CCC 2025 **Proceedings of the Creative Construction Conference** (2025)
Edited by: Miroslaw J. Skibniewski & Miklós Hajdu & Žiga Turk
https://doi.org/10.22260/CCC2025/0041

ENHANCING HUMAN-MACHINE INTERACTION WITH ON-DEVICE LARGE LANGUAGE MODELS IN CONSTRUCTION ROBOTICS: A CASE STUDY ON SAFETY APPLICATIONS

Muhammad Anas Gopee, Samuel A. Prieto, Borja García de Soto
S.M.A.R.T. Construction Research Group, New York University Abu Dhabi (NYUAD), United Arab Emirates (UAE)

Abstract

Large Language Models (LLMs) are increasingly being explored for robotics applications, enabling more natural and intuitive communication between humans and machines. In the construction industry, where automation has lagged behind other sectors and the workforce often lacks specialized robotics training, LLMs offer a promising solution for improving human-robot interaction. However, most existing implementations rely on cloud-based processing, which introduces challenges such as network latency, unreliable connectivity at remote sites, and concerns over security and privacy when transmitting sensitive project data. These limitations make cloud-dependent LLMs impractical for real-world construction environments, which are dynamic and unpredictable, requiring robust and responsive systems. With recent improvements in model efficiency, LLMs are becoming lighter and with better performance, making local execution on edge devices increasingly feasible. This study investigates the feasibility of deploying LLMs entirely on-device for construction robotics. We propose a locally executed framework that processes multimodal inputs such as speech and vision, directly on construction robots. By eliminating reliance on external servers, this approach ensures close to real-time responses, greater autonomy, and enhanced resilience in bandwidth-constrained settings. The research describes the framework, hardware integration, and the application of straightforward natural language prompts to enable practical multimodal processing in field conditions. Our results demonstrate how locally executed LLMs can provide a robust, responsive, and accessible interface for human-robot collaboration in construction. Additionally, we demonstrate in a case study how few-shot prompting can be used within our framework to enable a Local Intelligent Safety Assistant (LISA) that inspects work areas, interacts with workers in real time, and helps ensure compliance with on-site safety measures.

Keywords: multimodal, distilled models, edge computing, human-machine interaction, construction safety

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

1. Introduction

The construction industry is adopting technological solutions to address challenges such as safety risks, labor shortages, and project complexity [1]. Robotics and machine learning (ML) are increasingly integrated into construction workflows to enhance efficiency and safety. While traditional ML has been successfully applied in different construction domains [2], its reliance on application-specific training and extensive data gathering and pre-processing makes it time-consuming and inflexible for dynamic construction environments [3]. For instance, ensuring compliance with safety standards using ML models typically involves training algorithms on labelled data specific to each application, such as PPE (personal protective equipment) usage, making them less adaptable to unforeseen scenarios [4]. This limitation becomes particularly evident in dynamic construction sites where tasks and risks vary significantly, and it has become a barrier to adoption in the sector. Another barrier is the lack of intuitive human-machine interfaces (HMI) in current robotics solutions [5]. Conventional control methods often demand technical expertise beyond that of many on-site workers, hindering effective collaboration and technology adoption.

Corresponding author email address: nas@nyu.edu

Large Language Models (LLMs) offer a promising alternative to the limitations of traditional ML approaches in construction automation, as well as improving Human-Machine Interaction (HMI) on-site [6]. Unlike conventional ML systems, LLMs demonstrate remarkable generalizability and can perform complex reasoning tasks with minimal additional training, including zero-shot applications [7]. LLMs can also significantly enhance HMI in construction settings as they inherently possess excellent natural language interfaces, reducing the technical expertise required from operators, and their ability to integrate contextual information makes them well-suited for the dynamic conditions of construction environments. These versatile capabilities are increasingly being applied to various use cases within the construction industry [8]. Initially, LLMs were too large and resource-intensive for practical on-site deployment, requiring cloud-based solutions with high-performance GPUs that depend on a reliable internet connection [9], [10]. Construction sites, particularly those in remote locations or within structures with high signal attenuation, frequently experience connectivity challenges and bandwidth constraints that make cloud-dependent systems unreliable [11]. The latency involved in transmitting data to cloud servers and receiving responses can also compromise real-time applications. Recent advances, such as knowledge distillation, have enabled smaller, more efficient LLMs that maintain strong performance and can run locally on edge devices [12].

This raises a compelling question: How close are we to having an LLM capable of providing reliable, human-level reasoning for complex tasks in the construction industry, while also being lightweight enough for deployment on local edge devices? Achieving this could significantly advance the capabilities of construction robotics and expand the practical use of autonomous systems in complex, real-world environments. This paper investigates the potential of locally executed LLMs as an alternative to traditional ML to improve HMI and for safety applications in construction robotics. We present the Local Intelligent Safety Assistant (LISA), a framework designed to process multimodal inputs and deliver actionable safety insights based on user-defined guidelines with few-shot prompting. By focusing on scenarios like PPE compliance and safety advice, we assess whether local LLMs can offer a viable, adaptable, and easy-to-implement solution compared to traditional ML approaches.

2. Literature Review

This section provides a review of literature relevant to deploying large language models in robotics with a focus on edge computing implementation, as well as robotics in construction safety applications. The review looks into current approaches, identifies gaps, and establishes the foundation for our research contribution.

2.1. Large Language Model for Robotics

Recent surveys and review articles, including those by Zeng et al. [13], Kim et al. [14], and Kawaharazuka et al. [15], demonstrate the potential of large language models (LLMs) in robotics, particularly in enhancing adaptable reasoning and human–machine interaction. However, these advancements have traditionally relied on cloud servers for computational power, which introduces challenges such as unpredictable latency, connectivity dropouts in weak-signal environments, and data-privacy concerns. This reliance on the cloud remains a significant barrier to seamless deployment, especially in scenarios where real-time performance is critical.

Addressing these limitations, more focused studies, such as those by Zheng et al. [16] and Friha et al. [17], explore methods to adapt LLMs for on-device use. Techniques like quantization, pruning, optimized inference pipelines, and tight hardware–software co-design aim to shrink models while maintaining functionality. Quantization-aware training approaches, as demonstrated by Kong et al. [18], further reduce accuracy loss compared to naive compression methods, offering a more refined path to edge deployment. Sikorski et al. [10] investigate the feasibility of deploying a quantized LLaMA 2-7B model on edge devices to control a mobile robot, though their approach often struggles with command misinterpretations, inconsistent execution, and is limited to simple tasks. Another path, as proposed by Yu et al. [9], focuses not on shrinking LLMs, but rather on designing custom hardware architectures specifically optimized for efficiently running large models.

Despite these advances, challenges persist. Edge deployments that rely on pruning and post-hoc quantization often suffer noticeable drops in accuracy, as noted by Zheng et al. [16] and Friha et al. [17], and rarely undergo rigorous testing on truly multimodal or noisy inputs, which are common in real-world robotic applications. While quantization-aware retraining can partially mitigate these issues, it introduces additional complexity to the training process and still struggles with tasks requiring heavy reasoning in language or vision domains. Furthermore, work like that of Sikorski et al. [10], are increasingly being outpaced by newer distilled models that offer better efficiency, reduced memory footprints, and improved task performance, signaling a move away from traditional quantization techniques in edge-oriented LLMs.

Custom hardware solutions, such as the Cambricon-LLM architecture developed by Yu et al. [9], offer high efficiency but come with their own hurdles. These include the need for specialized application-specific integrated circuit (ASIC) designs, high development costs, and limited support from broader standardized ecosystems. Leveraging off-the-shelf edge platforms such as NVIDIA Jetson could simplify adoption and improve portability, making edge deployment more accessible and practical.

2.2. Robotics for construction safety

Recent academic studies highlight that robotics and machine learning can be used to enhance construction safety. For example, Bharathi et al. [19] showed that autonomous ground vehicles with deep learning and domain adaptation can automate the detection of safety guardrails, making inspections faster and more reliable. Dornik and Bartels [20] found that UAVs and legged robots improve site inspections and hazard detection, supporting safer work in hazardous or hard-to-reach areas. Lee and Chien [21] introduced a mobile robot system that combines deep learning with visual SLAM to detect PPE compliance and unsafe behaviors directly on site, offering improved flexibility and accuracy over fixed camera systems. Collectively, these advances automate hazard detection, minimize manual inspection risks, and support proactive safety management.

Despite these advancements, significant limitations hinder widespread adoption. One major issue is that many of these systems rely on data-intensive methodologies, requiring extensive, high-quality datasets for training. Such datasets are often unavailable in construction environments, which are characterized by variability, unpredictability, and limited access to annotated data. This scarcity compromises model development and usage, especially when identifying hazards under unique or changing conditions. Another challenge lies in integrating these systems into existing workflows. Construction sites are dynamic environments with established routines, and introducing new technologies without proper alignment can disrupt operations and reduce efficiency in an industry that is already resistant to change. Suboptimal human-machine interfaces (HMI) further exacerbate implementation barriers. Poorly designed HMIs may alienate workers unfamiliar with advanced technologies, leading to resistance or disengagement.

Overall, the state-of-the-art shows that most LLM implementations for robotics either depend on cloud-based methods, introducing latency and reliability issues, or use outdated model-shrinking techniques for edge deployment. Some applications leverage custom hardware, but these solutions often face high costs and limited portability. In the safety context, previous approaches have been hampered by a lack of application-specific training data and poor integration into construction workflows due to unintuitive human–machine interfaces. Building on these findings, our paper addresses these gaps by developing and evaluating an on-device LLM-based robotic framework for construction safety, focusing on appropriate distilled model selection, plug-and-play deployment, and few-shot multimodal processing to overcome latency, data scarcity, and integration barriers.

3. Methodology

The proposed Local Intelligent Safety Assistant (LISA) framework enables more natural human-robot interaction on construction sites by using a multimodal pipeline that interprets and processes natural language commands (*Figure 1*). Its core components are a large language model (LLM), automatic speech recognition (ASR), and text-to-speech (TTS) systems, all of which run locally on the device without relying on cloud-based APIs. This local setup improves privacy, reduces latency, and increases reliability in dynamic construction environments. By combining these components, LISA supports near real-time, bidirectional communication between robots and workers without network dependency. The LLM’s function-calling capability also allows users to control robotic functions through natural language.

This is shown with the camera function in *Figure 1* ('Call robot camera function'), which the system can call when visual input is needed. *Figure 1* illustrates the workflow to process spoken commands, determines if visual input is required, activates the camera function when necessary, and provides audio feedback to the user. This workflow is meant to be generic. Specific tools used are indicated in Section 4.

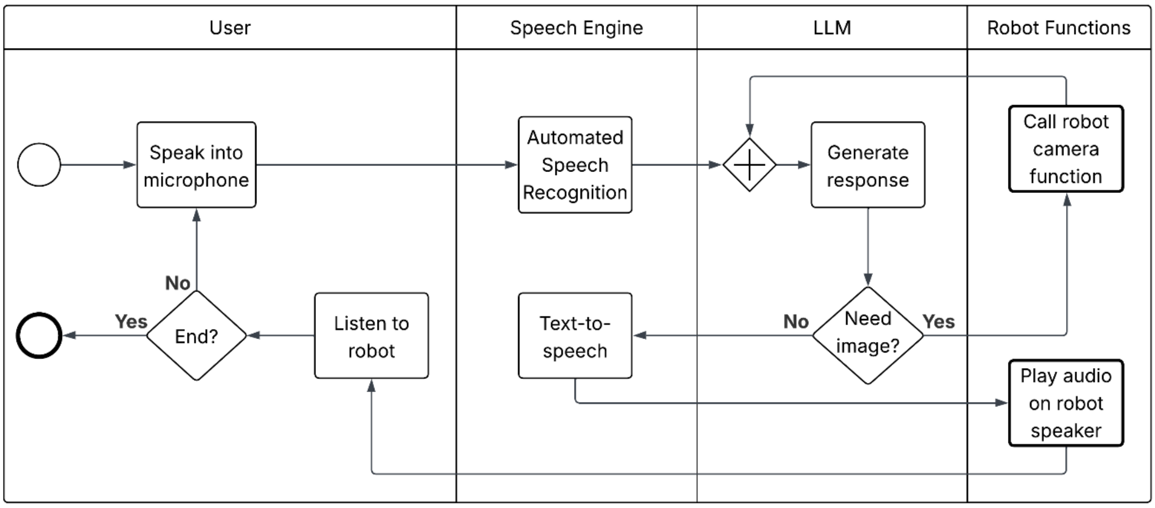

Figure 1: BPMN diagram of the key elements of the LISA framework

3.1. User interaction

User voice input uses a straightforward push-to-talk system: the user presses a remote button to start speaking and presses it again when finished. This approach works well in construction environments, where high background noise can disrupt other activation methods like wake-word detection. The audio recorded between button presses is sent to the Automatic Speech Recognition (ASR) module, which converts speech to text for the large language model (LLM). Once the LLM generates a response, the text is processed by a local text-to-speech (TTS) engine, turning it into audio that plays through the robot's speaker. This method enables natural communication with the robot and requires virtually no user training, unlike more complex methods such as tablet-based input for robot control.

3.2. LLM function calling

The initial system prompt gives the LLM detailed instructions on its purpose, functionality, and the functions it can access. This prompt can be customized for different applications and scenarios. When a user's request is received in text form, the LLM uses these instructions to determine whether visual input is needed to fulfil the request. For instance, the LLM can be configured to activate the camera if the user asks the robot to perform a safety inspection or inquire about the current construction environment. Modern LLMs support structured output, enabling them to reliably generate specific command keywords, such as a "function_call" tag, whenever a function call is required. These keywords are then parsed to execute the corresponding function by the system. When users provide ambiguous commands, the LLM asks for clarification. If the LLM fails to generate the expected "function_call" keyword due to unclear input, the system defaults to conversational mode without triggering function execution. This prevents unnecessary function calls while maintaining natural dialogue flow. This approach can be extended to any high-level robotic function, as long as it is properly defined in the system prompt, enabling more complex capabilities like waypoint navigation or gesture control.

4. Experimentation and Results

We evaluated the LISA framework on a Unitree Go2 robot equipped with an NVIDIA Jetson Orin NX 16GB, focusing on a real-world scenario: on-site personal protective equipment (PPE) compliance. The system was designed to run all inference tasks-language modeling, speech recognition, and speech synthesis, locally.

4.1. Experimental setup

4.1.1. LLM selection

Given the Jetson Orin NX's hardware constraints, we selected the Google Gemma 3:12B model, which is a recent multilingual open-weight model that offers a large 128K-token context window, supports structured outputs, function calling, and vision capabilities, and is designed for efficient deployment on resource-limited devices. This model enabled the use of detailed prompts and allowed the robot to interpret and act on natural language commands as well as process visual inputs, all while fitting within the memory and performance limits of the Jetson Orin NX. The only model parameter that was modified from the default values was the temperature, which was set to 0 for more deterministic output.

4.1.2. Automated Speech Recognition and Text-to-Speech

For automated speech recognition, we implemented OpenAI's Whisper base model due to its strong balance of speed and accuracy. Whisper runs fully offline, providing reliable transcription with minimal latency and maintaining robustness in the noisy conditions typical of construction sites. While it supports multiple languages, only English was tested in our application. For text-to-speech, we used Piper-TTS, which delivered fast and reliable speech synthesis on the Jetson Orin NX. Although Piper's voice quality is not as natural as some cloud-based solutions, it is comparable to widely used TTS systems and was adequate for providing clear, intelligible feedback in the field.

4.1.3. Input and output hardware

Audio input was captured using a wireless microphone system, with the user carrying a transmitter and the Jetson Orin NX receiving the signal via a connected receiver. A wireless USB clicker was used to activate the push-to-talk system, minimizing false activations in noisy environments. Visual input was provided by the Go2's onboard camera, and audio output utilized the robot's built-in speakers. *Figure 2* shows the hardware setup used to test the LISA framework.

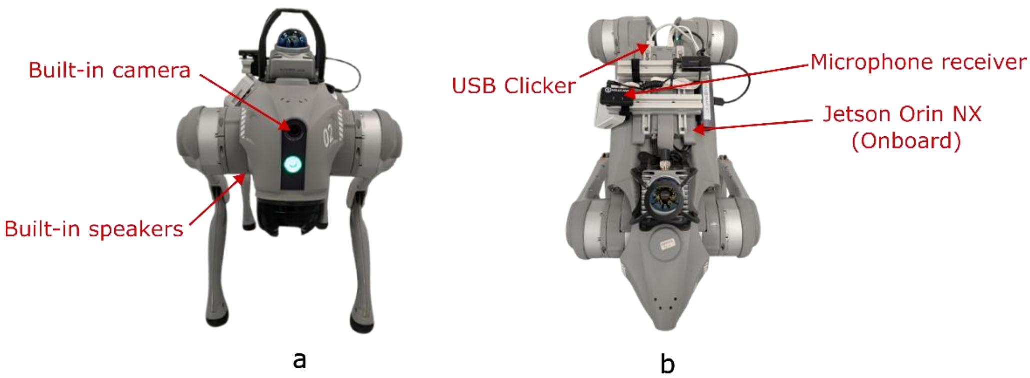

Figure 2: Hardware setup of the Unitree Go2 running the LISA framework; (a) front view; (b) top view with payload

4.2. Application prompt

The application used to test the LISA framework was centered on personal protective equipment (PPE) compliance within a construction environment. In this setup, the LLM was instructed by a system prompt to both answer user questions about appropriate PPE for specific tasks, and to perform safety analyses by evaluating images captured by the robot's camera. Three representative construction activities were included in the prompt: cutting wood with a saw, climbing a ladder, and using aerosol. For each of these tasks, the typical required PPE was clearly defined. The LLM's responsibilities included advising users on which PPE was necessary for each activity, analyzing photos of ongoing work to assess whether the correct PPE was being worn, and notifying users if any required equipment was missing. Readers interested in the prompts used in this study may request them from the authors.

4.3. Results

4.3.1. Human-machine interaction

The first part of the experiment focused on evaluating the quality and speed of the communication between users and LISA. Our aim was to determine whether the system could support natural, near real-time exchanges. We found that for typical queries, such as asking for advice on PPE requirements for specific construction activities, LISA responded within an average of 5 to 10 seconds, in a controlled laboratory environment with minimal background noise. When users posed more general questions outside the scope of the application prompt, such as "what is the average curing time for concrete?", the LLM was still able to provide answers, although response times tended to be slightly longer for these broader or less-defined queries. Overall, the system demonstrated reliable, conversational interaction suitable for a construction environment, as long as the microphone system is robust and not affected by background noise. While direct comparison with cloud-based LLMs is beyond this study's scope, preliminary findings from testing the framework has shown that cloud-based response times are highly network-dependent, with responses going well above 45 seconds in low bandwidth conditions and in several cases, there were issues with connectivity. LISA's local deployment ensures consistent performance regardless of connectivity conditions, which is critical for construction applications.

4.3.2. Safety analysis

The LLM was also instructed by the system prompt to initiate a visual inspection for PPE compliance whenever a user requested a safety analysis (e.g., "*Hi LISA, can you do a safety analysis?*"). Upon receiving such a request, the LLM recognizes the need for visual input and trigger the robot's camera function to capture an image of the ongoing activity. We tested this capability by performing a series of construction tasks in front of the robot and asking LISA to conduct a safety analysis. *Figure 3* shows the test situations, which involved the three target construction activities with varying levels of PPE compliance. We have found that the LLM was reliably performing the function calling to snap a picture with the robot's camera. *Table 1* summarizes the results, with a shortened version of the response provided by LISA. The actual responses were spoken to the worker in longer, more natural sentences (e.g., "*Hi there, you seem to be cutting wood. It looks like you have all the required PPE including gloves, a hard hat and a safety vest. You are fully compliant, good job!*"). The data for the prompt, the full responses and interaction between the worker and the robot can be provided by the authors upon request.

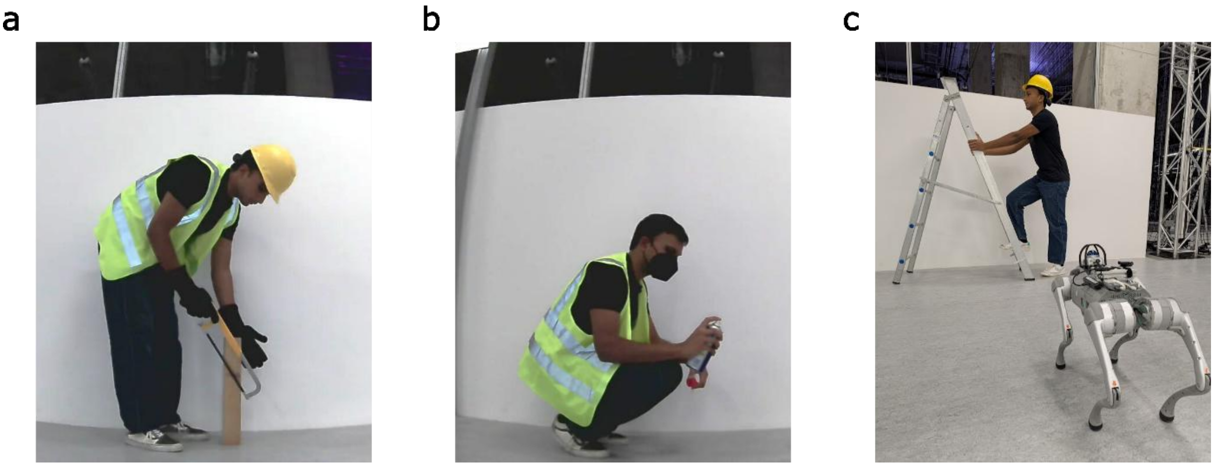

Figure 3: Construction activities with varying PPE compliance: (a) cutting wood and (b) aerosol use, both images taken from the robot's onboard camera, and (c) experimental setup showing LISA observing a subject climbing a ladder without proper PPE.

Table 1: Summary of LISA's simplified responses for safety analysis

Activity	Required PPE	Worn PPE	LISA Response
(a) Cutting wood	hard hat, safety vest, gloves	hard hat, safety vest, gloves	cutting wood with gloves, hard hat, and safety vest; compliant
(b) Aerosol use	hard hat, safety vest, mask	safety vest, mask	aerosol use with safety vest, face mask; non-compliant, missing hard hat
(c) Climbing ladder	hard hat, safety vest, harness	hard hat	climbing ladder with hard hat; non-compliant, missing safety vest, harness

5. Discussion, Conclusion and Future Work

The experiment demonstrates that a locally deployed LLM can effectively enhance human-robot interaction on construction sites, enabling natural communication without relying on network connectivity. The system performed reliably in general conversational tasks, with robust function calling, activity recognition and PPE compliance check. However, the testing was performed in a controlled laboratory setting, which may not fully capture the challenges of active construction sites, and the specific PPE application was relatively simple with only three construction tasks. Although formal accuracy comparison is outside this study's scope, it should be noted that larger cloud-based models are generally less prone to hallucinations and missed detections in the visual analysis tasks. Nonetheless, the results highlight the strong potential of locally executed LLMs to deliver robust, responsive solutions in bandwidth-constrained environments, paving the way for practical human-robot collaboration in construction. Looking ahead, future work will focus on several fronts to address the current limitations and further enhance system capabilities. These include deploying larger LLMs on more powerful edge hardware to improve reasoning and multimodal understanding, as well as fine-tuning the vision components for more accurate and reliable visual recognition. Additionally, retrieval-augmented generation (RAG) could be leveraged for more complex applications. Implementing and testing more advanced function calling, such as autonomous waypoint navigation, will also enable robots to perform increasingly complex tasks. Additionally, since all components of the LISA framework can support multilingual interaction, future development will also emphasize expanding and refining multilingual capabilities to better serve diverse teams on construction sites. These efforts aim to further improve the reliability, versatility, and inclusiveness of fully offline, network-independent human-robot systems for real-world construction applications.

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. Thanks also to the Core Technology Platform (CTP) at NYU Abu Dhabi, particularly the CTP's Kinesis Lab.

References

[1] H. Sadeh, C. Mirarchi, and A. Pavan, "Technological transformation of the construction sector: a conceptual approach," International Journal of Construction Management, vol. 23, no. 10, pp. 1704–1714, Nov. 2021, doi: 10.1080/15623599.2021.2006400.

[2] P. Ko, M. A. Gopee, L. Yu, Y. Garg, M. Mansour, and B. García de Soto, "Development of an instance segmentation model for the automatic identification of recyclable objects in construction sites," *South Korea*, 2022.

[3] Y. Xu, Y. Zhou, P. Sekula, and L. Ding, "Machine learning in construction: From shallow to deep learning," *Developments in the Built Environment*, vol. 6, p. 100045, May 2021, doi: 10.1016/J.DIBE.2021.100045.

[4] R. Azizi, M. Koskinopoulou, and Y. Petillot, "Comparison of Machine Learning Approaches for Robust and Timely Detection of PPE in Construction Sites," *Robotics*, vol. 13, no. 2, Art. no. 2, Feb. 2024, doi: 10.3390/robotics13020031.

[5] S. Park, X. Wang, C. C. Menassa, V. R. Kamat, and J. Y. Chai, "Natural Language Instructions for Intuitive Human Interaction with Robotic Assistants in Field Construction Work," *Automation in Construction*, vol. 161, p. 105345, May 2024, doi: 10.1016/j.autcon.2024.105345.

[6] S. A. Prieto and B. García de Soto, “PACA: A Personal Autonomous Construction Assistant to Facilitate Human-Robot Interaction,” in *Proceedings of the 4th Future of Construction Workshop at the International Conference on Robotics and Automation (ICRA 2025)*, Atlanta, USA, May 2025. [Online]. Available: https://construction-robots.github.io/papers/112.pdf

[7] S. A. Prieto and B. García de Soto, “Large Language Models for Robot Task Allocation,” in *Proceedings of the 3rd Future of Construction Workshop at the International Conference on Robotics and Automation (ICRA 2024)*, May 2024. doi: 10.22260/ICRA2024/0007.

[8] E. Ahmadi and C. Wang, “Transforming Construction Practices with Large Language Models,” presented at the EPiC Series in Built Environment, pp. 414–404. doi: 10.29007/hb2g.

[9] Z. Yu *et al.*, “Cambricon-LLM: A Chiplet-Based Hybrid Architecture for On-Device Inference of 70B LLM,” Sep. 24, 2024, *arXiv*: arXiv:2409.15654. doi: 10.48550/arXiv.2409.15654.

[10] P. Sikorski *et al.*, “Deployment of Large Language Models to Control Mobile Robots at the Edge,” Oct. 10, 2024, *arXiv*: arXiv:2405.17670. doi: 10.48550/arXiv.2405.17670.

[11] S. Tanabe, K. Sawai, and T. Suzuki, “Sensor Node Deployment Strategy for Maintaining Wireless Sensor Network Communication Connectivity,” *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 2, no. 12, Art. no. 12, Jan. 2012, doi: 10.14569/IJACSA.2011.021221.

[12] S. Lee *et al.*, “Quantification of Large Language Model Distillation,” Feb. 17, 2025, *arXiv*: arXiv:2501.12619. doi: 10.48550/arXiv.2501.12619.

[13] F. Zeng, W. Gan, Y. Wang, N. Liu, and P. S. Yu, “Large Language Models for Robotics: A Survey,” Nov. 13, 2023, *arXiv*: arXiv:2311.07226. doi: 10.48550/arXiv.2311.07226.

[14] Y. Kim, D. Kim, J. Choi, J. Park, N. Oh, and D. Park, “A Survey on Integration of Large Language Models with Intelligent Robots,” May 20, 2024, *arXiv*: arXiv:2404.09228. doi: 10.48550/arXiv.2404.09228.

[15] K. Kawaharazuka, Matsushima ,Tatsuya, Gambardella ,Andrew, Guo ,Jiaxian, Paxton ,Chris, and A. and Zeng, “Real-world robot applications of foundation models: a review,” *Advanced Robotics*, vol. 38, no. 18, pp. 1232–1254, Sep. 2024, doi: 10.1080/01691864.2024.2408593.

[16] Y. Zheng, Y. Chen, B. Qian, X. Shi, Y. Shu, and J. Chen, “A Review on Edge Large Language Models: Design, Execution, and Applications,” Feb. 23, 2025, *arXiv*: arXiv:2410.11845. doi: 10.48550/arXiv.2410.11845.

[17] O. Friha, M. Amine Ferrag, B. Kantarci, B. Cakmak, A. Ozgun, and N. Ghoualmi-Zine, “LLM-Based Edge Intelligence: A Comprehensive Survey on Architectures, Applications, Security and Trustworthiness,” *IEEE Open Journal of the Communications Society*, vol. 5, pp. 5799–5856, 2024, doi: 10.1109/OJCOMS.2024.3456549.

[18] J. Kong, L. Hu, F. Ponzina, and T. Rosing, “TinyAgent: Quantization-aware Model Compression and Adaptation for On-device LLM Agent Deployment,” presented at the Workshop on Efficient Systems for Foundation Models II @ ICML2024, Jul. 2024. Accessed: Apr. 14, 2025. [Online]. Available: https://openreview.net/forum?id=ntI0eq0urB

[19] V. Bharathi, S. A. Prieto, B. García de Soto, and J. Teizer, “Automating Construction Safety Inspections using Robots and Unsupervised Deep Domain Adaptation by Backpropagation,” *Proceedings of the 41st International Symposium on Automation and Robotics in Construction (ISARC 2024)*, pp. 855–862, Jun. 2024, doi: 10.22260/ISARC2024/0111.

[20] F. Dornik and N. Bartels, “Use of robotics to coordinate health and safety on the construction site,” in *Proceedings of the 41st International Symposium on Automation and Robotics in Construction*, IAARC, Jun. 2024, pp. 537–544. doi: 10.22260/ISARC2024/0070.

[21] M.-F. R. Lee and T.-W. Chien, “Intelligent Robot for Worker Safety Surveillance: Deep Learning Perception and Visual Navigation,” in *2020 International Conference on Advanced Robotics and Intelligent Systems (ARIS)*, Aug. 2020, pp. 1–6. doi: 10.1109/ARIS50834.2020.9205772.