2025 European Conference on Computing in Construction

CIB W78 Conference on IT in Construction

Porto, Portugal
July 14-17, 2025

A CONCEPTUAL FRAMEWORK FOR LLM-BASED MULTI-AGENT SYSTEMS IN CONSTRUCTION MANAGEMENT

Alok Rawat, Emlyn Witt, and Irene Lill
Tallinn University of Technology, Tallinn, Estonia

Abstract

Construction management faces challenges like delays and coordination issues. This research proposes a Large Language Model Multi-Agent System (LLM-MAS) framework tailored for construction to enhance automation and decision support. The framework features a Central Coordinator managing specialized LLM agents and utilizes a hybrid memory system combining Knowledge Graphs (KGs) for structured data (schedules, dependencies) and Vector Databases for documents, enabling advanced reasoning (Graph RAG). We demonstrate feasibility through an experiment simulating logistics delay detection using LangGraph, Neo4j (KG), and Gemini LLMs. The system successfully ingested simulated site data, updated the KG, used graph reasoning to identify the delay's impact on activities, and generated alerts. This work contributes the conceptual framework and provides initial validation, highlighting the potential of KG-integrated LLM-MAS to improve construction project coordination and efficiency.

Introduction

Construction management perennially grapples with significant challenges, including costly project delays, budget overruns, and the intricate coordination required among diverse stakeholders (Daoud et al., 2023; Dlamini & Cumberlege, 2021). While existing digital tools like Building Information Modeling (BIM) and the Internet of Things (IoT) offer valuable capabilities, they face limitations that hinder comprehensive solutions. Hurdles such as high implementation costs, the need for specialized skills, interoperability issues between platforms (Bui et al., 2016; Tang et al., 2019), and challenges with connectivity, data security, and scalability on dynamic construction sites (Altohami et al., 2021) highlight the need for more integrated and intelligent automation approaches. Globally, 70% of construction projects experienced time delays in 2017, with overruns ranging from 10% to 30% of the planned schedule (Memon et al., 2023).

Recent advancements in LLMs and Multi-Agent Systems (MAS) present a promising avenue to address these complexities. LLMs offer powerful natural language understanding and generation capabilities suitable for knowledge extraction and decision support, while MAS provide frameworks for coordinating multiple intelligent entities to solve complex, distributed problems (Guo et al., 2024). The integration of LLM reasoning capabilities within MAS architectures LLM-MAS holds significant potential to enhance the automation of complex tasks requiring nuanced understanding and dynamic coordination, moving beyond the limitations of earlier rule-based agent systems often used in construction (Xiang et al., 2022). The key advantage of using these systems is to allow automation and abstraction of complex workflows involving BIM, IoT and other technologies to make it easier to adopt.

However, there is a gap in the development and application of integrated LLM-MAS specifically tailored to the unique demands and data ecosystems of construction management. To address this gap, this paper aims to achieve two primary objectives. Firstly, it proposes a conceptual framework that integrates LLM-powered agents, multi-modal memory systems, and mechanisms for accessing construction-specific data sources and external tools. This framework seeks to leverage the combined strengths of LLMs and MAS to create a more intelligent and adaptive approach to managing construction projects. Secondly, the paper aims to demonstrate the practical feasibility of this approach by implementing and evaluating an agent system for a specific, relevant construction task: tracking construction material flow using site sensor data, computer vision, and project documents.

This work's primary contribution is therefore twofold: it provides a conceptual blueprint for applying LLM-MAS in the construction domain and offers initial empirical evidence of this approach's feasibility through a focused experiment on a logistics delay detection and impact analysis.

Consider the construction scenario depicted in Figure 1, where a manager needs to understand the impact of a 5-day delay in steel beam delivery. A generic LLM might offer a basic textual summary of likely consequences. Improving upon this, a Retrieval-Augmented Generation (RAG) workflow would typically react to this specific query by employing techniques like Chain-of-Thought prompting to identify needed information, then using tools to *retrieve* relevant data – querying a knowledge graph for dependent tasks (erect_steel_frame, install_roof_decking) and alternative material availability – to synthesize a more detailed and factually grounded response. This RAG process, however, remains fundamentally reactive and focused on answering the posed question. In contrast, the agent workflow, which embodies the principles of the LLM-MAS framework proposed in this paper, demonstrates a more proactive and autonomous capability. Instead of solely reacting to a query, an agent continuously integrates diverse, multi-modal inputs (such as IoT data flagging low current_stock of steel beams against a min_threshold, ERP data showing the Structural Phase starts in 1 day, and site camera feeds). This triggers an internal reasoning process ("Critical steel inventory below safety margin") leading to planned, goal-oriented *actions*. The agent utilizes tools not just for querying (MaterialCheckTool, ScheduleImpactTool) but for *execution*, such as initiating an UrgentResupplyRequest, potentially involving automated vendor communications and requests for human approval for expedited shipping. This example highlights the key shift: from RAG's reactive information retrieval to the agent's proactive, multi-modal data integration, internal state management, decision-making, and capacity for autonomous action execution, offering a more powerful paradigm for managing complex, dynamic construction environments.

Research Gap

Pre-2022 research (Before ChatGPT) extensively explored agent-based modeling (ABM) and MASs to tackle the complexities of construction management, as highlighted in reviews by (Ren & Anumba, 2004; Xiang et al., 2022). Key applications included resource allocation (Nair et al., 2018) , project scheduling (Ren & Anumba, 2004), and supply chain optimization. However, these earlier approaches faced limitations such as high computational demands, difficulties in accurately modeling complex human behaviors, and challenges in model validation due to the dynamic nature of construction (Xiang et al., 2022). Compared to current LLM-based agents, which can leverage vast amounts of data for more nuanced understanding and interaction, these older agentic systems relied on predefined rules and often struggled with adaptability and complex reasoning in dynamic real-world scenarios (Rojas & Mukherjee, 2006).

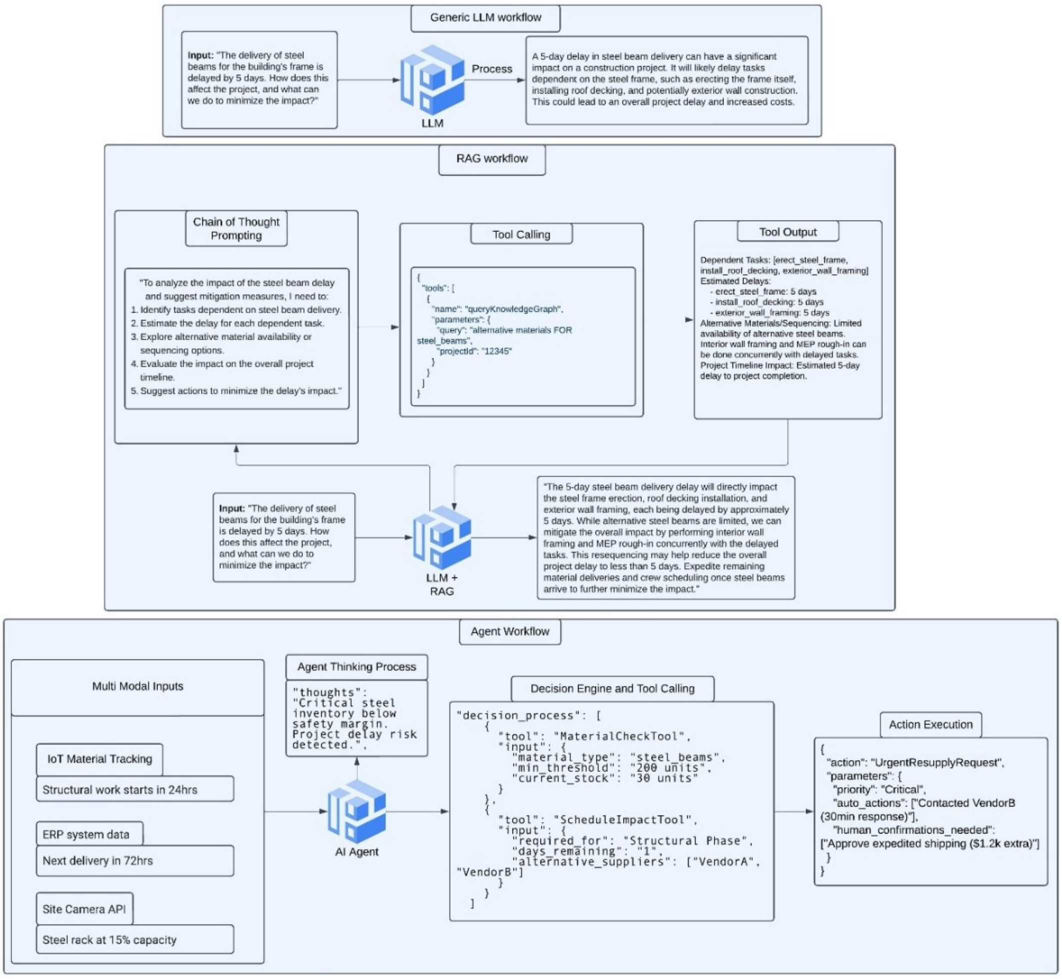

Figure 1: Comparison of Generic LLM, RAG and Agentic workflow

Methodology

The conceptual framework presented in this paper was developed following a focused literature review targeting recent advancements relevant to intelligent automation in construction. Key academic databases (Scopus, ArXiv) were searched for publications from 2022 onwards using keywords as (LLM AND agent*). Relevant papers were identified partly through manual review (ArXiv) and partly using an LLM (Gemma 3 27b-it via API) to filter metadata exported from Scopus, ensuring coverage of the rapidly evolving landscape of agentic AI systems.

The core findings from this review, encompassing LLMs for natural language processing and reasoning (He et al., 2024), MASs for distributed coordination (Guo et al., 2024), advanced agent memory architectures, Retrieval-Augmented Generation (RAG) for grounding responses in external data (Lewis et al., n.d.), and KGs for structured data representation (Pan et al., 2023), were synthesized to design the proposed system architecture, depicted in Figure 2. The goal of this synthesis was to create a cohesive architecture specifically tailored to address the challenges inherent in construction management, such as dynamic project environments, the need to process multi-modal data, and the management of complex task dependencies. The resulting framework thus integrates these concepts to enable more intelligent automation and decision support in construction.

Furthermore, recognizing the importance of standardized and modular tool integration for agentic systems, the framework design acknowledges emerging standards like the Model Context Protocol (MCP) Anthropic. (2024, November 24). MCP represents an open standard aiming to connect LLMs with diverse external data sources and tools through a client-server architecture. While specific construction-focused MCP tools are currently nascent, the protocol's emphasis on modularity, stateful context, and two-way communication aligns well with the principles of the agentic execution of tasks proposed in our framework, suggesting a potential future pathway for implementing robust and interoperable tool use for construction agents. A detailed discussion of specific tool implementation is deferred to the experimental setup section.

Conceptual Framework

The design of the proposed LLM-MAS framework (Figure 2) incorporates several architectural choices aimed at balancing capability with manageability within the complex construction domain. A key decision was the adoption of a centralized coordination model, facilitated by the Central Coordinator component acting as a hub. While decentralized agent architectures offer potential benefits in robustness and scalability (Li et al., 2024), a centralized approach was deemed more suitable as an initial blueprint for construction management. This structure mirrors the hierarchical nature of typical construction projects, simplifying the mapping of project goals and oversight responsibilities onto the agent system (Han et al., 2024; Li et al., 2024). Furthermore, the Central Coordinator provides a single point for managing overall system state, resolving conflicts between specialized agents, and ensuring coherent progress towards global project objectives, which can be challenging to achieve reliably through purely emergent peer-to-peer interactions in complex, dynamic environments. This centralized flow also significantly aids in tracing agent interactions and diagnosing issues during development and operation, a practical consideration for deploying complex AI systems. However, we acknowledge the potential limitation of the Central Coordinator acting as a performance bottleneck or single point of failure, a trade-off accepted in this conceptual model for enhanced control and debuggability (Li et al., 2024).

Another significant design choice is the mediated access to memory and execution tools, where agents primarily interact with the Memory Hub and Execution Layer via the Central Coordinator. The alternative would be to grant agents direct access. The mediated approach was chosen primarily to enforce data consistency and coherence, particularly crucial when multiple agents might need to access or update shared project information (e.g., schedules, resource status) stored in the Memory Hub. It also allows for centralized management of resources (e.g., API rate limits, computational resources for tool use), security policy enforcement, and provides a clear audit trail of actions performed by each agent, again facilitating easier debugging and analysis (Han et al., 2024). While this mediated access introduces potential latency compared to direct interaction, the benefits of control and consistency were prioritized in this initial framework design. Future iterations or specific implementations could explore hybrid models; for instance, granting certain trusted agents direct read-only access to specific Long-Term Knowledge stores or frequently used, low-risk tools might optimize performance for specific tasks, though this would require careful design of concurrency and consistency protocols.

The framework also proposes a differentiation between the Agent Pool between Core and Auxiliary Agents. Core Agents (e.g., Resource Allocation, Cost Management, Quality Control) are conceptualized to handle fundamental, domain-critical project management functions directly impacting project outcomes. Auxiliary Agents (e.g., Documentation, Reporting, Legal/Contract) provide essential supporting functions, often involving information processing, summarization, or compliance checks. This separation promotes modularity, allowing specialized development and independent updating of agents based on their function. It clarifies the system's structure and potentially facilitates scaling by adding new agents within either category without fundamentally altering the core operational logic managed by the Coordinator and Core Agents (Ren & Anumba, 2004).

Communication between the Central Coordinator and specialized agents, as well as interactions with the Execution Layer tools, primarily relies on structured data formats (e.g., JSON, as exemplified in Figure 1 and the

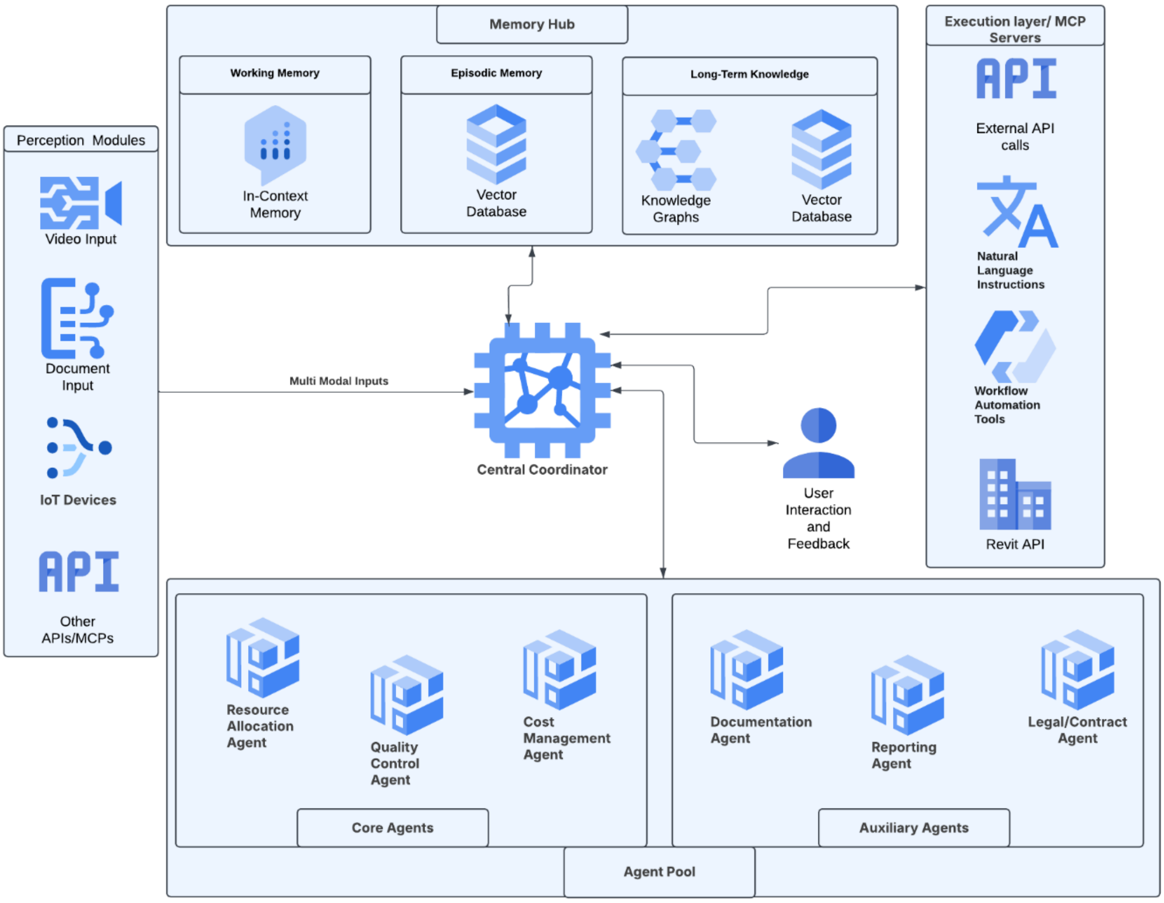

Figure 2: Proposed Conceptual Framework for Construction Management

experimental implementation) for core system operations. This approach is generally favored over free-form natural language communication for these interactions to ensure clarity, reduce ambiguity, and enhance reliability crucial for dependable automation in construction workflows. Structured data facilitates consistent parsing, validation, and seamless integration with tools requiring precise inputs. While direct natural language communication between agents for executing critical tasks poses risks of misinterpretation, NL could potentially serve specific roles within the system, always mediated and managed by the Central Coordinator. For instance, agents might leverage NL for complex collaborative problem-solving requiring nuanced negotiation before reaching a structured plan, or for generating descriptive summaries intended for other agents' consumption rather than direct action execution. However, for the majority of deterministic tasks, especially those involving tool use and requiring verifiable execution, structured communication remains the foundational modality within this framework to ensure operational integrity (Du et al., 2024; Sarmah et al., 2024).

Furthermore, the framework incorporates a multi-component memory structure within the Memory Hub, recognizing that effective agent reasoning and adaptation require different types of information storage and retrieval mechanisms, analogous to concepts in cognitive architectures (Wang et al., 2024). A single monolithic memory is insufficient for the diverse information needs of construction management tasks. The Working Memory component is designed to hold immediate context relevant to the active task, facilitating focused reasoning and efficient processing, potentially leveraging techniques like Cache Augmented Generation (CAG) (Chan et al., 2024). The Episodic Memory, implemented here as a vector database storing summaries of past interactions, queries, and successful actions, enables agents to learn from experience. By retrieving similar past episodes, agents can potentially improve decision-making or action selection over time (Han et al., 2024; Wang et al., 2024). This episodic memory enables agents to learn from experience by retrieving relevant past events. Crucially, for practical application and continuous improvement, this memory should explicitly store not only summaries of past actions but also the associated user feedback (such as confirmations, corrections, or outcome evaluations). This collected feedback provides a rich learning signal directly within the operational loop, and importantly, creates a valuable dataset that can subsequently be leveraged for offline fine-tuning of the underlying LLMs powering the agents, potentially using techniques like Reinforcement Learning from Human Feedback (RLHF), to further align agent behavior with domain-specific requirements and user preferences over time (Liesenfeld et al., 2023).

Finally, the Long-Term Knowledge store itself utilizes a dual approach, combining KGs and Vector Databases. KGs are included for their strength in representing explicit entities and relationships within structured construction data (e.g., project schedules, component dependencies), enabling complex, structured queries (Pan et al., 2024). Complementing this, Vector Databases facilitate efficient semantic search over large volumes of unstructured or semi-structured documents (e.g., contracts, specifications, daily logs) via Retrieval-Augmented Generation (RAG).

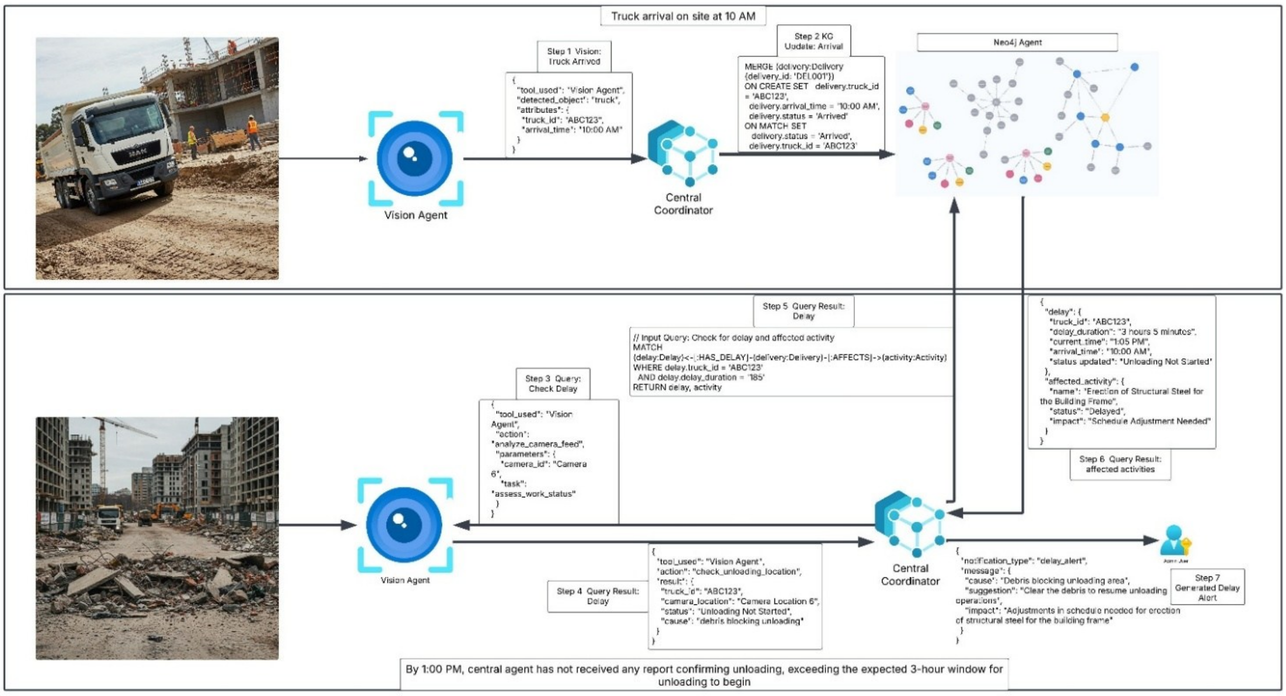

Figure 3: Experimental simulation of the proposed framework

This hybrid Long-Term Knowledge approach allows the system to leverage both the structured relationships and the rich textual information pervasive in construction projects (Chan et al., 2024; Gao et al., 2023).

Experimental work

To demonstrate the practical feasibility and evaluate key aspects of the proposed conceptual framework, an experimental setup was implemented focusing on a representative construction site scenario (Figure 3): the real-time detection of a logistics delay and the subsequent analysis of its impact on scheduled activities. This scenario was chosen to test the framework's ability to integrate perception data, update a knowledge base, perform graph-based reasoning, and generate timely notifications. The implementation leveraged LangGraph as the primary orchestration engine, simulating the role of the Central Coordinator, and utilized a Neo4j graph database accessed via the py2neo library to instantiate the Knowledge Graph component within the framework's Memory Hub. To ensure visually relevant and contextually appropriate inputs, representative construction site images were generated using Google's Imagen 3 model. Corresponding synthetic data, used to simulate site events and populate the initial state of the knowledge graph, was generated using the Google Gemini 2.0 Flash model, ensuring a controlled and reproducible test environment.

The LangGraph framework managed the state transitions and coordinated the interactions between different functional modules, embodying the Central Coordinator's function. The Neo4j database stored nodes representing relevant construction entities such as deliveries, trucks, site locations, scheduled activities, and potential issues (e.g., delays, obstructions), along with the relationships connecting them (e.g., HAS_DELAY, AFFECTS, LOCATED_AT). The py2neo library provided the interface for the LangGraph workflow to dynamically interact with this knowledge graph, enabling both state updates using Cypher MERGE commands based on incoming events, and complex information retrieval using Cypher MATCH queries to reason about dependencies and impacts.

Within this setup, specific agent functionalities were simulated as distinct nodes or processes within the LangGraph graph. The underlying intelligence for these simulated agents (e.g., Vision Agent interpretation logic, Coordinator reasoning, Notification Agent formatting) was powered by the Google Gemini 2.0 Flash model. Each agent's behavior was defined through carefully crafted system prompts outlining its role, inputs, expected outputs, and constraints. To promote consistent and predictable behavior suitable for an automated workflow, the temperature parameter for the Gemini 2.0 Flash model was set to 0.35, favoring more deterministic outputs. A 'Vision Agent' function, using the configured LLM, processed simulated visual inputs (represented by the Imagen 3 generated images and associated synthetic metadata) and output structured JSON data. The core reasoning and decision-making logic, representing the Central Coordinator possibly invoking specialized agents, resided within the LangGraph workflow, also leveraging the Gemini 2.0 Flash model. This logic interpreted inputs, determined actions, formulated Cypher queries, processed graph data, and decided on the final output. A final node simulated a 'Notification Agent', using the LLM to format results into structured alerts. All interactions were mediated through LangGraph's state management, adhering to the framework's principle of centralized coordination.

The experimental workflow demonstrated the end-to-end process: simulated perception data (JSON derived from synthetic inputs) triggered updates to the Neo4j knowledge graph via py2neo. Subsequently, based on predefined logic, the LangGraph coordinator initiated a Cypher query to identify the delay status and affected activities. The results retrieved from Neo4j were processed by the coordinator logic (using Gemini 2.0 Flash), and a final notification detailing the delay, cause, and impact was generated. This setup allowed for the evaluation of the framework's capacity for integrating event data, maintaining a dynamic knowledge base, performing graph-based reasoning, and generating actionable information using contemporary LLMs within an orchestrated workflow.

Discussion

The experimental simulation successfully demonstrated the integration of core components from the proposed LLM-MAS framework, utilizing LangGraph, Neo4j, and Gemini LLMs to address a representative logistics delay scenario. The implemented workflow effectively processed simulated perception data, updated the Neo4j KG to reflect the event, employed LangGraph for coordination, executed basic Cypher queries against the KG to identify the delay and its direct impact on a linked activity based on predefined relationships, and subsequently generated a structured notification. This process validates the fundamental roles envisaged for the central coordinator and the KG within the architecture for this type of monitoring task, showcasing a potential pathway for automating information flow and basic impact analysis, even though the graph queries used in this specific instance were relatively straightforward.

While this experiment confirms the viability of the centralized coordination approach for integrating these specific components in the demonstrated workflow, its scope was limited to a single scenario without complex multi-agent collaboration. Crucially, the simulation focused solely on the KG component of the proposed hybrid memory system; the utility and integration of vector databases for handling unstructured data, as envisioned in the conceptual framework, were not part of this evaluation and require separate investigation. Therefore, although these initial results obtained in a controlled environment are encouraging, further research involving real-world data, more complex reasoning and query patterns, testing the full hybrid memory architecture, and exploring dynamic multi-agent interactions is necessary to fully assess the potential of LLM-MAS to enhance construction project efficiency and proactive decision-making.

Limitations

Despite promising findings, several limitations must be acknowledged. The experiment relied on synthetic data (generated via Imagen 3 and Gemini 2.0 Flash), lacking the noise and complexity of real-world construction data; validation with actual project data is essential. The experimental scope was narrow, focusing on one scenario and not evaluating the full framework, including complex multi-agent interactions or sophisticated tool use. The 'Vision Agent' functionality was also simulated rather than using live computer vision.

Regarding the conceptual framework, the Central Coordinator poses potential scalability bottlenecks and a single point of failure, a trade-off made for initial control. Mediated access to memory and tools, while ensuring consistency, might introduce latency unsuitable for some real-time applications.

Technological limitations also persist. System performance depends on LLM reliability (Gemini 2.0 Flash), with risks of hallucination and challenges in explainability remaining. Experimentation with real word data will present challenges as well as opportunities for enhancing such agentic systems iteratively.

Conclusion

This research addressed the persistent challenges of inefficiency and complexity in construction management by proposing and partially validating an integrated framework based on LLM-MAS. We presented a conceptual architecture featuring a Central Coordinator, specialized LLM-powered agents, modular tool access, and a hybrid memory system incorporating Knowledge Graphs and Vector Databases for advanced reasoning and retrieval (Graph RAG). The primary contribution is twofold: the proposal of this tailored conceptual framework and the demonstration of its core feasibility through a focused experiment.

The experimental simulation, utilizing LangGraph, Neo4j, and Google's Gemini models, successfully demonstrated the system's potential for a practical construction scenario – detecting a logistics delay from simulated perception data and reasoning about its impact on scheduled activities using the Knowledge Graph. This provides initial empirical support for the framework's ability to integrate diverse data sources, maintain a dynamic knowledge base, and perform automated reasoning to generate actionable insights, moving beyond the capabilities of traditional methods or simpler automation approaches.

While acknowledging limitations related to the use of synthetic data, the experiment's limited scope, and potential scalability concerns inherent in the centralized architecture, this work establishes a valuable foundation. It highlights the promise of LLM-MAS, particularly when combined with structured knowledge representations like KGs, to enhance coordination, improve real-time awareness, and support more informed decision-making on construction projects. Future research should focus on validating the framework with real-world project data, expanding the repertoire of specialized agents and integrated tools (e.g., direct BIM API interaction), exploring hybrid coordination models to address scalability, and further investigating methods for ensuring reliability and explainability. Ultimately, the continued development and application of such intelligent systems hold significant potential to transform construction management practices, leading towards more efficient, predictable, and resilient project delivery.

Acknowledgements

This research has been co-funded by the Housing Decarbonisation Skills for Climate, Health, and Jobs (Skills4Deca) project—grant number: DIGITAL-2022-SKILLS-03-SPECIALISEDEDU-101123311 of the European Union's Digital Europe Programme and the Empowering Trainers for Digital innovation in construction Ecosystem (ET4D) project—grant number: 2024-1-IT01-KA220-VET-000249119 of the Erasmus Plus Programme of the European Union. The European Commission's support in producing this publication does not constitute an endorsement of the contents that reflect the views only of the authors, and the Commission cannot be held responsible for any use that may be made of the information contained therein.

References

Altohami, A. B. A., Haron, N. A., Ales@Alias, A. H., & Law, T. H. (2021). Investigating approaches of integrating BIM, IoT, and facility management for renovating existing buildings: A review. *Sustainability (Switzerland)*, *13*(7). https://doi.org/10.3390/su13073930

Bui, N., Merschbrock, C., & Munkvold, B. E. (2016). A Review of Building Information Modelling for Construction in Developing Countries. *Procedia Engineering*, *164*, 487–494. https://doi.org/10.1016/j.proeng.2016.11.649

Chan, B. J., Chen, C.-T., Cheng, J.-H., & Huang, H.-H. (2024). *Don't Do RAG: When Cache-Augmented Generation is All You Need for Knowledge Tasks*. https://doi.org/10.1145/3701716.3715490

Daoud, A. O., El Hefnawy, M., & Wefki, H. (2023). Investigation of critical factors affecting cost overruns and delays in Egyptian megaconstruction projects. *Alexandria Engineering Journal*, *83*, 326–334. https://doi.org/10.1016/j.aej.2023.10.052

Dlamini, M., & Cumberlege, R. (2021). The impact of cost overruns and delays in the construction business. *IOP Conference Series: Earth and Environmental Science*, *654*(1). https://doi.org/10.1088/1755-1315/654/1/012029

Du, H., Thudumu, S., Vasa, R., & Mouzakis, K. (2024). *A Survey on Context-Aware Multi-Agent Systems: Techniques, Challenges and Future Directions*. http://arxiv.org/abs/2402.01968

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, M., & Wang, H. (2023). *Retrieval-Augmented Generation for Large Language Models: A Survey*. http://arxiv.org/abs/2312.10997

Guo, T., Chen, X., Wang, Y., Chang, R., Pei, S., Chawla, N. V., Wiest, O., & Zhang, X. (2024). *Large Language Model based Multi-Agents: A Survey of Progress and Challenges*. http://arxiv.org/abs/2402.01680

Han, S., Zhang, Q., Yao, Y., Jin, W., Xu, Z., & He, C. (2024). *LLM Multi-Agent Systems: Challenges and Open Problems*. http://arxiv.org/abs/2402.03578

He, Y., Tang, Y., & Chen, T. (2024). A Study on Large Language Model-Based Approach for Construction Contract Risk Detection. *ACM International Conference Proceeding Series*, 136–141. https://doi.org/10.1145/3696500.3696523

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. (n.d.). *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks*. https://github.com/huggingface/transformers/blob/master/

Li, X., Wang, S., Zeng, S., Wu, Y., & Yang, Y. (2024). A survey on LLM-based multi-agent systems: workflow, infrastructure, and challenges. *Vicinagearth*, *1*(1), 9. https://doi.org/10.1007/s44336-024-00009-2

Liesenfeld, A., Lopez, A., & Dingemanse, M. (2023). Opening up ChatGPT: Tracking openness, transparency, and accountability in instruction-tuned text generators. *Proceedings of the 5th International Conference on Conversational User Interfaces, CUI 2023*. https://doi.org/10.1145/3571884.3604316

Memon, A. H., Memon, A. Q., Khahro, S. H., & Javed, Y. (2023). Investigation of Project Delays: Towards a Sustainable Construction Industry. *Sustainability (Switzerland)*, *15*(2). https://doi.org/10.3390/su15021457

Nair, A. S., Hossen, T., Campion, M., Selvaraj, D. F., Goveas, N., Kaabouch, N., & Ranganathan, P. (2018). Multi-Agent Systems for Resource Allocation and Scheduling in a Smart Grid. In *Technology and Economics of Smart Grids and Sustainable Energy* (Vol. 3, Issue 1). https://doi.org/10.1007/s40866-018-0052-y

Pan, S., Luo, L., Wang, Y., Chen, C., Wang, J., & Wu, X. (2023). *Unifying Large Language Models and Knowledge Graphs: A Roadmap*. https://doi.org/10.1109/TKDE.2024.3352100

Pan, S., Luo, L., Wang, Y., Chen, C., Wang, J., & Wu, X. (2024). Unifying Large Language Models and Knowledge Graphs: A Roadmap. *IEEE Transactions on Knowledge and Data Engineering*, *36*(7), 3580–3599. https://doi.org/10.1109/TKDE.2024.3352100

Ren, Z., & Anumba, C. J. (2004). Multi-agent systems in construction-state of the art and prospects. In *Automation in Construction* (Vol. 13, Issue 3, pp. 421–434). https://doi.org/10.1016/j.autcon.2003.12.002

Rojas, E. M., & Mukherjee, A. (2006). Multi-Agent Framework for General-Purpose Situational Simulations in the Construction Management Domain. *Journal of Computing in Civil Engineering*, *20*(3). https://doi.org/10.1061/(asce)0887-3801(2006)20:3(165)

Sarmah, B., Hall, B., Rao, R., Patel, S., Pasquali, S., & Mehta, D. (2024). *HybridRAG: Integrating Knowledge Graphs and Vector Retrieval Augmented Generation for Efficient Information Extraction*. http://arxiv.org/abs/2408.04948

Tang, S., Shelden, D. R., Eastman, C. M., Pishdad-Bozorgi, P., & Gao, X. (2019). A review of building information modeling (BIM) and the internet of things (IoT) devices integration: Present status and future trends. In *Automation in Construction* (Vol. 101, pp. 127–139). Elsevier B.V. https://doi.org/10.1016/j.autcon.2019.01.020

Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., Chen, Z., Tang, J., Chen, X., Lin, Y., Zhao, W. X., Wei, Z., & Wen, J. (2024). A survey on large language model based autonomous agents. In *Frontiers of Computer Science* (Vol. 18, Issue 6). Higher Education Press Limited Company. https://doi.org/10.1007/s11704-024-40231-1

Xiang, L., Tan, Y., Shen, G., & Jin, X. (2022). Applications of multi-agent systems from the perspective of construction management: A literature review. *Engineering, Construction and Architectural Management*, *29*(9), 3288–3310. https://doi.org/10.1108/ECAM-01-2021-0038