CCC 2025 **Proceedings of the Creative Construction Conference** (2025)
Edited by: Miroslaw J. Skibniewski & Miklós Hajdu & Žiga Turk
https://doi.org/10.22260/CCC2025/0061

LLM-DRIVEN HUMAN-ROBOT INTERACTION WITH DIGITAL TWINS FOR FACILITY MANAGEMENT

Peihang Luo[1], Samuel A. Prieto[2], Erika Parn[1], Borja García de Soto[2], Ioannis Brilakis[1]

1 University of Cambridge, Cambridge, United Kingdom
2 S.M.A.R.T. Construction Research Group, Division of Engineering, New York University Abu Dhabi (NYUAD), United Arab Emirates (UAE)

Abstract

This paper presents a framework for integrating user interaction, building digital twins, and robotic automation to enhance facility management. The system leverages a Large Language Model (LLM) as the central interface, enabling the user to intuitively retrieve data from the digital twin and command the robotic agents for facility inspection and monitoring tasks. Commands from the user are processed by the LLM, which translates them into actionable tasks. These tasks are then interpreted by the robotics middleware and executed autonomously by robots equipped with navigation and data acquisition capabilities. The collected data is then presented to human operators, who can use it to update the digital twin and inform maintenance decisions. By combining the natural language processing power of LLMs with digital twin-based data and robotic automation, the proposed framework reduces manual effort while streamlining facility inspections and supporting maintenance decision-making in facility management. A theoretical case study demonstrates the system's capabilities, illustrating its ability to process user queries, allocate robotic tasks, collect and deliver inspection data, and support informed decision-making. This approach bridges the gap between human decision-making, digital representations, and physical site operations, offering a user-friendly solution for modern facility management.

Keywords: digital twin, facility management, human-robot interaction, large language model, operations and maintenance.

Peer-review under responsibility of the scientific committee of the Creative Construction Conference 2025.

1. Introduction

Effective facility management relies heavily on routine inspection of building assets to ensure safety, functionality, and compliance. As buildings become increasingly complex, the demand for efficient, accurate, and timely inspection methods has grown. Manual inspection methods are time-consuming, labor-intensive, and prone to inconsistency. At the same time, maintaining the accuracy and timeliness of digital twins, which are virtual representations of physical assets, requires a continuous flow of updated, ground-truth information from the physical assets. This creates a growing need for more efficient, scalable, and intelligent approaches to facility inspection and digital twin alignment.

While robotic systems have the potential to automate physical facility inspection tasks, their operation often requires technical expertise and training, limiting their accessibility for non-technical users such as facility managers and building operators [1]. There is a lack of intuitive, user-friendly interfaces that allow non-technical users to control robotic systems in an intuitive way.

Recent advances in Large Language Models (LLMs) provide an opportunity to simplify this interaction. LLMs can offer a natural language interface that interprets and translates unstructured, high-level user commands into structured task plans for robotic execution [2]. When integrated with building digital twins and robotic systems, LLMs can help non-technical users command robots to perform inspection tasks using high-level natural language.

Corresponding author email address: pl452@cam.ac.uk

In this paper, we propose a framework that enables users to initiate facility inspection tasks through natural language commands, which are processed by an LLM in conjunction with a building digital twin. The LLM interprets the prompt, retrieves relevant spatial and semantic context, and generates structured instructions for robotic execution. After the inspection is completed, the results are returned to the user, who then approves the update for the digital twin record, with all changes logged for traceability. We demonstrate this approach through a theoretical case study of a single-story building, demonstrating the system's ability to support non-technical user interaction, automate inspection, and maintain digital twin timeliness and alignment.

2. Literature review

2.1. Digital twin-based robotic inspection

Ground robots have emerged as a practical solution for routine inspection and maintenance in buildings, driven by the need to reduce labor, improve safety, and maintain up-to-date digital records of facility condition. A recent systematic review of 269 publications by Halder et al. [3] shows that, after aerial systems, unmanned ground vehicles (UGVs) are the most frequently studied platform for built-environment inspection. The review also highlights autonomous navigation, multi-sensor data capture, and integration with BIM (Building Information Modeling) or digital twins (DTs) as dominant research themes. These findings establish a broad research agenda: coupling mobile robots with rich building information to close the feedback loop between physical assets and their virtual representations.

Early demonstrations of that coupling adopt a digital-twin-enabled teaming paradigm. Lu et al. [4] develop a bi-directional DT that streams real-time robot telemetry to facility managers while allowing operators to dispatch a mobile robot for patrolling, cleaning, and inspection. They manage to report faster anomaly detection and improved situational awareness compared with manual workflows. Baniqued et al. [5] extend the concept to multi-robot fleets in high-risk nuclear facilities, presenting an immersive DT that fuses building geometry, live sensor data, and mission plans in both desktop and virtual-reality interfaces. Operators can therefore monitor and re-task robots without entering hazardous zones. These studies underline the value of DTs as a shared information space that mediates human oversight and robotic autonomy.

Whereas DTs offer global context, knowledge-driven task reasoning ensures that the robot can act locally. Chen et al. [6] introduce the iSTA (integrated Scene-Task-Agent) model, which formalizes inspection know-how stored in a BIM and fuses it with human knowledge to plan robot missions such as fire-door and lighting checks. By querying the BIM for component semantics (location, type, inspection frequency) the robot generates waypoints and camera poses autonomously, reducing manual task programming. Follini et al. [7] also leverage BIM to pre-compute collision-free paths and embed inspection results back into the model, demonstrating how static design data can bootstrap robot deployment during both construction and O&M (Operations and Maintenance) phases.

Advances in on-board perception further increase robot utility. Ge and Sadhu [8] equip a UGV with LiDAR (Light Detection and Ranging) and monocular vision and apply deep learning to localize and classify cracks and spalling. The detected defects are automatically mapped onto a 3D point cloud to create an annotated digital twin for structural health monitoring. Such systems demonstrate that mobile robots can produce high-value semantic data (i.e., imagery, geometry and defects metadata) needed by facility managers.

Collectively, current research proves that ground robots, enriched by BIM and DT information, can autonomously collect facility data and feed it back to digital records. However, task requests still rely on predefined schedules, rule-based action planners, or direct teleoperation. None of the surveyed work offers a natural-language interface that recognizes information gaps in the DT, reformulates user intent into robotic tasks, and returns structured results to close the loop automatically. Addressing this gap, our framework introduces an LLM as a cognitive broker between the user, the digital twin, and the robot. By building on the technical foundations outlined above, the proposed LLM-driven architecture aims to deliver truly on-demand inspections and data updates for modern facility management.

2.2. LLM-based human-robot interaction and task planning

Recent advances in LLMs have enabled new possibilities for intuitive human-robot interaction by enabling users to issue high-level commands in natural language. Unlike traditional interfaces that rely on graphical user interfaces (GUIs), predefined commands or structured inputs, LLMs can interpret flexible and high-level prompts provided by the users and translate them into actionable steps that can be directly executed by robots. In the context of robotic systems, this capability supports the generation of task plans that can be executed autonomously, bridging the gap between human intent and robotic action. This section reviews recent work at the intersection of LLMs, natural language interfaces, and task planning for robotic applications.

Recent work has explored the integration of LLMs into robotic systems to support more natural, flexible, and efficient human-robot collaboration. Gkournelos et al. [9] propose an LLM-based manufacturing execution system that enables seamless communication between human operators and industrial robots in smart assembly tasks. This system combines a natural language interface, real-time data integration through digital twins, and behavior-based robotic control, significantly improving collaboration efficiency in real-world use cases. Similarly, LLMs have been applied to streamline robot programming and task execution, reducing reliance on rigid command structures by enabling free-form natural language instructions [10]. In addition, Prieto and García de Soto [11] developed a dual-LLM agent framework for robotic task allocation in construction. By combining a generator and a supervisor agent, the system is able to produce more accurate, constraint-aware task schedules. These studies demonstrate the growing potential of LLMs to bridge human intent and robotic action, especially in contexts where adaptability and ease of instruction are critical.

Beyond human-robot interaction, recent work has explored how LLMs can support robotic task planning in complex, real-world environments. Luo et al. apply a hierarchical code generation framework using LLMs to automate the control of construction assembly robots, integrating high- and low-level logic with domain-specific APIs to improve program reliability and flexibility across diverse tasks [12]. Prieto and García de Soto propose a multi-agent LLM architecture for task planning and allocation in construction robotics, where separate Planner and Supervisor agents collaborate to generate and validate instructions, improving execution accuracy and reducing hallucinations in dynamic environments [13]. In addition, the ELLMER (embodied large-language-model-enabled robot) framework integrates LLMs with sensorimotor feedback to enable robots to execute long-horizon tasks in unpredictable settings, combining retrieval-augmented generation and real-time contextual adaptation [14]. These works highlight the emerging role of LLMs in bridging high-level reasoning and low-level robotic control, offering solutions for robust and adaptable task planning.

Overall, the integration of LLMs into human-robot interaction and task planning presents a promising direction for enabling more flexible, intelligent, and user-friendly robotic systems. Current research demonstrates the potential of LLMs to interpret user intent, generate executable plans, and adapt to complex and dynamic environments. However, applying these models in the context of building digital twins introduces challenges related to the model’s ability to understand the building’s spatial layouts and semantic information, which requires effective integration with structured digital representations. Additionally, mechanisms are required for maintaining humans in the loop to validate and trace data updates. LLMs, in particular, have the potential to empower users to direct robotic agents in facility inspection and monitoring tasks, supporting the continuous alignment between physical conditions and their digital representations and helping to keep digital twins accurate and up to date.

3. Methodology and Implementation

This work presents a methodology that leverages LLMs to facilitate human-robot interaction for indoor facility inspection and monitoring tasks. These tasks typically involve identifying and evaluating physical elements such as HVAC (Heating, Ventilation, and Air Conditioning) units, electrical panels, fire extinguishers, and doors to support facility management activities like safety checks and condition assessments. The intended users are facility or building managers, who often lack operational knowledge of robotic systems. By enabling interaction through natural language commands, the system reduces the complexity of directing robotic agents and interpreting inspection outcomes.

An overview of the system methodology is shown in Figure 1. The architecture integrates a natural language interface powered by an LLM, a robotic middleware layer, and a robotic agent. Users initiate the process by submitting natural language prompts describing inspection needs. The system interprets these prompts to extract context, translates them into structured instructions, and communicates the tasks to the robotic agent via the middleware. The robot performs the required operations, such as navigation and data collection, and the collected data is then returned to the user for validation. Based on user input, updates are made to the digital twin of the facility, with all changes logged to support transparency and traceability.

To demonstrate the application of this approach, a single-story building was used as a testbed, focusing on facility elements such as furnishings, fire safety equipment, HVAC systems, and electrical infrastructure. When a user prompt is received (e.g., “Retrieve information about the fire extinguisher in the hallway”), the LLM interprets the task intent and retrieves relevant spatial and semantic metadata from the building’s IFC (Industry Foundation Classes) model. If the required data is available and up to date, it is retrieved and displayed to the user. If not, the system initiates a data collection process by generating a structured instruction in JSON format, which is later parsed by the robotic middleware and autonomously executed by the robot.

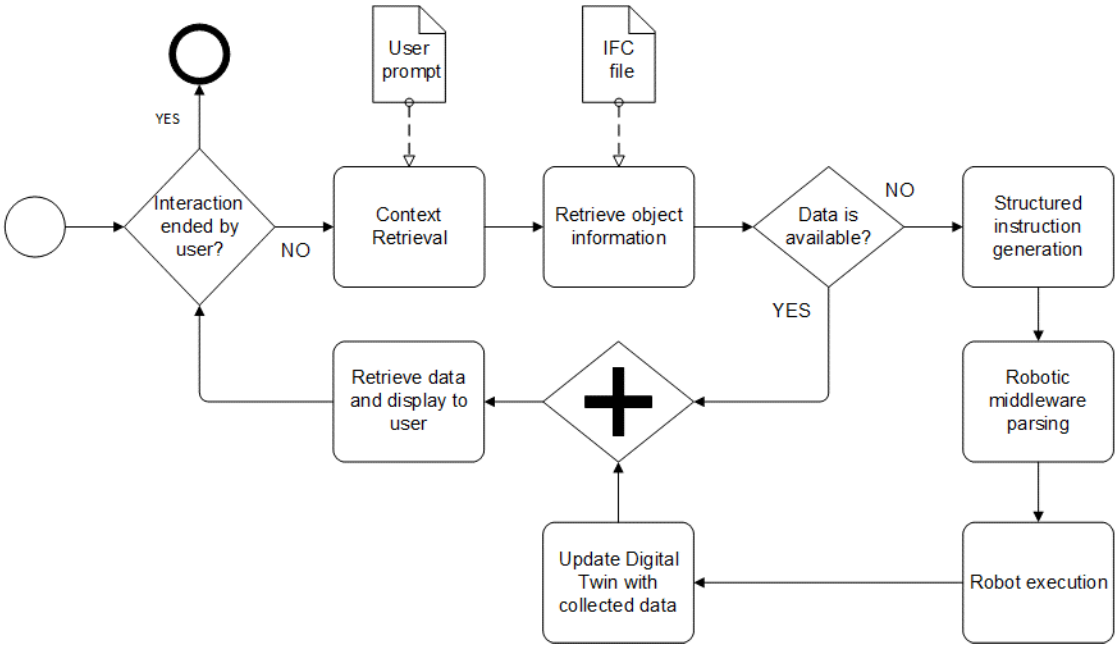


Fig. 1. Overview of the system methodology, showing the sequential flow from user interaction to robotic execution and digital twin update.

For new data collection, the LLM generates a structured instruction set specifying the inspection task, including the target object, its location (x, y, yaw), and required sub-tasks such as navigation and image capture. A representative output in JSON (JavaScript Object Notation) format may look like this:

```
{
  "task": "inspection",
  "target_object": "<target object name>",
  "location": {"x": <X coordinate>, "y": <Y coordinate>, "yaw": <YAW orientation>},
  "sub_tasks": ["navigate", "capture_image"]
}
```

This structured output is parsed by the ROS (Robot Operating System) middleware, which translates high-level instructions into commands for the robot's control system. The robotic agent executes the assigned operations and transmits the results to the system interface for user review. Upon validation, the data is used to update the digital twin. All updates are logged with metadata, such as timestamps, user identity, and affected components, to maintain a complete and auditable record of all changes.

4. Results and Discussion

To validate the proposed framework, we conducted a case study using a single-story residential building modeled in IFC format. The building contained common spaces including a bedroom, living room, kitchen, and bathroom, as shown in Figure 2. The inspection task focused on verifying the presence and condition of a fire extinguisher located in the hallway.

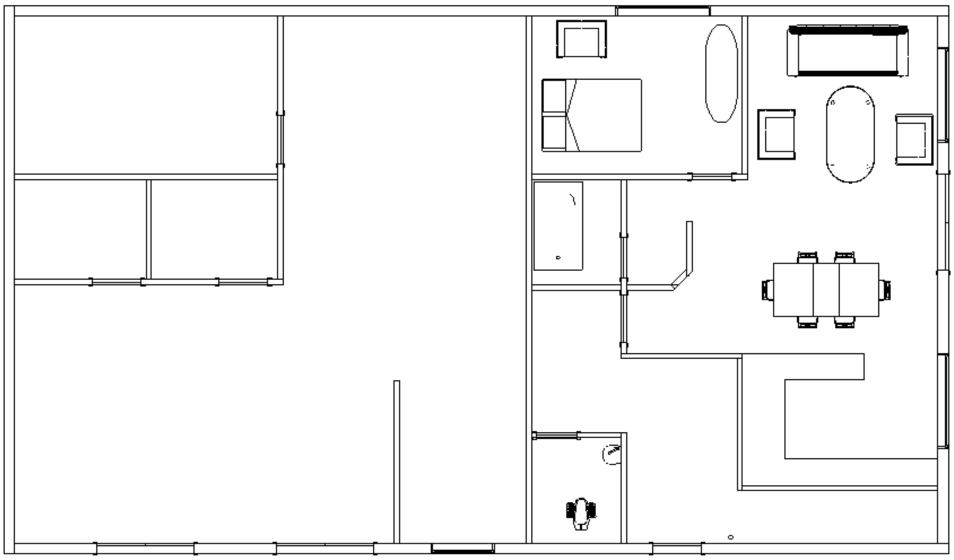
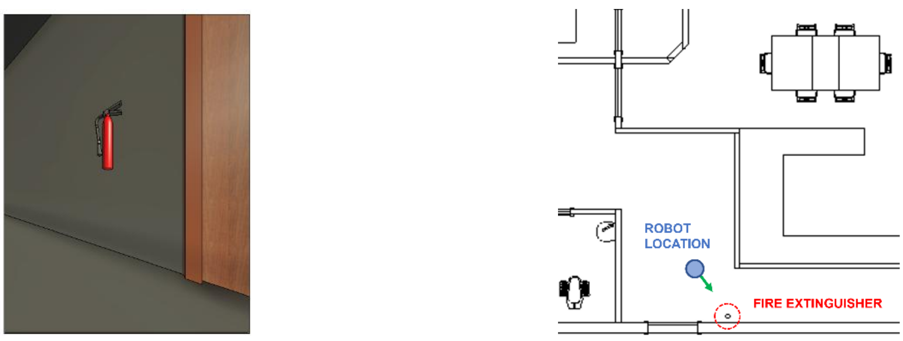


Fig. 2. Floor plan of the building.

A natural language prompt, "*Please inspect the fire extinguisher in the hallway*," was issued to the system. The LLM processed this input, identifying the task type (inspection), the target object (fire extinguisher), and the spatial context (hallway). By querying the IFC model, the system retrieved the relevant spatial layout and object metadata to generate a structured instruction set for robotic execution. The structured instruction generated was as follows:

```
{
  "task": "inspection",
  "target_object": "Fire_Safety-Nystrom-ABC_Fire_Extinguisher:EX-3002",
  "location": {"x": 3.538, "y": -1.848, "yaw": -0.820},
  "sub_tasks": ["navigate", "capture_image"]
}
```

This instruction was passed to the robotic middleware, which parsed the input and translated it into executable commands. The robot then autonomously navigated to the hallway using the coordinates provided in the instructions, and captured an image at the given yaw orientation as shown in Figure 3.



a b

Fig. 3. Inspection result, (a) Image taken by the robot and (b) location of the robot with respect to the fire extinguisher when taking the picture.

The inspection result, which is a photograph of the fire extinguisher (as shown in Figure 3a), will be returned to the user. After reviewing the captured data, the user will then validate it and update the digital twin record, confirming the fire extinguisher's presence and adding a timestamped log entry linked to the inspection task.

This case study demonstrates the feasibility of using LLMs to bridge user intent and robotic execution in facility inspection tasks. It highlights the system's ability to retrieve spatial context from digital models, generate interpretable instructions, and support human-in-the-loop digital twin updates. While the task was relatively simple, the workflow illustrates key strengths of the approach: reducing user effort, simplifying robot operation, and supporting continuous digital twin alignment.

5. Conclusion

This paper presented a framework that utilizes an LLM to enable natural language interaction with robotic agents for facility inspection tasks, supported by a building digital twin. Through a case study of a facility inspection task in a single-story building, we demonstrated the feasibility of translating user prompts into structured robotic instructions, executing autonomous inspection tasks, and enabling human-in-the-loop digital twin updates.

The results highlight the potential of LLM-based interfaces to make robotic inspection systems more accessible and user-friendly, particularly for non-technical facility managers. By integrating semantic and spatial information from digital building models, the framework supports context-aware task planning and simplifies the process of maintaining up-to-date digital twin records.

Future research will explore expanding the system's capabilities to support more complex inspection workflows, multi-object task planning, anomaly detection during inspection, and more dynamic update mechanisms between physical environments and their digital counterparts. Further investigation is also needed to enhance the reliability, safety, and interpretability of LLM-driven robotic operations in real-world facility management scenarios.

Acknowledgements

This research was partially supported by different Centers at NYUAD. In particular, the Center for Sand Hazards and Opportunities for Resilience, Energy, and Sustainability (SHORES) funded by Tamkeen under the NYUAD Research Institute Award CG013, the Center for Interacting Urban Networks (CITIES), funded by Tamkeen under the NYUAD Research Institute Award CG001, and the Center for Artificial Intelligence and Robotics (CAIR), funded by Tamkeen under the NYUAD Research Institute Award CG010. The authors also gratefully acknowledge the EPSRC, Bentley Systems and the University of Cambridge for funding this research through the EPSRC Centre for Doctoral Training in Future Infrastructure and Built Environment: Resilience in a Changing World (EPSRC grant reference number EP/S02302X/1).

References

[1] J. M. Davila Delgado *et al.*, "Robotics and automated systems in construction: Understanding industry-specific challenges for adoption," *Journal of Building Engineering*, vol. 26, p. 100868, Nov. 2019, doi: 10.1016/j.jobe.2019.100868.

[2] C. Zhang, J. Chen, J. Li, Y. Peng, and Z. Mao, "Large language models for human–robot interaction: A review," *Biomimetic Intelligence and Robotics*, vol. 3, no. 4, p. 100131, Dec. 2023, doi: 10.1016/j.birob.2023.100131.

[3] S. Halder and K. Afsari, "Robots in Inspection and Monitoring of Buildings and Infrastructure: A Systematic Review," *Applied Sciences*, vol. 13, no. 4, Art. no. 4, Jan. 2023, doi: 10.3390/app13042304.

[4] W. Lu, J. Chen, Y. Fu, Y. Pan, and F. A. Ghansah, "Digital twin-enabled human-robot collaborative teaming towards sustainable and healthy built environments," *Journal of Cleaner Production*, vol. 412, p. 137412, Aug. 2023, doi: 10.1016/j.jclepro.2023.137412.

[5] P. D. E. Baniqued *et al.*, "Multimodal immersive digital twin platform for cyber–physical robot fleets in nuclear environments," *Journal of Field Robotics*, vol. 41, no. 5, pp. 1521–1540, 2024, doi: 10.1002/rob.22329.

[6] J. Chen, W. Lu, Y. Fu, and Z. Dong, "Automated facility inspection using robotics and BIM: A knowledge-driven approach," *Advanced Engineering Informatics*, vol. 55, p. 101838, Jan. 2023, doi: 10.1016/j.aei.2022.101838.

[7] C. Follini *et al.*, "BIM-Integrated Collaborative Robotics for Application in Building Construction and Maintenance," *Robotics*, vol. 10, no. 1, p. 2, Dec. 2020, doi: 10.3390/robotics10010002.

[8] L. Ge and A. Sadhu, "Deep learning-enhanced smart ground robotic system for automated structural damage inspection and mapping," *Automation in Construction*, vol. 170, p. 105951, Feb. 2025, doi: 10.1016/j.autcon.2024.105951.

[9] C. Gkournelos, C. Konstantinou, and S. Makris, “An LLM-based approach for enabling seamless Human-Robot collaboration in assembly,” *CIRP Annals*, vol. 73, no. 1, pp. 9–12, Jan. 2024, doi: 10.1016/j.cirp.2024.04.002.

[10] N. Dimitropoulos, P. Papalexis, G. Michalos, and S. Makris, “Advancing Human-Robot Interaction Using AI – A Large Language Model (LLM) Approach,” in *Advances in Artificial Intelligence in Manufacturing*, A. Wagner, K. Alexopoulos, and S. Makris, Eds., Cham: Springer Nature Switzerland, 2024, pp. 116–125. doi: 10.1007/978-3-031-57496-2_12.

[11] S. A. Prieto and B. García de Soto, “Large Language Models for Robot Task Allocation,” *ISARC Proceedings*, vol. ICRA 2024 Future of Construction Workshop Papers, pp. 17–20, May 2024, doi: 10.22260/ICRA2024/0007.

[12] H. Luo, J. Wu, J. Liu, and M. F. Antwi-Afari, “Large language model-based code generation for the control of construction assembly robots: A hierarchical generation approach,” *Developments in the Built Environment*, vol. 19, p. 100488, Oct. 2024, doi: 10.1016/j.dibe.2024.100488.

[13] S. A. Prieto and B. García de Soto, “Collaborative Large Language Models for Task Allocation in Construction Robots,” Jan. 14, 2025, *Social Science Research Network, Rochester, NY*: 5097309. doi: 10.2139/ssrn.5097309.

[14] R. Mon-Williams, G. Li, R. Long, W. Du, and C. G. Lucas, “Embodied large language models enable robots to complete complex tasks in unpredictable environments,” *Nature Machine Intelligence*, pp. 1–10, Mar. 2025, doi: 10.1038/s42256-025-01005-x.