Human Factors in Design, Engineering, and Computing, Vol. 159, 2024, 1819–1828
https://doi.org/10.54941/ahfe1005749

Enhancing Communication Transparency and Teaming in Human-Autonomy Systems: Integrating Large Language Models for On-Site Construction Operations

Di Liu and Rui Liu

University of Florida, Gainesville, FL 32608, USA

ABSTRACT

The integration of autonomous systems in the construction industry is rapidly advancing, driven by the need to enhance efficiency, safety, and productivity on-site. With autonomous agents evolving from mere tools to collaborative teammates, there is a critical need to understand the dynamics of human-autonomy teaming (HAT) in complex construction environments. This research addresses the gap in knowledge surrounding the communication and collaboration between human workers and autonomous agents, specifically within the context of tower crane operations, where effective teamwork is crucial for safety and task success. Despite the growing implementation of autonomous systems, research on human-autonomy teaming in construction remains limited, particularly regarding the interdependencies of key attributes like team composition, communication transparency, interaction dynamics, trust, and overall team performance. Current studies on autonomous crane operations predominantly focus on the technical aspects, such as smart sensing, perception, and motion control, with little attention given to the team-based human factors involved in these operations. The lack of comprehensive studies on how various human psychological and behavioral factors interact with autonomous systems during real-time operations presents a significant challenge, which this research aims to address. This paper proposes and evaluates a novel VR-based system architecture designed to enhance understanding of the impact of communication transparency on human-autonomy teamwork in construction settings. In addition, the integration of a Large Language Model (LLM)-enabled autonomous agent adds a critical layer to the HAT dynamic by facilitating real-time communication and decision-making. The LLM processes inputs such as sensor data, crane motion sequences, and operator commands to provide actionable insights and hazard warnings, further improving transparency and reducing the cognitive load on the operator. The agent's ability to generate predictive models also assists with obstacle detection and motion planning, enhancing the overall safety and efficiency of crane operations. Organized into a three-layered HAT framework—(1) Team Goals (Bottom Layer), (2) Team Synchronization (Middle Layer), and (3) Team Outcome (Top Layer)—the VR environment simulates a construction job site where a semi-autonomous tower crane operates alongside a human worker. The system is designed to assess how varying levels of communication transparency, based on the situational awareness transparency model, affect the cognitive states, trust levels, and task performance of human operators during dynamic role allocation. This research contributes to both academic and industrial communities by offering a foundational architecture for studying human-autonomy teaming in construction, while the VR-based system serves as a powerful tool for future experimental research. For industry practitioners, the study provides insights into designing more effective and transparent communication systems that enhance safety and efficiency in autonomous crane operations, ultimately contributing to safer and more reliable construction practices.

Keywords: Human-autonomy teaming, Tower crane, Transparency, On-site construction, Large-language model

INTRODUCTION

In recent years, cutting-edge technologies such as automation, teleoperation, and robotics have revolutionized construction on-site operations. In particular, there are increasing studies on tower crane teleoperation(TCT) due to the benefit of human-in-the-loop (HITL), which allows autonomous systems to augment human capability without replacing human workers (Sitompul, 2022). TCT is a safety-critical process and the operator should strictly follow federal and state guidelines (*1910.179|OSHA*, 2016). Traditional crane operation training includes classroom training, simulator-based training, hands-on onsite operations with guidance from experienced operators, and certifications and testing (*NCCCO*, 2023). Specifically, virtual crane operation allows trainees to use crane operation simulators to mimic real-world conditions and to practice in various scenarios without risk and high costs. Moreover, virtual crane operation prepares next-generation crane operators on how to teleoperate tower crane systems.

Traditionally, a teleoperation system allows a human operator to remotely control a construction machine. Such a system includes a construction robot (e.g., an autonomous or semi-autonomous tower crane), a human operator, and the control interface. Research on TCT includes Human-Machine Interfaces (HMIs), such as augmented reality (AR) displays, and multi-sensory feedback systems, aiming to enhance operator perception and control (Wang *et al.*, 2021). However, several challenges in the TCT system cannot be overlooked. The first challenge is the limited situational awareness and excessive cognitive load of the operator (Sitompul, 2022). Unlike on-site operation where operators can use multiple sensory inputs and intuitively sense the task environment, teleoperation relies heavily on cameras, sensors, and other visualization tools, thus blind spots or delays in responding to unexpected situational changes occur (Si and Niu, 2024). The second challenge is the limitation of human-machine interface design since the majority are still using traditional control interfaces such as joysticks (Yu *et al.*, 2021). The third challenge is delayed communication, which may hinder precise operation in a complex job site. The last challenge is caused by the growing gap between the skills required by traditional crane operation and current teleoperation.

Studies on Human-Autonomy Teaming (HAT) and the implementation of prompt-based communication methods, e.g., large-language models (LLMs) shed light on addressing these challenges, especially in reducing the cognitive load and improving intuitive interface design. At a theoretical level, the development of HAT can provide a systematic understanding of human-autonomy interaction, cognition, team coordination, design, and evaluation for leveraging integrative performance. In a human-autonomy team, construction workers typically assume roles where they oversee autonomous systems, either directly through the HITL interface or indirectly through decision-making. Real-time communication, trust, and communication transparency can ensure role clarity so that humans can intervene effectively when needed. A well-defined interaction model should outline when and how human inputs are required and how the system communicates its state, intent, or alerts back to the human operator. Additionally, using prompt-based communication, such as LLMs to interpret and streamline sensor data can help reduce cognitive load and improve decision-making (Chen *et al.*, 2024). Incorporating LLMs into the operation system can enhance communication by generating intuitive interfaces that provide clear, context-aware instructions and explanations. For example, in construction crane operations, LLMs could help translate real-time sensor data into meaningful insights for operators, promoting trust and reducing mental workload by offering explainable features. Thus, it has the potential to bridge the gap between human cognition and machine autonomy, reduce the cognitive burden, and enhance team transparency and performance.

This paper aims to leverage the HAT theory and implementation of LLMs to advance TCT which can well prepare next-generation crane operators. Specifically, two research questions (RQ) are addressed:

- **RQ1**: where humans should fit in the dynamic of human-smart-crane teaming so that benefits TCT?
- **RQ2:** how does a prompt-based communication method, such as LLMs, promote the intuitiveness of communication, and further leverage trust, transparency, and team performance of HAT in TCT?

BACKGROUND

Human-Autonomy Teaming in Construction

O'Neill et al. (2022) defined Human-autonomy teaming (HAT) as "HAT can be defined as interdependence in activity and outcomes involving one or more humans and one or more autonomous agents, wherein each human and autonomous agent is recognized as a unique team member occupying a distinct role on the team, and in which the members strive to achieve a common goal as a collective. The "autonomy" aspect of human–autonomy teaming refers to the autonomous agent" (O'Neill *et al.*, 2022). Based on the classification of Level of Autonomy, HAT as an advanced stage of human automation interaction, has been widely accepted and implemented in areas of defense, aviation, autonomous driving, and manufacturing. In the construction domain, HAT theoretical development is still in the infant stage. Valuable research efforts related to HAT are made on developing the taxonomy for human-robot collaboration in construction (Liang Ci-Jyun *et al.*, 2021; Rodrigues *et al.*, 2023), and the framework of human-digital twin collaboration (Agrawal *et al.*, 2023). Nevertheless, there is a strong demand for HAT theoretical development in construction, especially focusing on the interdependence between key HAT variables such as communication, cognition, role allocation, transparency, and trust at the team level.

HAT in construction in this paper is defined as a collaborative concept in which human workers and autonomous systems (including construction robots, AI-driven tools, and other automated agents) cooperate to achieve construction goals.

LLMs Implementation in Construction

LLMs based on zero-shot or few-shot learning can understand the context and task with very few or no examples. Thus, LLMs such as ChatGPT, have demonstrated remarkable capabilities in understanding and generating human-like text, which relies on translating heterogeneous data, generating reasonable suggestions, and following instructions in natural language. Such an interaction is similar to the decision process of human beings and makes communication much more intuitive. Efforts were made to adopt LLMs in interactions between human users and autonomous systems such as robots or BIM to enhance task planning and user experience. LLMs can take multiple roles to support human users such as assisting (analyze, optimize, inform), coworking (recommend, suggest), or supervising (educate) (Rane, 2023). Kim et al. developed a novel framework for LLM-based construction task planning by translating heterogeneous input data (e.g., BIM data, robot task specifications, and instructions from human experts) and generating actionable task plans for robots via conversational robot-LLM communication (Kim *et al.*, 2024). BIMS-GPT, a prompt-based virtual assistant framework is developed to enhance information search in BIM, and it allows users to use natural language to retrieve building information in the BIM platform without deep knowledge of BIM data structure or graphical user interfaces (GUIs) (Zheng and Fischer, 2023). Chen et al. developed an AR-based prompt-tuning interactive interface that allows users to intuitively interact with and enhance situational awareness(Chen *et al.*, 2024).

Overall, since LLMs are based on in-context learning, which is learning from analogy, it can communicate explicitly with autonomous systems via machine language and implicitly with human users via natural language. LLMs have great potential in various construction directions such as project management, scheduling, cost estimating, data fusion, and automation. Furthermore, LLMs as an interpretable interface have the potential to advance the communication, cognition, transparency, and trust of HAT in teleoperation without additional hardcoded engineering efforts.

RESEARCH METHOD

To address RQ1, this paper first proposes a HAT interaction model for tower crane teleoperation to outline how the human operator fits into HAT dynamic for TCT. Moreover, to address RQ2, an LLMs-enabled autonomous agent is designed to serve as an unmanned team member collaborating with the human operator in a human-autonomy team for virtual crane operation. Lastly, a VR-based virtual TCT is designed for the evaluation which will be conducted in future case studies.

HAT Interaction Model in Tower Crane Teleoperation

The proposed HAT interaction model is composed of three layers: Layer01_Team Goals (Bottom Layer), Layer02_Team Synchronization (Middle Layer), and Layer03_Team Outcome (Top Layer) shown in **Figure 1**.

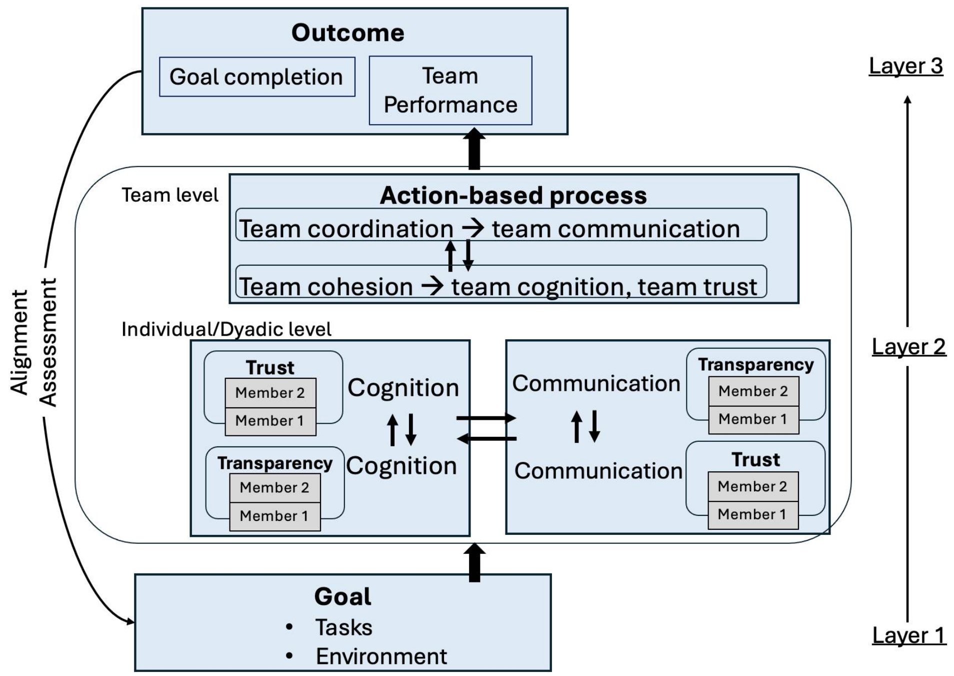

Figure 1: Bi-directional HAT Interaction Framework.

At layer 1, the team goal includes tasks and environmental factors. For example, managing environmental uncertainties such as obstacles and potential hazards is part of the goal.

At layer 2, it is noted that there are two types of team compositions, namely, dyadic level (between two team members), and team level (more than two team members). Communication is defined as multi-sensory information exchanges to enable knowledge updates for each party for goal achievement including task completion. Communication may occur in verbal, visual, or haptic formats. Communication is goal-driven. Cognition is defined as a background or inner process out of communication. Cognition constructs include intent, mental workload, attention, and situational awareness. Note that situational awareness is a cognition construct that involves a subjective understanding of current status and future prediction. Cognition may or may not be goal-driven. The influence of cognition and communication is bi-directional at the dyadic and team levels. Trust and transparency are relational constructs and occur at the dyadic or team level. Transparency is a quality of an interface to support the comprehension of an agent's intent, future plans, and the reasoning process (Chen, 2023). Interface transparency allows cognition constructs (e.g., intent) to be interpretable and facilitates communication and other cognition constructs (e.g., situational awareness). Trust is a goal-driven attitude toward other team members at the dyadic or team level, especially in exposure to uncertainty or vulnerability (Lopez *et al.*, 2023). The proper amount of trust enhances the efficiency of communication and optimizes cognition. Trust is associated with reliability, affection, and emotions. Trust and transparency occur at the dyadic and team levels (e.g., between two individuals) and affect cognition at the individual level and communication at the dyadic level. Trust and transparency are goal-driven. Trust is time-dependent, dynamic, and contextual (Li, Kamaraj and Lee, 2023), while transparency is more static and less contextual. At the team level, team coordination represents team communication and their actionable results. Team cohesion is a representation of team cognition including shared mental model and team trust. Team cohesion reflects how smoothly the different types of relations progress over time and if it positively affects team coordination.

At layer 3, outcomes include goal achievement, which should be assessed by alignment with the goal at layer 1. Outcomes also include team performance such as the assessment of HAT variables at level 2, and how they affect goal achievement. The alignment between layers 1 and 3 formulates the feedback loop.

Note that synchronization occurs between layer 1 and layer 2, meaning that team goals drive cognition, communication, and action-based processes. Overall, the bi-directional HAT framework is contextual and changed over time. The correct comprehension of the current status and prediction for the future status could greatly affect goal achievement.

Failure of cross-layer synchronization may lead to inefficiency or task incompletion. Other vulnerabilities of the process within layer 2 include unreliable noise from communication channels, insufficient transparency, improper amount of trust, or trust dynamic changing over time. Notably, both the Synchronization (layer2) and Goal (layer1) are dynamic and require efficient communication protocols and real-time decision-support systems.

Design an LLMs-Enabled Autonomous Agent to Facilitate the HAT Dynamic in TCT

The LLM-based agent is designed to facilitate the HAT dynamic in several aspects. First, at Layer 1 (team goals), an LLM-based agent can help assess environmental conditions and refine team goals. It is designed to translate sensor data (e.g., weather changes, obstacle detection) into actionable insights, ensuring that human operators understand how environmental factors might affect crane operation. Second, at Layer 2 (team synchronization), the LLM-based agent assists with communication, transparency, and trust. As for communication, it provides real-time feedback, alerting human operators to potential issues while offering recommendations for adjustments. It achieves transparency by explaining the autonomous system's decisions in real-time to assist the operator's comprehension of the current status. As for trust, the agent can track, calibrate, and inform the level of trust over time based on the integrative analysis of historical interactions and operational outcomes. The agent generates predictive models, offering insights into potential future risks based on current crane operations and environmental factors, affecting the next-step crane motions. At Layer 3 (team outcome), by using an LLM-based agent, performance metrics such as decision-making speed, the accuracy of commands, and operator satisfaction with the system's transparency can be recorded. The LLM can log decision-making processes and generate trust trajectories, enabling a review of team performance to further optimize the HAT interaction.

Table 1 shows the integration of the LLMs-enabled agent into the HAT model at each layer of the HAT dynamic to facilitate TCT. Heterogenous input data is collected and mapped into a data manager that connects to an external GPT server, and Prompt communication between the human operator (Q) and the agent (A).

Table 1. Integration of LLMs-enabled agent in HAT interaction model to facilitate TCT.

HAT Level	Input Data for LLM	LLM Prompt Template(Q&A)
Layer 1: Team Goals	- sensors: Real-time environmental data (obstacle locations, weather) - human: task descriptions (e.g., "Move cargo to location X") - crane system: pre-programmed motion sequences - Target locations	Q: "Translate sensor data to task plan. Identify potential hazards." A: Optimized route generated for crane operation. [List of potential hazards]
Layer 2: Team Synchronization	- From crane: real-time crane status (position, load) - From human: manual commands (e.g., switch to autonomous mode) - From LLM: recommended next steps for tasks and obstacle warnings - Target updates	Q: "What are the next steps for transporting cargo to location X?" A: "Move crane arm to 45-degree angle for optimal loading"
Layer 3: Team Outcomes	- From human: feedback on task success or failure - From sensors: confirmation of task completion (e.g., cargo delivered) - From crane: task status (e.g., cargo successfully dropped)	Q: "Was the task successfully completed?" A: "Completion status with key performance metrics - time taken to complete, deviations from optimal path. Trust updates based on human performance feedback"

VR-Based Intelligence TCT System and User Feedback

To evaluate the proposed HAT model, the human and agent dynamic affecting TCT, a VR-based intelligence TCT system is designed in Unity3d. Virtual Reality (VR) is widely recognized as an effective simulation tool for replicating close-to real environment to allow users to have a realistic and immersive experience and ensure safety at the same time (Ali *et al.*, 2024). The key feature of the proposed VR-based system is that a human operator team with an LLMs-enabled unmanned agent in the tower crane operation. The unmanned agent is connected with the virtual sensors to collect real-time operational data, including cargo status, motion status, and site environment, to analyze these task and environment data and to communicate with crane operators via a virtual conversational chat box connecting to an external GPT server. Through the natural language interaction, the agent can provide recommendations for the next step of action to facilitate human decision-making, especially in terms of motion planning and collision avoidance (Figure 2). Real-time data, such as crane and cargo positions, are recorded and displayed as the rationale to support the recommendation.

One user experienced the preliminary design of the operation system and provided valuable feedback. The user felt that the system was intuitive and natural to use, and the interaction was straightforward with clear visual cues. With the LLMs-enabled conversational interface, the natural language interaction allowed the user to have less stress and easier navigating during the operation. Meanwhile, functional improvements are needed such as reducing the glitches and enhancing the system stability.

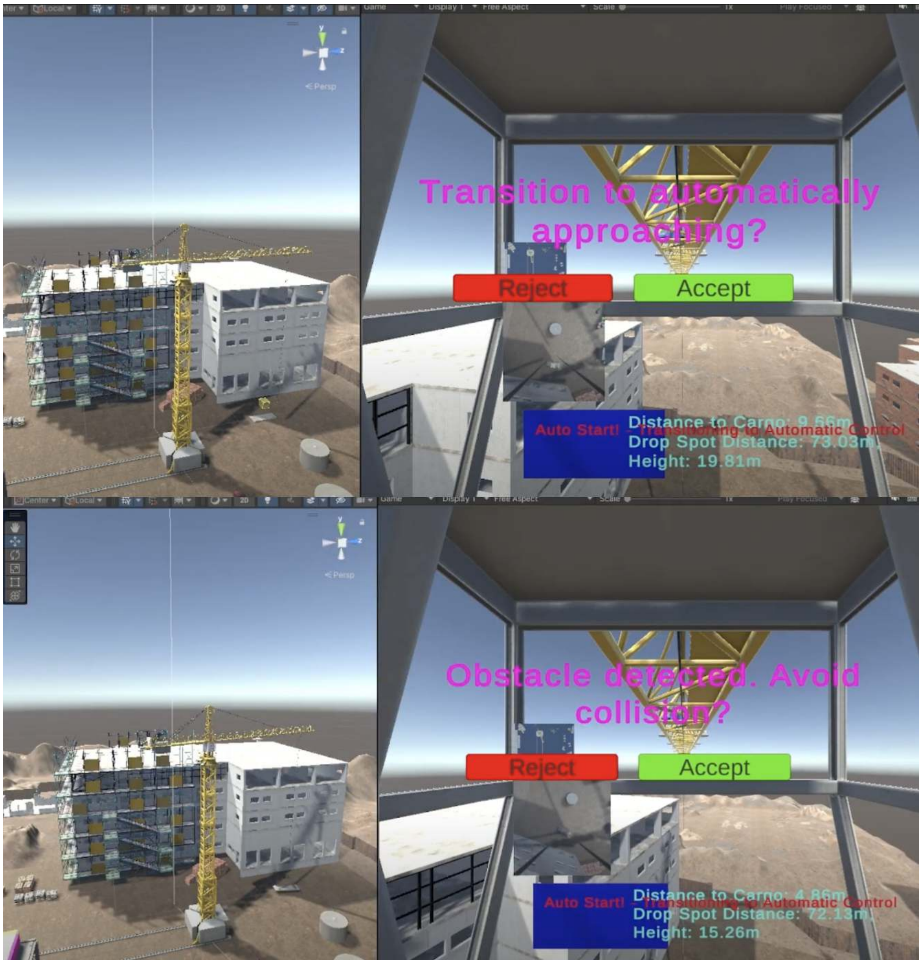

Figure 2: Preliminary design of the VR-based Intelligence TCT system.

NEXT STEP WORK

The next steps in this research will involve both empirical testing through human-subject experiment and technical improvements to the system. The human-subject experiment will focus on evaluating task performance, trust, transparency, and mental workload, which are key factors of the HAT interaction model, and their interdependence. Concurrently, the system will undergo several iterations of evaluation and improvement, including stability enhancements to minimize glitches during operation and refinement of the control mechanisms for more precise and seamless crane movement.

CONCLUSION

This study has proposed a comprehensive framework for enhancing Human-Autonomy Teaming (HAT) in tower crane teleoperation (TCT) through the integration of Large Language Models (LLMs). By addressing the key challenges of teleoperation, including limited situational awareness, high cognitive load, and delayed communication, we have developed a bi-directional HAT interaction model that incorporates a layered approach to team goals, synchronization, and outcomes. The model was further enhanced with an LLM-enabled autonomous agent that supports real-time decision-making, communication transparency, and trust calibration between the human operator and the autonomous system. The testing of the VR prototype based on the proposed framework is ongoing.

This research contributes to both academic and industrial communities by providing a novel framework that integrates LLMs into construction teleoperation systems. First, the study advances the theoretical understanding of HAT in construction by offering a model that captures the interdependence between communication, cognition, trust, and transparency. Second, the VR-based teleoperation system designed in this study can serve as a useful tool for training next-generation crane operators, improving operational efficiency, safety, and task accuracy. Future empirical tests and system refinements will further solidify the practical applicability of the proposed framework.

ACKNOWLEDGMENT

The authors received no financial support for the research of this article.

REFERENCES

1910.179 | OSHA (2016) *Occupational Safety and Health Standards - Overhead and gantry cranes*. Available at: https://www.osha.gov/laws-regs/interlinking/standards/1910.179 (Accessed: 16 September 2024).

Agrawal, A. *et al.* (2023) 'Digital Twin: Where do humans fit in?', *Automation in Construction*, 148, p. 104749. Available at: https://doi.org/10.1016/j.autcon.2023.104749.

Ali, A. H. *et al.* (2024) 'Tower crane safety technologies: A synthesis of academic research and industry insights', *Automation in Construction*, 163, p. 105429. Available at: https://doi.org/10.1016/j.autcon.2024.105429.

Chen, H. *et al.* (2024) 'Augmented reality, deep learning and vision-language query system for construction worker safety', *Automation in Construction*, 157, p. 105158. Available at: https://doi.org/10.1016/j.autcon.2023.105158.

Chen, J. Y. C. (2023) 'Transparent Communications for Human–Machine Teaming', in *Handbook of Human-Machine Systems*. *Handbook of Human-Machine Systems*, IEEE, pp. 105–114. Available at: https://doi.org/10.1002/9781119863663.ch10.

Kim, K. *et al.* (2024) 'Context-Aware and Adaptive Task Planning for Autonomous Construction Robots Through Llm-Robot Communication'. Rochester, NY. Available at: https://doi.org/10.2139/ssrn.4827728.

Li, M., Kamaraj, A. V. and Lee, J. D. (2023) 'Modeling Trust Dimensions and Dynamics in Human-Agent Conversation: A Trajectory Epistemic Network Analysis Approach', *International Journal of Human–Computer Interaction*, 0(0), pp. 1–12. Available at: https://doi.org/10.1080/10447318.2023.2201555.

Liang Ci-Jyun *et al.* (2021) 'Human–Robot Collaboration in Construction: Classification and Research Trends', *Journal of Construction Engineering and Management*, 147(10), p. 03121006. Available at: https://doi.org/10.1061/(ASCE) CO.1943–7862.0002154.

Lopez, J. *et al.* (2023) 'The complex relationship of AI ethics and trust in human–AI teaming: insights from advanced real-world subject matter experts', *AI and Ethics* [Preprint]. Available at: https://doi.org/10.1007/s43681-023-00303-7.

NCCCO (2023) *National Commission for the Certification of Crane Operators*. Available at: https://www.nccco.org/nccco/certifications/tower-crane-operator/certification-overview.

O'Neill, T. *et al.* (2022) 'Human–Autonomy Teaming: A Review and Analysis of the Empirical Literature', *Human Factors*, 64(5), pp. 904–938. Available at: https://doi.org/10.1177/0018720820960865.

Rane, N. (2023) 'Role of ChatGPT and Similar Generative Artificial Intelligence (AI) in Construction Industry'. Rochester, NY. Available at: https://doi.org/10.2139/ssrn.4598258.

Rodrigues, P. B. *et al.* (2023) 'A multidimensional taxonomy for human-robot interaction in construction', *Automation in Construction*, 150, p. 104845. Available at: https://doi.org/10.1016/j.autcon.2023.104845.

Si, W. and Niu, L. (2024) 'Enhancing Human Reliability Prediction in Smart Tower Crane Interfaces: A Refined Approach Using Simplified Plant Analysis Risk–Human Reliability Assessment and the Decision Making Trial and Evaluation Laboratory–Analytic Network Process', *Buildings*, 14(4), p. 1083. Available at: https://doi.org/10.3390/buildings14041083.

Sitompul, T. A. (2022) 'Human–Machine Interface for Remote Crane Operation: A Review', *Multimodal Technologies and Interaction*, 6(6), p. 45. Available at: https://doi.org/10.3390/mti6060045.

Wang, R. D. *et al.* (2021) 'A system boundary-based critical review on crane selection in building construction', *Automation in Construction*, 123, p. 103520. Available at: https://doi.org/10.1016/j.autcon.2020.103520.

Yu, Z. *et al.* (2021) 'Approaches for Motion Control Interface and Tele-Operated Overhead Crane Handling Tasks', *Processes*, 9(12), p. 2148. Available at: https://doi.org/10.3390/pr9122148.

Zheng, J. and Fischer, M. (2023) 'Dynamic prompt-based virtual assistant framework for BIM information search', *Automation in Construction*, 155, p. 105067. Available at: https://doi.org/10.1016/j.autcon.2023.105067.