was higher for inquiries with clear structure or phrasing patterns similar to previously encountered examples. In contrast, performance declined for inquiries that were vague, combined multiple project scopes, or included fragmented or informal language. Quantity errors mainly resulted from incomplete or ambiguous inputs. Many inquiries lacked explicit numerical requirements or used inconsistent units. These findings highlight the importance of structured input and human oversight when deploying LLMs in quotation workflows.

Limitations and Future work

This study is limited by its focus on a single company and service domain, which may affect generalisability to other construction contexts. Although models were evaluated using structured outputs, the quality of generated text was not assessed, despite its relevance in operational settings. Classical baselines and retrieval-augmented generation pipelines were excluded due to project scope and the requirement for flexible, generative outputs. Additionally, quantity estimation accuracy remains low across all models, highlighting the need for structured inputs and complementary rule-based tools. Future work could address these gaps by evaluating across multiple organisations, benchmarking against hybrid approaches, and assessing text quality and compliance in practical deployment.

Conclusions

This paper addresses the acceleration of quotation preparation using LLMs, where expert judgement is required for selecting appropriate service items and drafting quotation text. We designed an LLM-supported quotation preparation workflow and empirically evaluated several LLMs on a real dataset of inquiry-quotation pairs for geotechnical investigations, comparing commercial models (GPT 4.1, GPT 5, Claude Sonnet 4.5) with an open-source baseline (QwQ 32B) and its fine-tuned version (QwQ 32B-FT).

The results show that commercial LLMs significantly outperform open-source models in item selection, with GPT 5 (+ attachments) achieving the highest F1 score and Claude Sonnet 4.5 offering the best balance between F1 score and processing time. However, quantity accuracy is consistently low across all models, indicating that reliable quantity determination remains the main barrier to higher levels of automation.

Overall, the findings support the adoption of LLMs as drafting and decision-support tools within a human-in-the-loop process, while highlighting the need for hybrid architectures that combine LLM generation with deterministic quantity computation and validation. These findings underscore the value of HITL approaches not only as operational safeguards, but as foundational design choices that align AI support tools with the needs of high-stakes engineering domains. Future work should extend evaluation to multiple organisations, quantify text quality and compliance, and integrate rule-based or tool-assisted quantity estimation to reduce risk and improve end-to-end quotation reliability.

Acknowledgments

This research was supported by the Slovenian Research and Innovation Agency (ARIS) under the Young Researcher funding program and research program E-Construction (E-Gradbeništvo: P2-0210).

References

Brelih, A., Robnik Šikonja, M., Klinc, R. (2025) The Role of AI Agents in Construction Project Management, Tagungsband des 36. Forum Bauinformatik : 24.-26. September 2025 = Proceedings of the 36th Forum Bauinformatik 2025 / herausgegeben von Baris Özcan, Hristo Vassilev, p. pages 147–154. Available at: https://doi.org/10.18154/RWTH-CONV-254903.

Devlin, J., Chang, M-W., Lee, K., Toutanova, K. (2019) BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv. Available at: https://doi.org/10.48550/arXiv.1810.04805.

Gatto, C., Mirarchi, C., Pavan, A. (2025) Towards Automation in Cost Estimation: LLM-based Methodology for Classifying and Extracting Cost Data from Tender Documents, in. 2025 European Conference on Computing in Construction. Available at: https://doi.org/10.35490/EC3.2025.370.

He, Y., Wang, E., Rong, Y., Cheng, Z., Chen, H. (2024) Security of AI Agents. arXiv. Available at: https://doi.org/10.48550/arXiv.2406.08689.

Kotsiopoulos, T., Papakostas, G., Vafeiadis, T., Dimitriadis, V., Nizamis, A., Bolzoni, A., Bellinati, D., Ioannidis, D., Votis, K., Tzovaras, D., Sarigiannidis, P. (2024) Revolutionizing defect recognition in hard metal industry through AI explainability, human-in-the-loop approaches and cognitive mechanisms, Expert Systems with Applications, 255, p. 124839. Available at: https://doi.org/10.1016/j.eswa.2024.124839.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W-T., Rocktäschel, T., Riedel, S., Kiela, D. (2020) Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks, in Advances in Neural Information Processing Systems. Curran Associates, Inc., pp. 9459–9474. Available at: https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html.

Liu, Z., Yao, W., Zhang, J., Yang, L., Liu, Z., Tan, J., Choubey, P. K., Lan, T., Wu, J., Wang, H., Heinecke, S., Xiong, C., Savarese, S. (2024) AgentLite: A Lightweight Library for Building and Advancing Task-Oriented LLM Agent System. arXiv. Available at: https://doi.org/10.48550/ARXIV.2402.15538.

Mikhridinova, N., Yurdakul, F.A., Wolff, C. (2024) Using Large Language Models for Project Staffing: Evaluation of GPT-Based Mapping of Teams to Projects, in Proceedings of the 12th IPMA Research Conference Project Management in the Age of Artificial Intelligence. 12th IPMA Research Conference Project Management in the Age of Artificial Intelligence, International Project Management Association – IPMA, IPMA USA, pp. 1–19. Available at: https://doi.org/10.56889/cykr4175.

Ruan, J., Chen, Y., Zhang, B., Xu, Z., Bao, T., Du, G., Shi, S., Mao, H., Li, Z., Zeng, X., Zhao, R. (2023) TPTU: Large Language Model-based AI Agents for Task Planning and Tool Usage. arXiv. Available at: https://doi.org/10.48550/ARXIV.2308.03427.

Taboada, I., Daneshpajouh, A., Toledo, N., de Vass, T. (2023) Artificial Intelligence Enabled Project Management: A Systematic Literature Review, Applied Sciences, 13(8), p. 5014. Available at: https://doi.org/10.3390/app13085014.

Turk, Ž. (2006) Construction informatics: Definition and ontology, Advanced Engineering Informatics, 20(2), pp. 187–199. Available at: https://doi.org/10.1016/j.aei.2005.10.002.

Xi, Z., Chen, W., Guo, X., He, W., Ding, Y., Hong, B., Zhang, M., Wang, J., Jin, S., Zhou, E., Zheng, R., Fan, X., Wang, X., Xiong, L., Zhou, Y., Wang, W., Jiang, C., Zou, Y., Liu, X., Yin, Z., Dou, S., Weng, R., Cheng, W., Zhang, Q., Qin, W., Zheng, Y., Qiu, X., Huang, X., Gui, T. (2025) The rise and potential of large language model based agents: a survey, Science China Information Sciences, 68(2), p. 121101. Available at: https://doi.org/10.1007/s11432-024-4222-0.