

Tan TAN, Xin LIU, Alexander N. WALZER, Ming Shan NG, Daniel M. HALL

Human–robot interaction design with augmented reality in construction: A systematic review of frameworks, applications, and future directions

© The Author(s) 2026. This article is published with open access at link.springer.com and journal.hep.com.cn

Abstract Research in human–robot Interaction (HRI) has increasingly demonstrated how Augmented Reality (AR) enables better interactions between humans and robots. However, the design of HRI remains less understood. Through a systematic literature review of 53 related papers, this research provides an overview of the emerging applications and trends for AR and identifies three types of AR interfaces as follows: 1) remote modular interface, 2) proximal modular interface, and 3) proximal integral interface. The review indicates potential future directions of construction-oriented and human-centric interaction design studies, leading to four pairs of subsystems, which are frequently modularised or integrated, and three conceptual frameworks for HRI interfaces are proposed. Moreover, this research contributes to the theoretical exploration of interaction design. Future applications can adapt to various tasks by using the proposed three conceptual frameworks for interfaces, as well as combining the four proposed subsystem pairs to suit specific task requirements

in the construction sector.

Keywords augmented reality, human–robot interaction, interaction design, human-computer interaction, literature review

1 Introduction

Continued innovation in robotics is expected to drive digital transformation into the growth phase, leading to wider adoption (Bock, 2015). The construction industry has one of the lowest levels of digitalisation among all other industries (Agarwal et al., 2016). The adoption of robotic technologies is still in its early stages and faces numerous challenges due to their complex operations and diverse applications (Graser et al., 2021; Tan et al., 2023; Walzer et al., 2025). The construction industry is a key driver of job creation and economic development for many governments due to its labor-intensive nature (Chiang et al., 2015). Achieving full automation in the industry remains a distant goal in the sector due to challenges in both technical development and management for automation in construction. Recent research has proposed hybrid collaboration between human labor and machinery power to different extents (Wang et al., 2023; Xiang et al., 2021). This could serve as a transition in preparation for a robotic transformation to equip job skills and technical development in preparation for full automation implementation in the sector. However, this hybrid collaboration between humans and robots requires an understanding of both the emerging technical development and, most importantly, the corresponding human–robot Interaction (HRI) interfaces that suit the current state-of-the-art robotic implementation and job skills in the sector.

human–robot Collaboration (HRC) is a core subset of the broader HRI paradigm, which has received academic attention in the fields of industrial and service robotics

Received Apr. 17, 2025; revised Jul. 1, 2025; accepted Jul. 18, 2025

Tan TAN (✉), Xin LIU

Department of Real Estate and Construction, The University of Hong Kong, Hong Kong SAR, China; National Center of Technology Innovation for Digital Construction Hong Kong Branch, The University of Hong Kong, Hong Kong SAR 999077, China
E-mail: tant@hku.hk

Alexander N. WALZER

School of Architecture, Design and Civil Engineering, Zurich University of Applied Sciences, 8400 Winterthur, Switzerland; Department of Civil, Environmental and Geomatic Engineering, ETH Zurich, Zurich 8093, Switzerland

Ming Shan NG

Center for the Possible Futures, Kyoto Institute of Technology, Kyoto 606-0951, Japan

Daniel M. HALL

Department of Management in the Built Environment, Delft University of Technology, Delft 2628 BL, Netherlands

This research was supported by Teaching Development Grant from the University of Hong Kong, China.

(Galín and Meshcheryakov, 2020). HRC refers to a series of collaborative behaviors between humans and robots (Ajoudani et al., 2018). Recent reviews of HRC in construction mainly emphasized the potential of robotic systems to enhance productivity and safety through collaborative operation (Bloss, 2016; Boschetti et al., 2022; Pérez et al., 2020). However, although these studies emphasized the importance of integrating robots into construction workflows, they typically only focused on system-level collaboration strategies without detailing the mechanisms for implementing real-time interaction. HRI specifically targets these mechanisms, such as interaction methods and communication protocols (Frijns et al., 2023; Lunghi et al., 2019). Therefore, reviewing HRI in depth can help translate high-level collaboration concepts into practical interaction strategies for on-site construction applications.

In the construction industry, the deployment of robots is still limited, mainly focused on simple or structured tasks such as automatic painting or unmanned bulldozing, and the role of HRI has become particularly important (Al Masri et al., 2024; Khan et al., 2025; Tan et al., 2024). Most real-world construction environments are unstructured, dynamic, and labor-intensive, making it difficult to achieve complete robot autonomy in the short-term. HRI plays a crucial role in achieving effective supervision, guidance, and collaboration between human operators and semi-autonomous machines. A well-designed HRI interface can ensure immediate ease of use and safety and lay the foundation for expanding task complexity, increasing operational flexibility and user acceptance of robots (Adamides et al., 2014; Lunghi et al., 2019). It is important that the design and functionality of the HRI interface be customized for specific robot application scenarios (Krupke et al., 2018). For example, in autonomous systems such as unmanned bulldozers, HRI mainly involves supervisory control and safety coverage to support remote monitoring (Khan et al., 2025). In contrast, task-oriented robots, such as robot spray painting arms or assembly units, require more interactive and context-aware interfaces that can utilize gesture recognition, voice commands, and spatial input coordination (Brosque et al., 2020).

Existing review studies primarily examined the technical aspect of the interaction level of HRI from three dimensions: (i) spatial dimension: the location where robots and humans perform tasks, (ii) temporal dimension: the simultaneity of task performance by robots and humans and (iii) direction: the direction of commands between humans and robots. These dimensions provide a general overview of interaction design. Among this existing research, it is clear that the design for HRI interfaces is essential to enable successful and efficient interactions between humans and robots. Recent construction-specific studies have contributed to this taxonomy, yet exhibited

notable gaps. For example, Han et al. (2021) focused on co-located collaborative construction in architectural contexts, offering insight into spatial proximity but without a systematic analysis of how spatial arrangements influence interface requirements or cognitive workload. Fu et al. (2024) and Zhang et al. (2023) explored off-site manufacturing and on-site assembly, respectively, yet these works tend to emphasize process integration rather than the interaction modalities and their ergonomic or task-performance implications. In terms of temporal interaction, Liang et al. (2021) proposed five levels of human–robot collaboration, ranging from programming to full autonomy, yet stopped short of mapping these levels to specific interaction design needs or user capabilities in construction environments. Similarly, while directionality was addressed by Wei et al. (2023) through classifications based on robot intelligence (e.g., passive, reactive, proactive systems), such distinctions remain abstract and are seldom grounded in empirical interface performance data. Burden et al. (2022) highlighted implementation barriers in collaborative robotics but offered limited discussion on how interaction directionality constrains or enables effective task handover and shared control in dynamic site conditions. Overall, while existing studies provided valuable classification schemes and foundational insights into the dimensions of HRI, most remained conceptual or descriptive in nature, lacking critical interrogation of their applicability in complex, real-world construction settings. Few studies empirically assessed how interaction dimensions affect user performance, safety, or task efficiency under varying site constraints. Moreover, the integration of human-centered design principles such as cognitive load, adaptability to diverse user roles, and context-specific usability remains insufficiently addressed. This highlights a significant research gap between theoretical frameworks and practical implementation. Our review aims to bridge this gap by critically analyzing the usability and design implications of HRI systems in construction, offering a more application-oriented perspective.

Additionally, limited research has systematically explored and developed HRI interface designs that can be generalized and applied by both researchers and industry practitioners to fully realize the collaborative potential between humans and robots. Among all, visualization technology, including immersive technologies, could enhance the Architectural, Engineering, and Construction (AEC) process by improving communication and collaboration among stakeholders (Bouchlaghem et al., 2005), and between humans and robotics (Suzuki et al., 2022). Among various immersive visualization technologies, AR can facilitate construction support, progress monitoring, assembly, and safety in construction (Davila Delgado et al., 2020a), enabling a hybrid existence of the real world and virtual environment. AR's adoption and easy use by the operators face many technical barriers and are

constrained by existing hardware design and software platforms (Davila Delgado et al., 2020b). Though researchers want to use AR to facilitate the manipulation of robots, immersive technologies could be challenging to use in terms of user experience and are not human-friendly. Suzuki et al. (2022) systematically reviewed AR and robotics, but it was for general purposes rather than for use in the construction sector.

To address this gap, this research aims to address the following two questions through a systematic literature review.

- What are the key conceptual frameworks for interfaces supporting HRI with AR in construction?
- What are the future research directions to advance HRI with AR and robotics in construction?

The first question aims to identify and evaluate the different interfaces used in construction settings for HRI with AR. The second question explores future trends in integrating AR and robotics within the construction industry. Theoretically, this study investigates the concept of interaction design in HRI with AR research, enhancing understanding and knowledge from this perspective. It classifies knowledge patterns using modularity theory and establishes three conceptual frameworks of interfaces, forming a basis for future research and theoretical exploration.

This study adopts a systematic literature review approach, combining thematic synthesis and framework synthesis methods. A total of 53 relevant articles were retrieved through a structured search using the Web of Science database, focusing on the intersection of AR, robotics, and construction. The inclusion and exclusion criteria were carefully defined to ensure relevance and quality. Thematic synthesis was employed to extract and cluster key interaction trends and subsystem relationships, while framework synthesis enabled the development of conceptual frameworks for AR-HRI interfaces in construction. This dual-method approach ensures a comprehensive and theory-informed review process that underpins the conceptual contributions of the study. This study is organized into six main sections. Section 2 introduces the key concepts underlying augmented-reality human-computer interaction and robotic construction. Section 3 explains the systematic-review methodology. The results are split across Section 4, which maps application trends, and Section 5, which analyses interaction interfaces and conceptual frameworks. Section 6 discusses the theoretical and practical contributions of these findings, while Section 7 concludes the paper and outlines future research directions.

2 Research backgrounds

The concept of HRI with AR involves three categories of objects: humans, robots, and AR. Some studies have used

different taxonomies for them. Han et al. (2021) interpreted human factors in HRI in terms of number (e.g., individual or collective), training level (e.g., trained operators or untrained consumers), and function level (e.g., supportive or disruptive). Walzer et al. (2022) categorised robots in construction (e.g., morphology, color, and material) and established the relationship between human perceptions and robot types. AR hardware systems can be classified into different types, such as head-mounted displays, mobile device-based systems, and projector-based systems (Furht, 2011). Distributed cognition theory provides a holistic perspective on understanding these three fundamental objects for its focus on the whole environment (Hollan et al., 2000). It claims that interaction can enhance and extend individual cognitive abilities by sharing information and coordinating actions (Hutchins, 2020). Accordingly, HRI with AR may enhance construction operators' efficiency and safety by distributing cognitive tasks through interaction between humans, tools, and the environment. Therefore, HRI with AR is not only about decoding the three objects and understanding their internal subsystems but also about the relationship between humans, robots and AR, as well as the relationship generated by the internal subsystems of the three objects.

Some studies focus on the interaction and frameworks of two of these three objects. Existing reviews focus on the purpose of interaction and visualization in interactive AR. Song et al. (2021) identified three types of AR use in digital fabrication, including 3D holographic instruction, data sharing, and HRI. Chen and Xue (2022) claimed that in the context of AR usage, information browsing, and tangible interaction are more commonly utilized than collaborative interaction and hybrid interaction. Some reviews from other industries, though focused on the taxonomy of interaction methods but did not establish connections with construction scenarios. For example, Hertel et al. (2021) classified AR-based interaction techniques into two dimensions, namely task and modality. The task has five sub-categories: creation, selection, geometric manipulation, abstract manipulation, and text input. Modality includes tactile interaction, gestures, voice, gaze, and brain-computer interfaces (Kaufmann et al., 2013). Although these classifications are widely applied in other domains such as manufacturing, health-care, and education, their direct transfer to the construction industry is neither straightforward nor sufficient. Compared to these relatively controlled environments, construction sites present distinct challenges, including constant physical changes in the environment, unpredictable weather, high noise levels, dust, vibration, and strict safety requirements (Almaskati et al., 2024). Voice commands may fail in noisy conditions, while gesture recognition can be hindered by protective clothing, gloves, or occluded body movements. Furthermore, the fragmented and temporary nature of construction teams, as well as the varying levels of digital literacy among

workers, make it difficult to ensure consistent interaction design across different sites (Zulu et al., 2023). These characteristics are rarely addressed in conceptual frameworks developed for other industries. Therefore, there is a critical need to systematically explore and adapt AR-HRI modalities in light of the construction sector's operational complexity, environmental instability, and safety-critical context.

Regarding HRI, while existing classifications help understand the degree of interaction, they have not discussed the specific relationships between interaction systems and lack systematic classification and conceptualisation. For example, Kopp et al., (2021) proposed three degrees of increasing interaction within HRI based on spatial dimension, temporal dimension, and directions: human–robot Coexistence, where humans and robots share a workspace without engaging in common tasks; human–robot Cooperation, where they work toward the same purpose under the same time and space requirements using advanced sensing technologies; and human–robot Collaboration, involving direct or contactless task collaboration, with humans and robots performing complex tasks together, requiring mutual coordination and communication. This classification provides a way to understand the degree of interaction. However, some deeper and more specific relationships between interaction elements have not been addressed or evaluated. For example, the relationship between interaction modes and the direction of interaction commands, as well as how factors such as the number, distance, and size of humans and robots affect the overall depth of interaction. These issues are crucial for establishing human-centered HRI relationships but have not been classified or systematically conceptualised.

This perspective echoes and fits with systems theory. From the system theory perspective, the relationships between subsystems are more critical for the system than the subsystems themselves and the purposes of the interaction themselves (Forrester, 1997; Von Bertalanffy, 1968). In this study, we conceptualise HRI with AR as a socio-technical system comprising seven subsystems, including human, AR system, manipulator, sensors, computational model, physical material, and designed artifact in its environment. Systems theory informs the way these subsystems are identified, linked, and synthesized into interface structures, and it further guides the interpretation of modularity and interdependence patterns, as well as the scenario-based evaluation of their performance in varying construction contexts. Thus, systems theory is not only a conceptual anchor but a methodological compass that shapes both framework construction and comparative analysis. In general use applications, not specific for construction, Suzuki et al. (2022) reviewed interaction dimensions within HRI with AR, such as the position where AR is worn, the

characteristics of the robot, and the modes of interaction. Although this classification comprehensively listed many interaction dimensions and purposes, it did not categorise the relationships between subsystems within AR-HRI. Some studies proposed technical or conceptual frameworks for their technical studies in construction. For example, Wang et al. (2023) proposed a conceptual framework for a head-mounted AR-HCI system named human-cyber-physical system with four phases among three agents, namely designers, HoloLens users, and robot operators. Ootsubo et al. (2016) developed a teleoperation AR-robot system for remote construction. Xiang et al. (2021) proposed a mobile projection AR system, MPAR, for collaborative robot construction. There is a gap in critically and systematically reviewing all these specific frameworks. Nevertheless, there are no studies to the authors' knowledge that investigate a comprehensive taxonomy or frameworks, which could assist the feasibility, effectiveness, and successful adoption of HRI with AR in current practice.

In addition to Suzuki's work, several recent reviews have expanded the scope of HRI research to real-world construction environments. For example, Zhang et al. (2023) conducted a comprehensive review of human–robot collaboration on construction sites, highlighting the practical challenges, task taxonomies, and robotic roles across various field scenarios. However, their analysis primarily focuses on robotic task delegation and physical coordination, while the design of AR interfaces as mediators in the interaction process remains largely unaddressed. More recently, Pan et al. (2024) critically reviewed the integration of extended reality and robotics in construction. Their work identifies research gaps in system integration and discusses interface-related limitations such as perception instability, cognitive overload, and synchronization challenges. Although this review touched on AR applications, it did not present a structured HRI interface typology or systematically analyze interaction frameworks as this study does. Compared to these works, this study fills a unique gap by synthesizing AR-HRI interaction patterns specific to construction, categorizing them into distinct interface frameworks, and mapping them to spatial-temporal interaction dimensions. This theoretical lens contributes to both conceptual clarity and practical guidance for AR-HRI design in on-site construction settings.

Johns et al. (2014) proposed a conceptual interface framework for augmented materiality and embodied computation, which indicates six subsystems in HRI with AR, including manipulators, humans, physical material, computational model, sensors, and designed artifacts in its environment. These six subsystems represent the main components of HRI with AR and can be used to construct the interfaces, relationships, and conceptual framework. The relationships between these six subsystems can be

defined through some specific interaction dimensions, such as the interaction in HRI with AR identified by Suzuki et al. (2022), human factors identified by Han et al. (2021), and AR task elements identified by Hertel et al. (2021). For example, for manipulators, different levels of interaction (no interaction, implicit interaction, explicit and indirect manipulation, explicit and direct physical manipulation) describe how manipulators are used in an AR environment. For humans, the relationship between humans and robots, the skill level of humans using AR (trained, untrained), and the number of users (individual, group) can indicate the complexity and requirements of the interaction, representing this element. All these existing results and knowledge provide a basis to further build a new taxonomy for HRI AR conceptual interface frameworks.

In contrast to prior research that either focuses on isolated perspectives, this study provides a systematic review of HRI with AR for the construction industry. Unlike existing surveys that provide taxonomies without contextualising them in construction tasks or human factors, our review adopts a modular systems perspective to synthesize subsystem relationships and proposes three conceptual interface frameworks (remote modular, proximal modular, and proximal integral). Furthermore, this study uniquely categorises four pairs of subsystems (e.g., human-AR, manipulator-sensor) by their degrees of modularity and integration, offering a scalable and transferable lens for understanding AR-HRI design across diverse construction scenarios. This theoretical advancement not only bridges technical fragmentation in the literature but also provides a structured foundation for future interaction design and implementation strategies in human-robot collaboration within the AEC sector.

3 Methodology

This study adopts a systematic literature review method. Thematic synthesis and framework synthesis were used

for the taxonomy of interaction interface and conceptual frameworks. Thematic synthesis involves extracting, clustering, and synthesizing themes from literature into analytical themes to answer research questions (Xiao and Watson, 2019). Framework synthesis, also referred to as “best fit” framework synthesis (a derivative of the method), involves structuring the coding of the literature by establishing an initial conceptual model, which is then modified based on new evidence to produce a revised framework (Carroll et al., 2013; Dixon-Woods, 2011). The sample of this research mainly includes articles that simultaneously and combinedly use AR and robotics in the construction industry. The articles retrieved are limited to those in English. There are no restrictions on the publication date of the articles. Web of Science was used to search the sample articles. The sampling considers journals, conference papers, and book chapters. Although conference papers and book chapters are not often included in SLR, considering that both AR and robotics are very emerging topics, many real-time studies might not have been published in journal papers yet. Including these types of literature can provide a more comprehensive and latest analysis of the relevant fields. In terms of keyword selection, besides including terms related to AR, robotics, and construction, this study also incorporates keywords associated with Mixed Reality (MR) and Extended Reality (XR) to reflect the broader technological scope. As seen in Fig. 1, XR is a comprehensive concept that encompasses AR, MR, and Virtual Reality (VR) (Rauschnabel et al., 2022). Specifically, AR overlays digital content onto the real-world environment without anchoring it to a physical surface, thereby enhancing user perception without interfering with their perception of the surrounding environment (Yuen et al., 2011). MR advances this technology by enabling real-time interaction between digital elements and physical objects, providing a more integrated user experience (Speicher et al., 2019). In contrast, VR generates a fully immersive virtual environment that completely replaces the physical world (Anthes et al., 2016). Given that human-computer

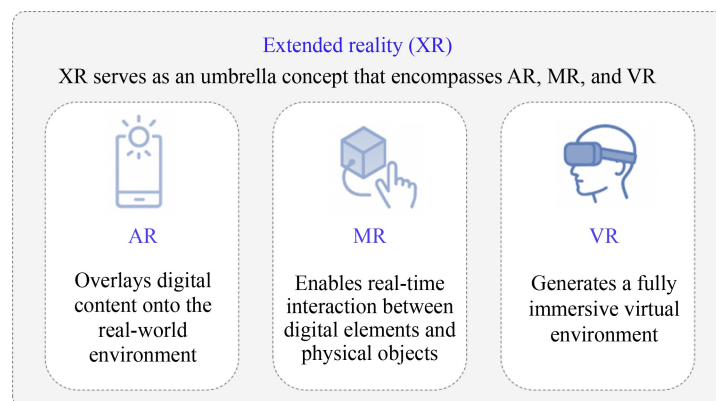


Fig. 1 Definitions between XR, AR, MR and VR in the contexts of immersive technologies.

interaction on construction sites requires real-world interaction, spatial perception, and physical environment (Wang et al., 2021), VR was specifically excluded from both the keyword list and the scope of this study due to its limited applicability in such scenarios. The final set of retrieval keywords employed in the Web of Science database was: (“augmented reality” OR AR OR “mixed reality” OR MR) AND robot* AND (“construction” OR “AEC” OR “Architectural Engineering and Construction”), with the search conducted across all fields.

This research mainly used abductive and inductive codings (see Appendix A). Abductive coding interprets data by comparing it to existing theories for the best explanations, while inductive coding generates codes directly from raw data without pre-existing theories (Adu, 2019). This research adopts abductive coding to localize the existing taxonomy of AR, but it recognizes that none of them is fully suitable for understanding. At the same time, the use of inductive coding could be better for understanding some specific categories that exist in the construction industry. For those categories that adopted abductive coding, this research combines eight interaction dimensions from Suzuki et al. (2022), human elements from Han et al. (2021), and task elements from Hertel et al. (2021). All these results were then used to understand the interaction conceptual frameworks from six subsystems derived from Johns et al. (2014): 1) manipulators, 2) human, 3) physical material, 4) computational model, 5) sensors, and 6) designed artifact in its environment. Additionally, a new subsystem, 7) AR system, was added. Subsequently, two authors (i.e., the first and the second author) conducted the code analysis by categorising all results into MS Excel separately and finally synthesizing them through a last-round review. Through the coding process of these 16 dimensions, trends in interaction design in this field can be understood, especially by establishing relationships with the seven subsystems to know how the interactions of these three objects are characterized. Subsequently, by classifying the characteristics and relationships of these seven subsystems, the study aims to achieve a comprehensive framework, abstracting several typical interaction conceptual frameworks from existing research.

Systems theory, with its long history of constructing conceptual frameworks across various disciplines, provides a lens to construct, evaluate, and interpret different interaction interfaces and conceptual frameworks for framework synthesis. Systems theory encompasses key constructs such as elements, boundaries, inputs, outputs, feedback, self-organization, and the interactions within and between systems and their environments (Luhmann et al., 2013). These constructs are used to systematically analyze the components, relationships, and feedback loops within the HRI with the AR conceptual interface framework. Then, the modularity lens, mainly developed

since Baldwin and Clark (2000), were used to investigate the dependence and independence between these elements within the whole system, which contributes to the categorisation of different frameworks. Furthermore, systems theory provides the epistemological grounding for our framework synthesis. It allows us to structure our coding around systemic components (subsystems), their boundaries, and their interactions. This approach facilitates a deeper understanding of how modular and integral relationships among technical and human elements emerge across different AR-HRI applications in construction. The framework synthesis process thus reflects the systemic view by treating interface configurations as outcomes of system interdependencies rather than isolated technological functions.

Figure 2 illustrates the overall research workflow adopted in this study, encompassing three major stages: data retrieval, data preprocessing, and data analysis. The process began with a structured keyword-based search on the Web of Science database, resulting in an initial pool of 426 articles. These were filtered through well-defined inclusion and exclusion criteria, yielding 49 studies, with 4 additional papers identified via snowball sampling, leading to a final data set of 53 articles. In the preprocessing stage, a coding matrix was developed to extract 16 interaction-related dimensions, such as AR hardware placement, interactivity levels, and task types. These were systematically mapped to seven conceptual subsystems grounded in systems theory. Subsequently, the data analysis phase focused on categorising application domains, identifying subsystem interface patterns, and synthesizing three conceptual interface frameworks. This structured workflow ensures methodological rigour and provides a replicable pathway for future studies on AR-HRI in construction contexts.

4 Categorisation of applications

For the overall trend, publications gradually increased after 2018 (see Fig. 3). This may be due to 2017 being a breakout year for AR technology, as Apple and Google launched ARKit and ARCore, respectively. Microsoft also released two generations of AR products around this time: the first-generation HoloLens in 2016 and the second-generation HoloLens in 2019. Approximately 45.3% (24 papers) of the sample employed the HoloLens AR system. It can be anticipated that with the introduction of new AR devices, such as Apple Vision Pro, research in this field will continue to grow. Among the 33 papers that report material information, 54.5% (18 papers) of the sample focus on digital timber fabrication or assembly, which dominates the most portion among all other materials, like steel, stone, and concrete. For the application building life cycles, as shown in Fig. 4, 22.6% (12 papers) of the studies are about manufacturing tasks,

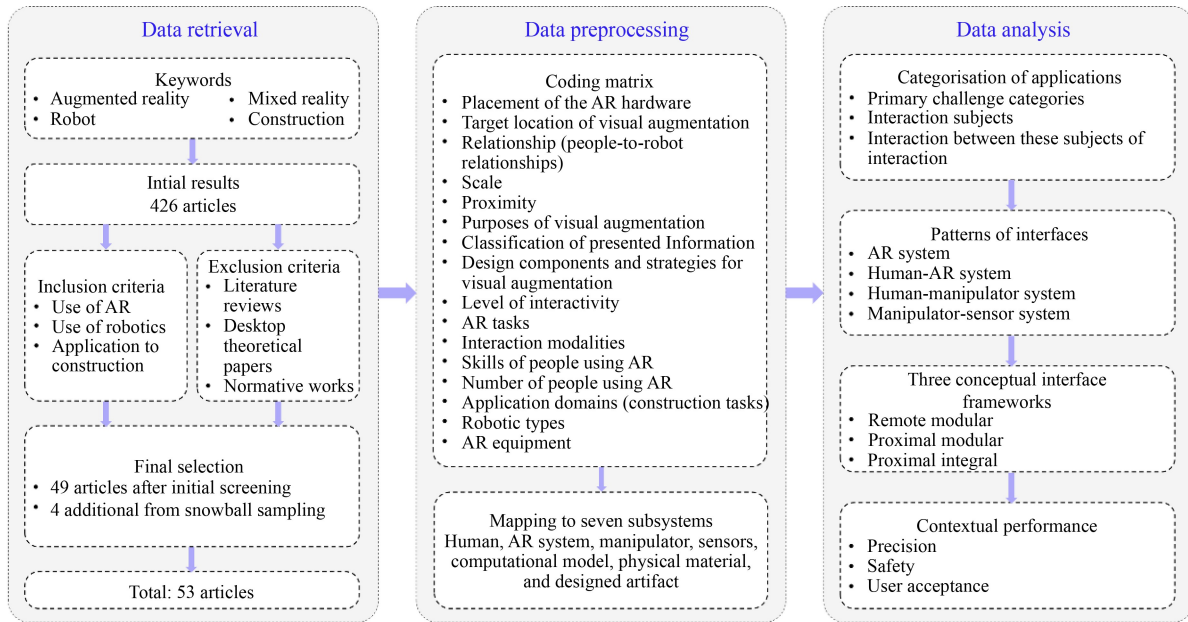


Fig. 2 Overview of the research workflow.

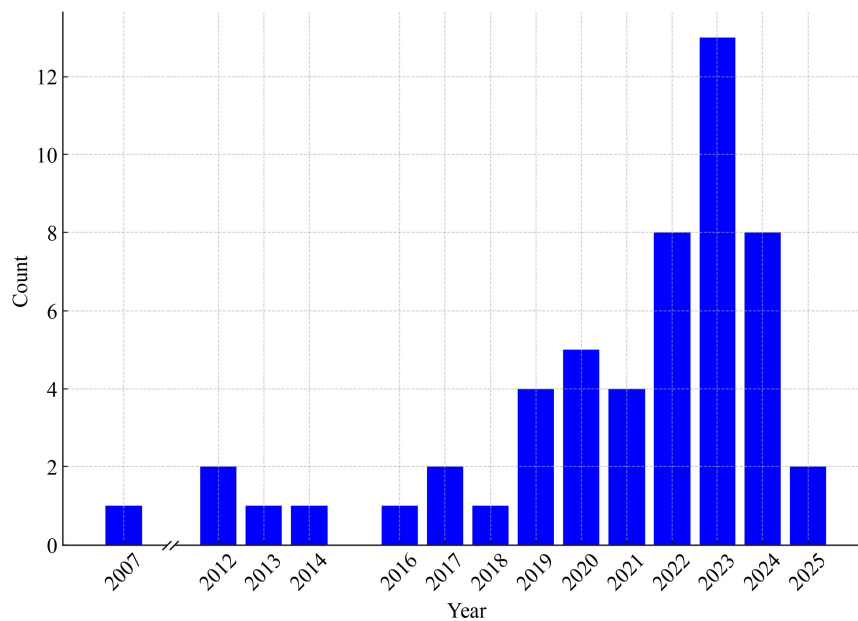


Fig. 3 Yearly publication histogram of AR-HRI research.

39.6% (21 papers) address assembly tasks, and 26.4% (14 papers) focus on construction site management, such as monitoring (9 papers), waste sorting (2 papers), crane control (1 paper), slope shaping (1 paper), and plastering (1 paper). Five papers address the design phase to assist with design visualization, and one paper addresses the maintenance phase, specifically damage detection and barricade installation (Bavelos et al., 2024). The key challenges in construction-related HRI that AR technologies aim to address. Four primary challenge categories emerged across the 53 studies: (i) enhancing visual communication for operational understanding (e.g.,

displaying plans, structure, or process information), (ii) improving accuracy and efficiency in task execution (e.g., assisting robotic assembly or manipulation), (iii) facilitating seamless human-robot coordination (e.g., real-time interface augmentation), and (iv) enabling safety and remote operation (e.g., risk reduction via telepresence or environment monitoring). Among the reviewed papers, 11 focused on addressing communication and visualization issues, 11 emphasized operational efficiency and manipulation precision, 20 targeted interface augmentation for collaboration, and 11 were oriented toward remote or hazardous operation contexts.

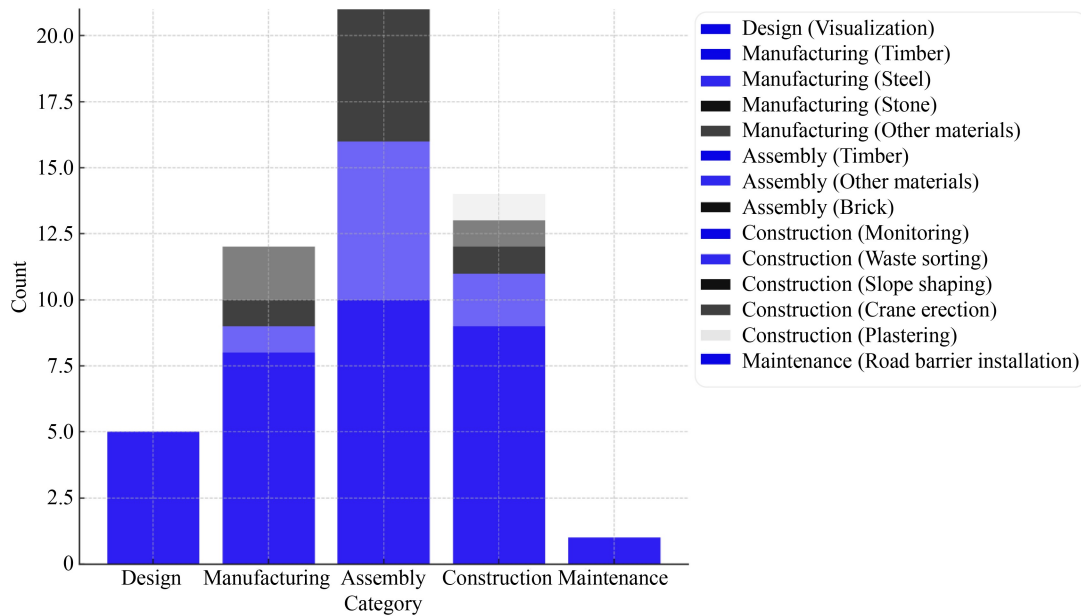


Fig. 4 Distribution of application domains by lifecycle phase.

The second dimension concerns interaction subjects (i.e., humans, robots, and AR). For human factors, all samples adopted trained AR users with professional construction skills. This is related to the application scenarios currently involved in AR-HRI research, which are all production activities rather than for non-professionals. For AR equipment (see Fig. 5, outer ring), a total of 62.3% (33 papers) of the studies used head-mounted AR (i.e., HoloLens, HTC Vive, and Magic Leap), forming an integral human-AR system. About 37.7% (20 papers) used loosely coupled AR displays and cameras. For robot sizes (see Fig. 5, middle ring), the majority were tabletop-scale (24 papers) or body-scale (24 papers), with only five papers employing large-scale robots, such as excavators (Ootsubo et al., 2016, Ootsubo et al., 2013) and cranes (Chi et al., 2012). For the AR-equipped human-robot ratio (see Fig. 5, inner ring), 73.6% (39 papers) of the sample involved a 1:1 ratio of AR-equipped individuals to machines, indicating that research on multi-human and multi-machine collaborative operations is still in its very early stages. From a functional standpoint, different types of AR devices present distinct advantages and limitations, influencing their suitability for specific HRI tasks in construction. Head-mounted AR devices, such as HoloLens or Magic Leap, offer immersive visual overlays aligned with the user's line of sight, making them highly effective for close-range, high-precision tasks like robotic assembly, material placement, and inspection (Zari et al., 2023). These systems enable hands-free operation and real-time guidance, but may also cause user fatigue, limited field of view, or discomfort during prolonged use. In contrast, hand-held AR (e.g., smartphones or tablets) and projection-based systems provide

more flexible and collaborative interfaces for design review, site walkthroughs, and remote supervision tasks (Choi and Kim, 2013). These devices reduce ergonomic strain and allow multiple users to interact simultaneously, although they may lack the precision alignment and spatial immersion required for complex manipulations. Therefore, head-mounted AR systems are more suited for on-site operations requiring direct interaction with manipulators, while non-wearable systems excel in planning, coordination, and inspection contexts where mobility and user comfort are critical.

The third dimension is about the interaction between these three subjects of interaction. 52.8% (28 papers) of the sample used AR visual augmentation to provide information on plans and activities. Additionally, AR also provides supplementary information, such as structural data of models, evaluation information (Chi et al., 2012), and placement guidelines (Kyaw et al., 2024). As shown in Fig. 6 outer ring, the three major categories of AR's purposes for visual augmentation are: 1) augmenting information to facilitate human operations (52.8%); 2) augmenting the interface between humans and robots (37.7%); and 3) augmenting intelligence for robotic operations (9.4%). After providing AR information, interaction is usually completed through explicit and indirect manipulation between humans and manipulators. Regarding the interaction between AR and humans (see Fig. 6, middle ring), in 45.3% (24 papers) of the samples, there was little direct interaction. 39.6% (21 papers) of the samples used a single modality of interaction, mainly the "touch" method, while 15.1% (8 papers) used multimodal interaction, mostly adding "spatial gesture," with interaction modalities like gaze, voice, and proximity rarely

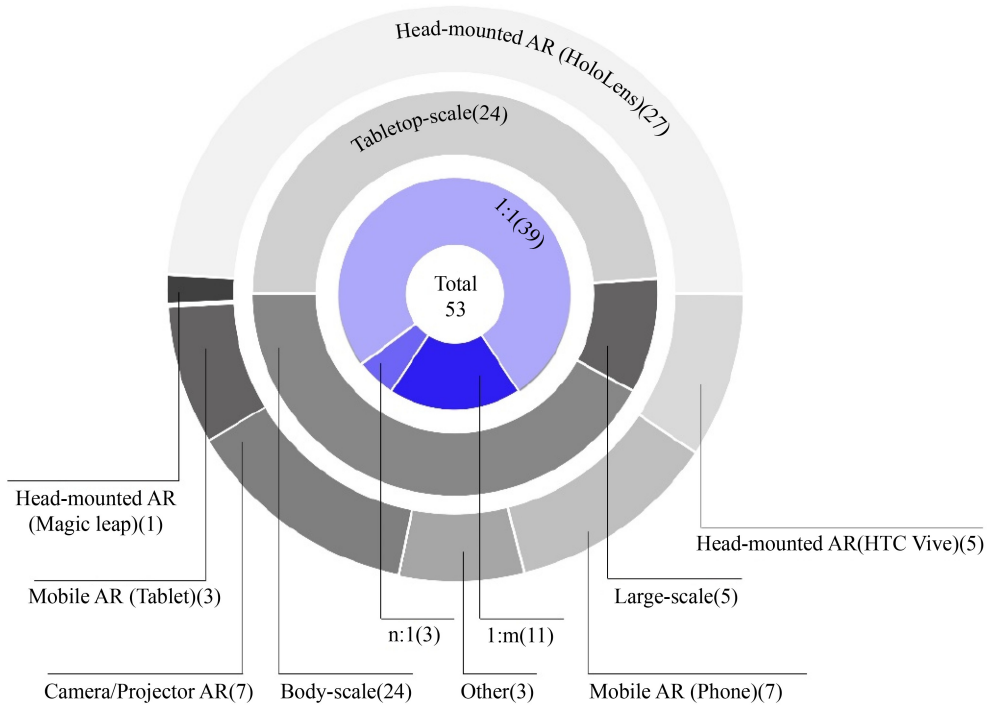


Fig. 5 Dimensions of interaction: equipment, size, and scale.

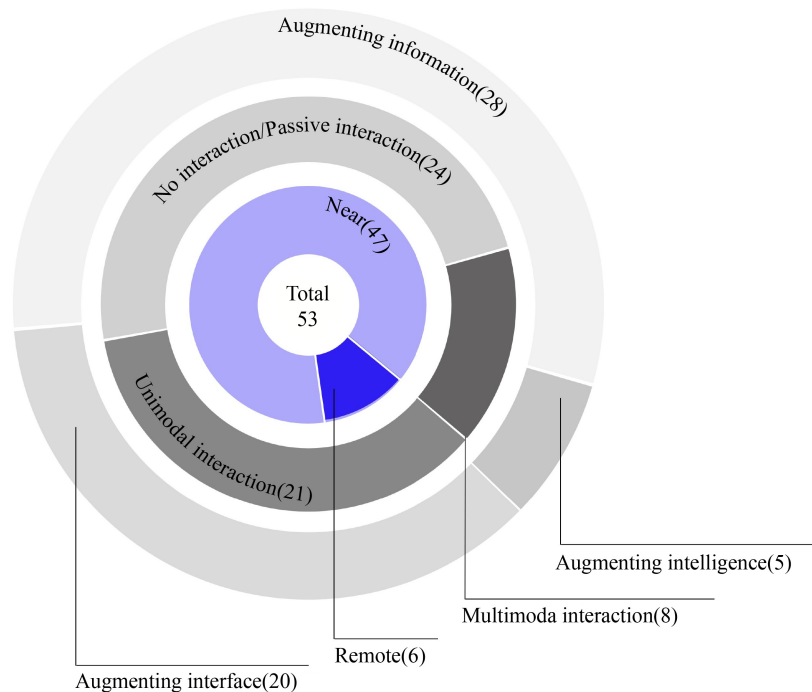


Fig. 6 Interaction modalities.

mentioned or designed. The purpose of these interactions is mainly to utilize robots for selection and geometric manipulation, such as translation (manipulation of an object's position) and rotation (manipulation of an object's orientation). For the human-robot distance (see Fig. 6 inner ring), 88.7% (47 papers) of the studies involved close-range machine operation, while 11.3% (6 papers)

dealt with remote operation.

5 Patterns of interfaces

By synthesizing the dependence and independence between seven subsystems within HRI with AR, this

study identifies four pairs of subsystems that are frequently modularised or integrated: 1) AR camera-AR display system; 2) Human-AR system; 3) Human-manipulator system; 4) Manipulator-sensor system. Dependence and independence are determined by spatial distance and technical structure. Independent and customisable components characterize modular systems, while integral systems exhibit high levels of integration and interdependence, often seen in commercial AR products and unified human-robot interactions. It can be adapted to solve various architectural tasks by reconfiguring the combination of these subsystems.

In a modular AR camera-AR display system, the data collection and display devices operate independently, allowing for flexibility in monitoring different aspects of projects. For example, in large-scale construction sites, independent AR cameras can be strategically placed around the site to monitor progress and gather data, such as the work by Halder et al. (2022) and Bavelos et al. (2024). The data is then transmitted to separate control rooms where displays are customised to present the information as needed. This setup is particularly useful in environments where spatial coverage is extensive and specific monitoring points are crucial. Approximately 17% (9 papers) of the sample adopted a modular AR camera-AR display system, while the remaining 83% (44 papers) used an integral system. In an integral AR camera-AR display system, data collection and display devices are highly integrated, ideally for tasks requiring high precision and immediate data visualization.

A modular human-AR system accounts for about 57% (30 papers) of the studies. In this system, humans and AR devices function independently with minimal dependence. For instance, in a construction project's design and planning phases, projection-based AR devices can be used to display architectural models and blueprints on flat surfaces (Mitterberger et al., 2022). This setup is particularly advantageous for collaborative settings where multiple stakeholders need to view and discuss the project simultaneously. Humans and AR devices are highly integrated in an integral human-AR system, which accounts for about 43% (23 papers). An example of this would be the use of wearable AR devices, such as smart glasses, that provide real-time augmented information directly in the user's field of vision. This setup is highly beneficial for on-site construction work, where workers can receive immediate visual instructions and safety alerts without having to look away from their tasks.

An integral human-manipulator system features a high level of integration between humans and manipulators. In contrast, a modular human-manipulator system refers to setups where humans and manipulators operate separately, often through remote interfaces. For example, in construction tasks that involve hazardous environments or remote locations, operators might use remotely controlled robotic dogs to perform inspections or carry

out specific tasks (Halder et al., 2022), which exemplifies a modular system. In addition, all studies using robotic arms belong to the integral type, as the user is not burdened with the robotic arm system. In an integral human-manipulator system, humans and manipulators are highly unified and closely integrated. For example, in construction sites where heavy machinery such as an excavator is used (Ootsubo et al., 2016, 2013), the operators are often seated within the machine, using an integrated system of controls, sensors, and displays to manipulate the equipment with high precision. This integration allows for real-time feedback and control, enabling the operator to perform complex tasks such as lifting heavy materials or precise excavation work.

In a modular manipulator-sensor system, sensor modules operate independently and provide data that various manipulators can utilize. There are 27 articles mentioning the relationship between sensors and robots, among which 11 are about modular systems. For instance, the work by J. Chen et al. (2023) and Bavelos et al. (2024), in a construction site, standalone sensors might be deployed to monitor environmental or construction conditions. These sensors collect data and transmit it to a central system where different operators or robotic manipulators can access it. This setup allows for flexibility in data usage, as different manipulators can be programmed or controlled based on the specific data they need. In an integral manipulator-sensor system, sensors and manipulators are co-designed to work together. There are 16 articles of this kind. For example, in robots used for construction tasks such as welding, assembling, or bricklaying, sensors are embedded within the robot to provide real-time feedback on position, force, and movement (Ootsubo et al., 2016, 2013; Pedersen et al., 2020). In addition, it also allows for mobile site inspection, such as a camera-embedded robot dog (Halder et al., 2022). This integration enables the manipulator to make immediate adjustments based on the sensor data.

6 Discussion

6.1 Three conceptual interface frameworks

This study thereafter identified three patterns of interface structures that originate from observed subsystem integration patterns and represent the interaction structures, including 1) remote modular interface, 2) proximal modular interface, and 3) proximal integral interface.

Proximal Integral Interface refers to the close-range collaboration between humans, AR, and robots, facilitated by wearable AR devices, and allows for direct physical control of robots. As shown in Fig. 7, the AR system mediates digital control and infographics between humans and computational models, while sensors report data to computational models and provide digital sense

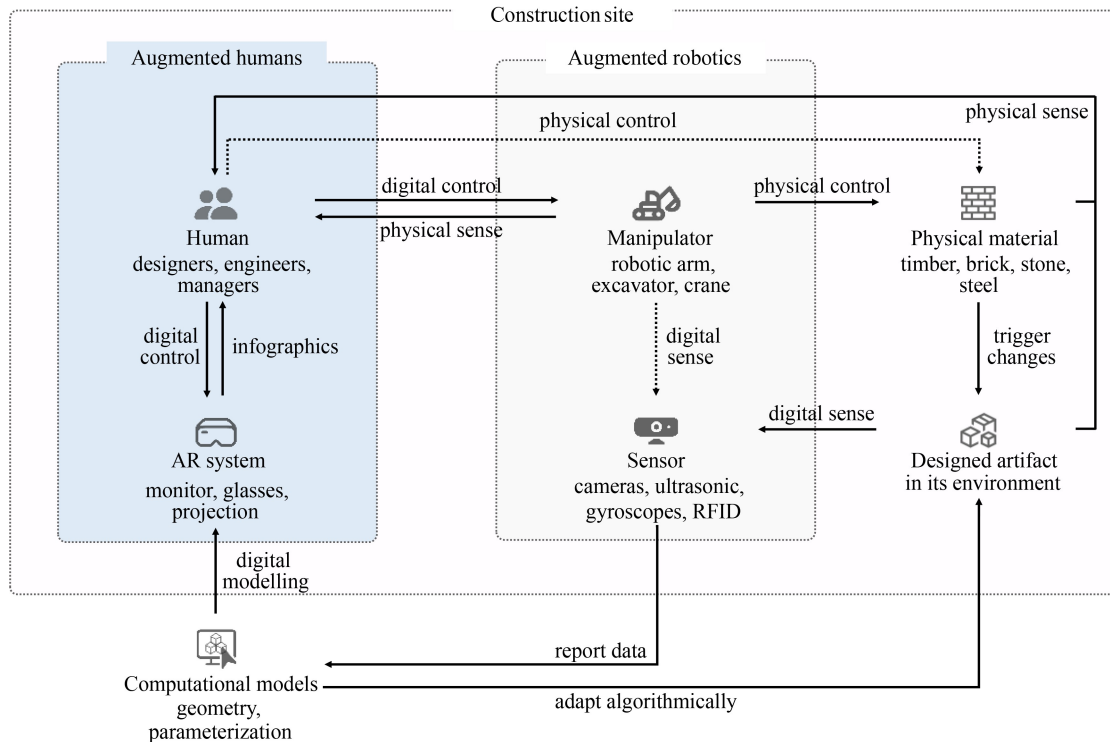


Fig. 7 Proximal integral interface.

feedback to manipulators. Manipulators exert physical control over materials, triggering changes in the designed artifact within its environment. The interaction loop encompasses digital and physical senses, control, and feedback mechanisms to adapt and optimise construction processes. It is a highly integrated system architecture where all subsystems are more closely and multiply connected, with a higher degree of interdependence.

Proximal integral interaction is the most typical way to adopt AR-HRI, with more than half of the sampled studies belonging to this pattern. Kyaw et al. (2024) investigated using gesture interaction to facilitate digital timber fabrication and assembly through the on-site construction of Unlog Tower. Chen et al. (2023) adopted AR for construction waste sorting by collaborating close-range with robot arms. Song et al. (2022) developed a framework for the design and on-site assembly of masonry components. Buyruk and Çağdaş (2022) combined parametric design and human-assisted digital stone fabrication using AR and robot arms. All other studies share similarities in terms of AR-HRI characteristics, like close-range on-site operations and wearable AR. These construction tasks are normally conducted by on-site laborers performing nearby work. In these studies, by introducing visual augmentation through AR and operational augmentation through robotics, AR-HRI helps complete complex construction tasks more efficiently. Key characteristics include real-time augmented information, direct physical contact with objects, and hybrid operational capability compatible with both human and manipulator power.

However, drawbacks include high equipment costs, potential discomfort from prolonged use, limited field of view, and technical complexity.

Proximal Modular Interface refers to proximate, loosely coupled interaction among humans, AR, and robots in on-site works, assisted by non-wearable AR, which allows for direct physical control with robots but more flexible interaction with AR. Unlike the integral interaction, which focuses on the strong dependence between interaction elements, the proximal modular interaction, as shown in Fig. 8, highlights distinct functional modules that are loosely coupled, allowing for more flexible and segmented control and response. Approximately one-quarter of the studies belong to this pattern. These studies used hand-held screen devices (such as smartphones or tablets) or projection equipment to establish an augmented visualization embedded into the environment of the construction site. For example, Mitterberger et al. (2022) developed an interactive robotic plastering system by integrating a mobile projector-based AR interface with a robotic spraying system, enabling users to design and fabricate plasterwork in situ. Johns (2014) explores a prototypical process that integrates projector-based AR, real-time computer simulation, and robotic fabrication to enable spontaneous and intuitive architectural design and manufacturing using recursive wax forming. Some of the other studies used mobile hand-held AR systems. For example, Li et al. (2023) developed a hand-held AR system to assist in self-building construction.

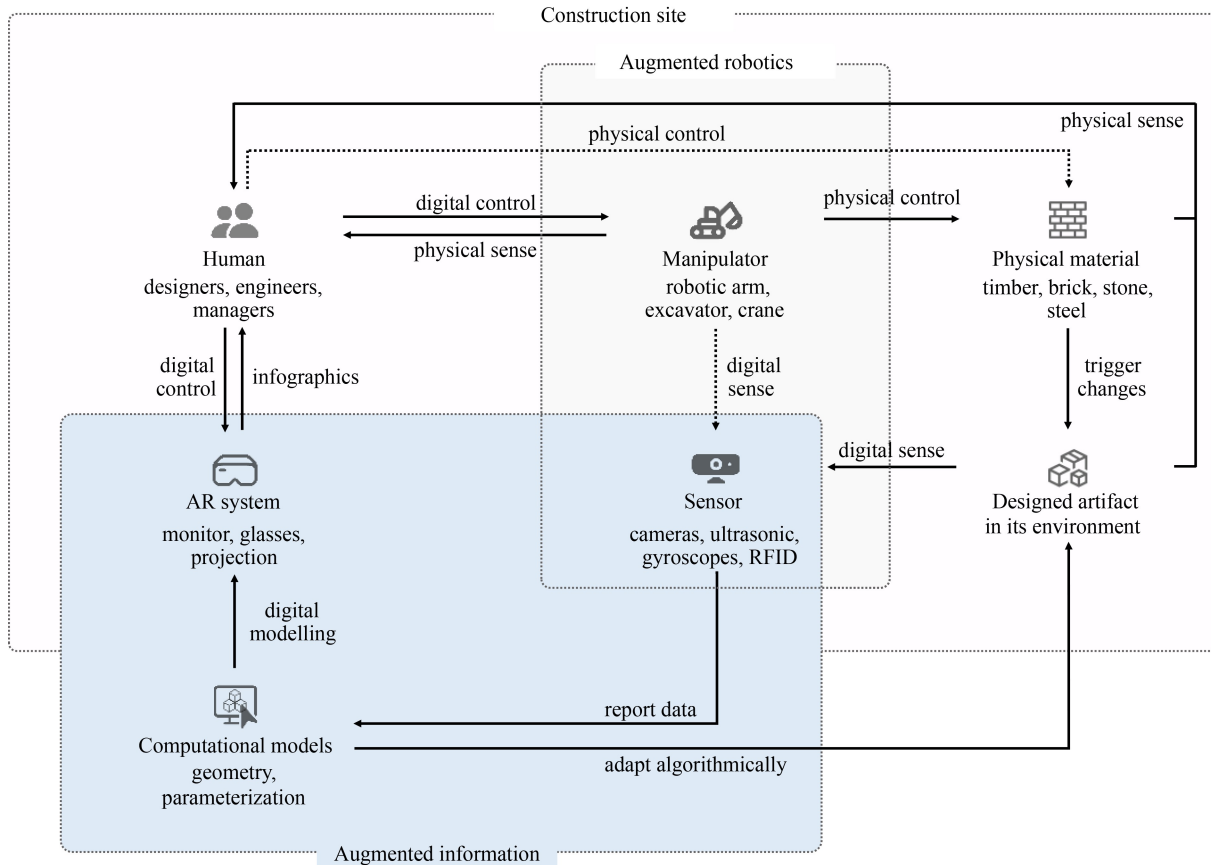


Fig. 8 Proximal modular interface.

A practical real-world application of the proximal modular interface can be seen in the Aurora project for China's Solar Decathlon (Liu et al., 2022). This full-scale building, featuring a steel structural system and complex non-standard bamboo façades, was constructed by student builders using Microsoft HoloLens and the Fologram platform to receive holographic instructions. The MR system enabled real-time adjustments during envelope assembly and rooftop solar structure installation under dynamic outdoor conditions. Through flexible, screen-based, and head-mounted AR interfaces, the implementer completed the installation with sub-centimeter tolerances. The case demonstrated how loosely coupled AR-HRI systems can adapt to material heterogeneity, environmental uncertainty, and human errors, thus validating their potential for scalable deployment in real construction projects.

In this pattern, AR visualization is placed in the environment rather than directly aligning with the operator's line of sight, which makes their relationship more independent. Key interaction characteristics include close-range operations within the same physical space, the use of hand-held or projection devices, and AR content displayed in the environment. The advantages are enhanced user comfort by eliminating head-mounted devices, flexible interaction methods, and simplified

equipment requirements. However, the limitations include the restricted field of view, the potential impact on operation precision, and environmental light affecting projection clarity. These systems are suitable for indoor navigation and information display, construction site inspections, design review and collaboration, and training and demonstrations.

Remote Modular Interface refers to an off-site interaction pattern where humans and AR systems are in control rooms and connect remotely with manipulators working at the construction site on the designed artifact. Unlike proximal interaction, remote modular interaction has two distinctive spatial sites, as shown in Fig. 9. Subsystems within these two parts are strongly connected and interdependent, but the interaction between these two parts is relatively independent and modular, connected only through sensors and control systems.

A few studies (about four) adopted uniquely designed AR systems that separate information collection and display, achieving remote robot operation. For example, Ootsubo et al. (2013) and Ootsubo et al. (2016) developed an AR-assisted teleoperated servo-controlled construction robot with hydraulic cylinders for slope shaping. Chi et al. (2012) proposed an AR-based user interface for a tele-operated crane to reduce erection time. Halder et al. (2022) developed an AR system for the remote control of

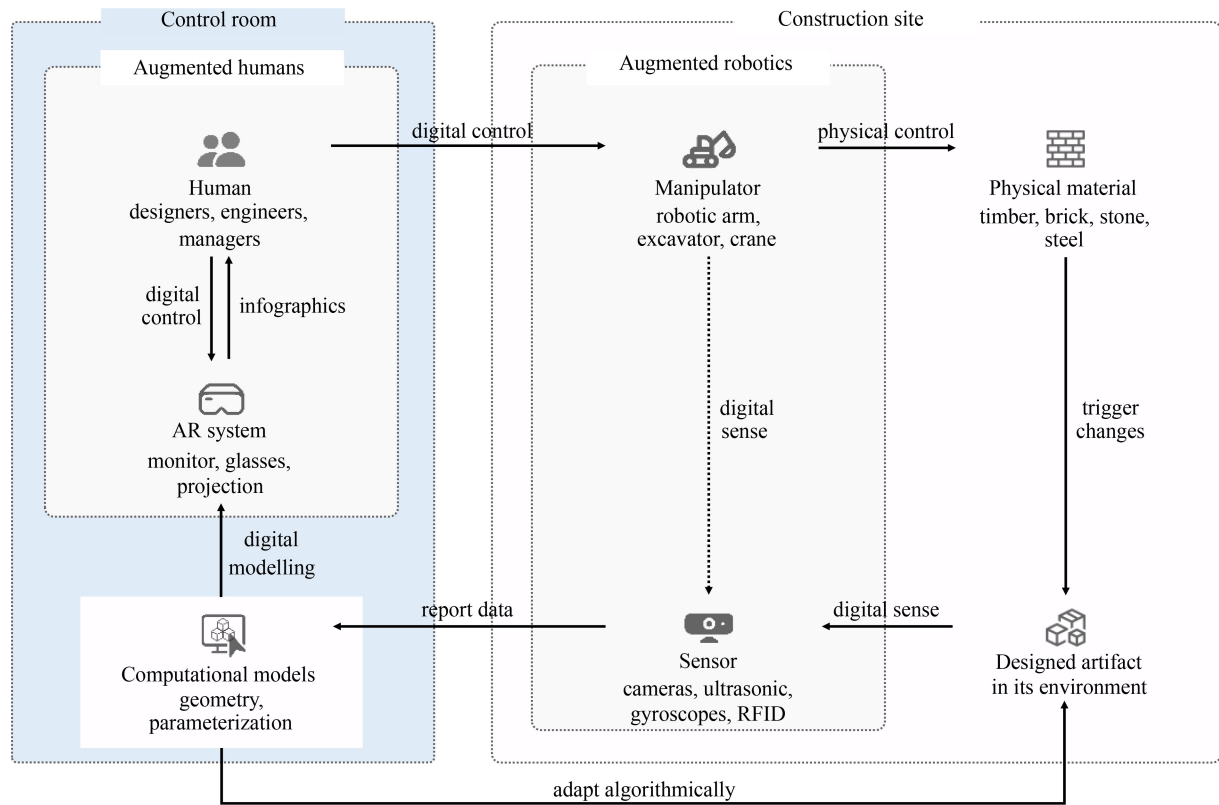


Fig. 9 Remote modular interface.

the Spot Robot for construction progress monitoring. These systems rely on AR devices for remote control and real-time feedback, utilizing network connections and sensor technologies. The advantages of this interface include geographical flexibility, enhanced safety by keeping operators away from hazardous environments, and optimised resource utilization through centralised management. However, they also depend heavily on stable network connections, may experience feedback delays, and have limited surrounding perception.

6.2 Understanding interaction design through a modularity perspective

Drawing from systems theory, we interpret each interaction framework as a composition of interrelated subsystems. This theoretical lens allows us to move beyond component-level analysis and instead examine how systemic coupling, feedback loops, and coordination structures affect interaction quality and contextual fit. The interaction in HRI with AR is largely influenced by the relationships between interaction entities, including the direction and closeness of these relationships, as shown in Fig. 10. For example, in projection-based AR, the relationship between AR and humans is not close; there is no direct interaction between AR and humans, as AR serves only as a display function. However, the relationship between the projection and the designed artifact in its environment

is strong and direct, even adapting to changes. While this study categorises the interaction relationships in HRI with AR into several types, there may be subtle differences in the connections among the seven subsystems in specific studies. The study classifies these relationships based on the direction of data and command transmission as well as the physical spatial distance. From the perspective of modularity theory, the technical frameworks under different interaction design models represent varying degrees of dependence and independence among the seven subsystems. For example, in the remote modular interaction model, the elements are more modular, and their relationships are more independent. A typical feature of this is the separation of the AR display system and the AR data collection system, allowing the system to operate independently and enabling remote functionality.

This modular design also facilitates easier hardware upgrades and replacements (Baldwin et al., 2000). For instance, the independent camera in an AR system can be upgraded without affecting the display system, and the display system's monitor can be upgraded independently to change interaction modes and methods. A simple example is upgrading a monitor from a mere display to a touchscreen, adding touch interaction without affecting data collection. This is difficult or impossible to achieve in integrated systems like the Vision Pro, which combines collection and display functions. The convenience brought by modularisation allows for more

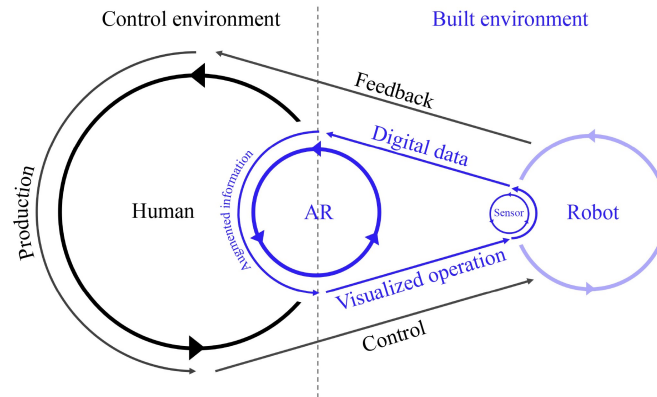


Fig. 10 Interrelationship between human, AR, sensor, and robot.

targeted interaction design (including software and hardware) for specific construction tasks, providing flexibility for future projects. However, while customisation capabilities are enhanced, this also poses challenges for software and hardware compatibility, safety, and reliability, necessitating more considerations in interaction design.

In contrast to modular relationships, proximal integral interaction reflects a more unified and integrated characteristic among the seven subsystems. Whether it is the AR device, the distance between humans and machines, or the distance between humans and construction objects, these interactions are much closer. This integration can lead to improvements in user experience, reliability, and safety. For instance, many existing interaction technologies (such as gestures and eye-tracking) are integrated into unified AR devices (such as Vision Pro or HoloLens), whereas modular, customised AR devices lack this mature and established hardware-software integration environment. This close proximity construction (humans, machines, and construction objects being in close range) and wearable/head-mounted devices (close proximity between humans and AR) reduce the path and latency needed for data transmission, enhancing real-time performance. It also offers more possibilities for interaction, as in this scenario, compared to remote modular interaction, humans can directly interact with artificial objects and materials in the environment without entirely relying on manipulators as intermediaries. Moreover, human senses can be involved directly rather than relying solely on sensors. However, integrated interaction methods may also face issues such as a lack of extensibility and heavy dependence on hardware devices. Therefore, there is no universal interaction design method. Each construction task will have its unique interaction design requirements, influencing how the technical elements must be structured.

6.3 Comparative analysis of interaction frameworks in construction scenarios

Building on the modularity perspective, this section evaluates the three proposed interaction frameworks

across diverse construction scenarios. By analyzing their performance in task efficiency, safety, adaptability, and user acceptance. We highlight context-specific strengths and limitations, as seen in [Table 1](#).

The proximal integral interface excels in high-precision tasks requiring direct physical manipulation, such as timber fastening or robotic arm control, where sub-millimeter accuracy is critical. For instance, research on AR-assisted timber fastening with Microsoft HoloLens 2 demonstrates that AR overlays and real-time feedback improve accuracy in screw and nail placement, reducing human errors ([Fazel and Adel, 2024](#)). However, its reliance on wearable AR devices like HoloLens introduces trade-offs, such as extended setup times and ergonomic strain. Moreover, studies integrating AR with BIM for construction inspections highlight that while head-mounted AR devices enhance monitoring accuracy, they still pose usability concerns related to prolonged usage ([Pan and Isnaeni, 2024](#)). A BIM-based human-robot collaboration framework for building inspections further validates this approach. Studies have shown that integrating MR headsets with quadruped robots improves the visualization, control, and allocation of inspection tasks, enhancing overall efficiency ([Tandon et al., 2025](#)). The proximal modular interface prioritizes flexibility, utilizing hand-held devices or projection-based AR for rapid deployment in experimental tasks and real-world construction projects. For instance, the Aurora project ([Liu et al., 2022](#)) features steel modular structures and geometrically complex bamboo facades, demonstrating that AR interfaces can support untrained workers to efficiently assemble building envelope structures in outdoor environments. The holographic overlay system can adapt to tolerance deviations and support real-time adjustment, demonstrating the practicality of proximal modular HRI-AR in medium and large-scale construction projects. Meanwhile, the remote modular interface demonstrates scalability in large-scale operations. The RoBétArmé Project enhances the application of shotcrete by combining human-robot collaboration with AR, enabling remote operators to effectively guide robot construction tasks

Table 1 Synthesis of framework trade-offs.

Framework	Best for	Avoid for
Proximal integral	High-precision assembly, real-time feedback	In dynamic environments, untrained workers
Proximal modular	Rapid inspections, collaborative design reviews	High-tolerance tasks, low-light conditions
Remote modular	Hazardous environments, multi-robot coordination	Time-critical adjustments, unstable networks

(Kostavelis et al., 2024). However, its reliance on stable networks can lead to delays, thereby limiting its real-time response capability in dynamic environments.

Safety outcomes further differentiate these frameworks. A VR-based study examined how the presence of quadruped robots in construction environments affects workers' attention, risk perception, and attitudes, providing valuable insights for safer collaboration (Albeaino et al., 2025). Although the proximal integral interfaces can achieve real-time hazard feedback, workers face collision risks due to their close proximity to the machine. Remote modular interfaces reduce such risks by isolating operators from hazardous environments, but there is a perception gap that may overlook dynamic on-site situations. Proximal modular interfaces achieve balance, using non-wearable augmented reality to guide workers to maintain a safer distance. The adaptability to material and environmental complexity also varies. Proximal integral interfaces perform best in standardization tasks, but perform poorly when dealing with heterogeneous materials such as recycled wood that require improvisation. The Mixed Reality Carpentry study emphasizes how AR can utilize irregular wood scraps to promote manufacturing and compensate for material inconsistencies through visual overlay and real-time guidance (Jahn et al., 2024). Remote modular interfaces are material-independent and suitable for monitoring or remote operation, but may fail under unpredictable conditions. Proximal modular interfaces have diverse functions, but their accuracy is insufficient for high-tolerance tasks. User acceptance further underscores practical limitations. Proximal integral interfaces demand skilled operators, limiting accessibility in low-skilled labor markets, while their ergonomic strain raises workforce concerns. The use of MR in prefabrication has demonstrated efficiency improvements, yet studies indicate that reducing time uncertainty through AR visualization may centralize control and limit worker autonomy (Sandagomika et al., 2024). Proximal modular interfaces, with intuitive touch/gesture interaction, can enhance novice efficiency. Conversely, remote modular interfaces risk workforce displacement by centralizing control, a tension highlighted in prefabrication studies. In synthesis, the efficacy of each framework hinges on task-specific demands, emphasizing that no single framework universally outperforms others.

6.4 Future research directions

The interaction of AR-HRI has multiple modalities, abilities, and systems, as seen in Fig. 11. Future research

directions could focus on interaction design in relation to two dimensions: construction tasks and humans. Despite the significant impact of interaction methods on task execution outcomes and their application in other industrial scenarios, this area remains unexplored in AR-HRI research within the construction industry. A major barrier to the industrial application of AR-HRI is the lack of user-friendly interaction methods. Future research can explore the relationship between interaction methods and construction tasks. For instance, different interaction method solutions can be investigated for the same construction task. Chi et al. (2012) is the only research within the sample that compares a proposed AR-based user interface with a conventional operation interface to compare erection time using a tele-operated crane system. This research methodology could be adopted for other studies, too. Taking the work by Chen et al. (2023) on waste sorting as an example, various types of interaction designs, such as on-site personnel integrated AR-HRI's proximal integral interaction and remote modular interaction, can be compared to study their impact on the waste sorting task. Such research can accumulate relevant knowledge, enabling the design of more suitable interactions for construction tasks. Additionally, with the evolution of AR devices and robots, utilizing more interaction modalities or intelligent agents for collaborative interaction has become an emerging field. Most existing studies primarily use touch interaction, while a few have started to focus on gesture interaction. With the introduction of Apple Vision Pro, eye-tracking interaction is expected to become a new hands-free solution, along with cutting-edge brain-computer interaction. Investigating the impact of these different interaction modalities on construction tasks is another important direction for the future.

On the other hand, humans are a crucial part of HRI with AR. Although there has been extensive research on human impact in interaction design, none of the 53 existing HRI with AR studies have explored the relationship between interaction design and human impact. Future research could focus on human-centric interaction design and human-centric human-robot interaction rather than solely robot-oriented design. For example, remote modular interaction can isolate operators from hazardous sites, demonstrating how interaction methods can directly affect humans. Interaction design has introduced many concepts and theories that emphasize human factors, such as emotional design and experience design (Hassenzahl, 2013; Ho and Siu, 2012). These emerging theories and approaches have not yet been discussed in the context of AR-HRI in the construction

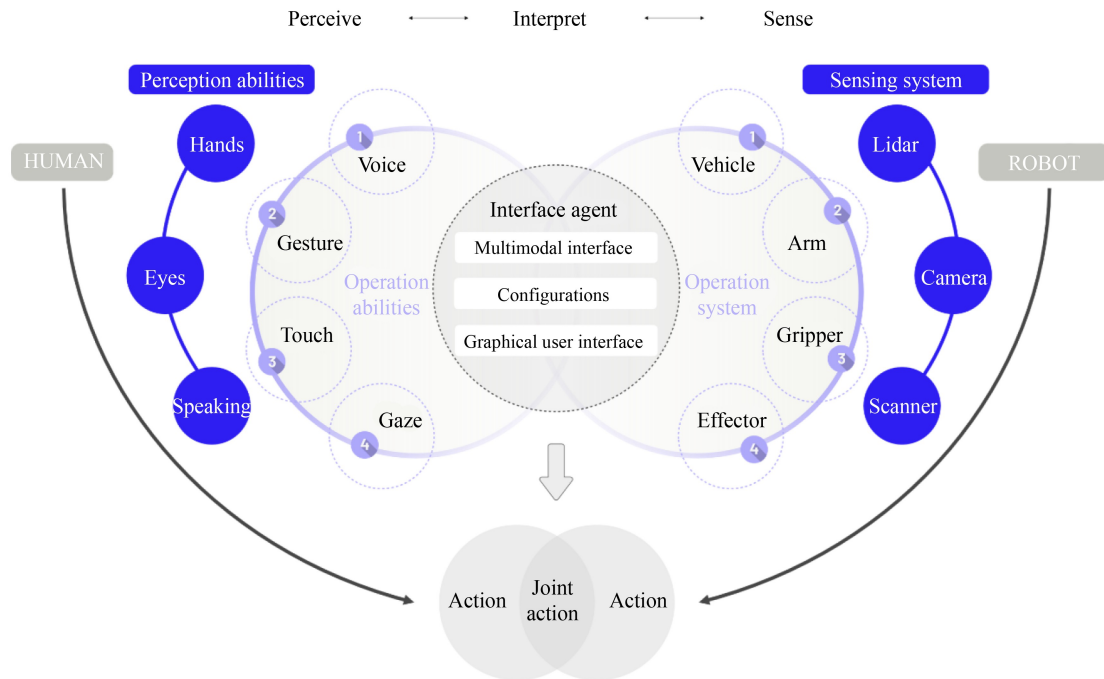


Fig. 11 Multimodal Interaction Framework of HRI with AR.

industry. Currently, most AR-HRI research focuses on trained construction professionals performing specialized tasks (e.g., manufacturing, assembly).

To improve the real-world applicability and user-friendliness of AR-HRI systems in construction, future research should more explicitly incorporate human-centered design principles, with particular attention to user needs, ergonomics, and cognitive load. Although AR technology has the potential to improve efficiency, research has shown that poor availability and high mental load can seriously hinder its application in industrial environments (Palmarini et al., 2018; Rogers and Preece, 2011). Ergonomic evaluations should be systematically integrated into interaction design (Wu et al., 2023). In addition, adopting cognitive load theory in AR interface design can help information presentation, reduce visual confusion, and support decision-making under complex conditions (Makransky et al., 2019). Research should also emphasize participatory design methods by involving end-users early in development to ensure alignment with real-world workflows and preferences. In addition, emotional and experiential aspects such as comfort, perceived usefulness, and trust are crucial for user acceptance and engagement. To further advance AR-HRI design, emotional or experiential features such as comfort and confidence cues should be considered as evaluation metrics and active design inputs (Wu et al., 2024). Incorporating these features in the early stages of design can enhance user engagement and trust, especially in high-pressure construction scenarios (Johansen et al., 2024). These factors are particularly important in high-pressure construction environments, as human perfor-

mance is often limited by fatigue and environmental stress factors. Future studies are therefore encouraged to evaluate task-based outcomes and subjective experience metrics such as usability ratings, workload scales, and ergonomic assessments to inform more human-centered AR-HRI solutions.

7 Conclusions

This study presents a review of HRI with AR in the construction field, focusing on interactions among humans, AR, and robots. It identifies and synthesizes patterns of interaction relationships within technical systems by adopting a modularity theoretical lens. From the perspective of applications, the study reveals the impact of hardware development on interaction methods. As AR devices evolve, HRI with AR interaction methods transition from augmenting information to augmenting interfaces, thereby promoting human–robot collaboration rather than mere coexistence. It is anticipated that with advancements in automation and robotics, HRI with AR interaction research will gradually shift to augmenting intelligence. As machines gain a certain level of autonomy, the focus will shift to enhancing the application of intelligent autonomy. Currently, most studies exhibit a strong experimental nature, often using exploratory materials such as timber, biological, or model materials. Most real construction scenarios, such as civil buildings made of concrete and steel structures, have not received sufficient attention and application. Group interactions involving multi-human and multi-machine collaborative operations

have also received little attention. These gaps highlight the areas in need of further research in this field.

From the interaction perspective, this study explores the relationships among seven subsystems of HRI with AR: manipulators, humans, physical materials, computational models, sensors, designed artifacts in their environments, and AR. Applying modularity theoretical concepts of dependence and independence among system subsystems, the study classifies these relationships based on the direction of information and command transmission as well as the physical spatial distance. It identifies three main patterns of interactions among the seven subsystems: 1) remote modular interface, 2) proximal modular interface, and 3) proximal integral interface. Four subsystem pairs that can either be modularised or integrated are identified: AR system, human-AR system, human-manipulator system, and manipulator-sensor system. The study explores their respective characteristics and suitable application scenarios by identifying and analyzing the technical dependencies and independencies among these

different elements.

Theoretically, this study attempts to introduce the concept of interaction design into HRI with AR research, enhancing the understanding of related studies from this perspective. It classifies the identified knowledge patterns and establishes three conceptual frameworks of interfaces, laying a foundation for future interaction design research and theoretical exploration. From a practical significance perspective, future HRI with AR applications can utilize the three interaction relationship patterns for different tasks. According to specific task requirements, the four subsystem pairs can be combined by establishing independence to form technical modules or by establishing dependence to create integration. This approach allows the HRI with an AR system to adapt to various task needs by forming different element relationship characteristics. Lastly, the study highlights the research gaps and future directions under the interaction perspective of HRI with AR, specifically in construction-oriented and human-centric interaction design.

Appendix A: Codes used for data analysis

No.	Category	Sub-categories	Codes	References
1	Approaches to AR	Placement of the AR hardware	1. On-body 2. On-environment 3. On-robot	Suzuki et al. (2022)
2		Target location of visual augmentation	1. Augmenting robots 2. Augmenting surroundings	
3	Characteristics of augmented robots	Relationship (people-to-robot relationships)	1. 1:1 2. 1:m 3. n:1 4. n:m	
4		Scale	1-4 (handheld-scale, tabletop-scale, body-scale, large-scale)	
5		Proximity	1-3 (on; near; remote)	
6	Purposes of visual augmentation		N/A (Inductive coding)	N/A
7	Classification of presented information		1. Robot's internal information 2. External information about the environment 3. Plan and activity 4. Supplemental content	Suzuki et al. (2022)
8	Design components and strategies for visual augmentation		UIs and widgets Spatial references and visualisations Embedded visual effects	
9	Level of interactivity		1. No interaction 2. Implicit interaction 3. Explicit and indirect manipulation 4. Explicit and direct physical manipulation	
10	AR tasks		1. Creation 2. Selection 3. Geometric manipulation 4. Abstract manipulation 5. Text input	Hertel et al. (2021)
11	Interaction modalities		1. Tangible 2. Touch 3. Pointer and controller 4. Spatial gesture 5. Gaze 6. Voice 7. Proximity	Suzuki et al. (2022)

(Continued)

No.	Category	Sub-categories	Codes	References
12	Human Factors	Skills of people using AR	1. Trained (e.g. operators) 2. Untrained (e.g. consumers and passerby)	Han et al. (2021)
13		Number of people using AR	1. Individual 2. Collective	
14	Application domains (construction tasks)		N/A (Inductive coding)	N/A
15	Robotic types			
16	AR equipment			

Competing Interests The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made.

The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adamides G, Christou G, Katsanos C, Xenos M, Hadzilacos T (2015). Usability guidelines for the design of robot teleoperation: A taxonomy. *IEEE Transactions on Human-Machine Systems*, 45(2): 256–262
- Adup P (2019). *A step-by-step guide to qualitative data coding*. Routledge
- Agarwal R, Chandrasekaran S, Sridhar M (2016). *Imagining construction's digital future*. McKinsey & Company
- Ajoudani A, Zanchettin A M, Ivaldi S, Albu-Schäffer A, Kosuge K, Khatib O (2018). Progress and prospects of the human–robot collaboration. *Autonomous Robots*, 42(5): 957–975
- Al Masri A, da Costa B B F, Vasco D, Boer D, Haddad A N, Najjar M K (2024). Roles of robotics in architectural and engineering construction industries: Review and future trends. *Journal of Building Design and Environment*, 2(1): 28029
- Albeaino G, Jeelani I, Gheisari M, Issa R R A (2025). Assessing proxemics impact on human–robot collaboration safety in construction: A virtual reality study with four-legged robots. *Safety Science*, 181: 106682
- Almaskati D, Kermanshachi S, Pamidimukkala A, Loganathan K, Yin Z (2024). A review on construction safety: Hazards, mitigation strategies, and impacted sectors. *Buildings*, 14(2): 526
- Anthes C, García-Hernández, R J, Wiedemann M, Kranzlmüller D (2016). State of the art of virtual reality technology. In: *IEEE Aerospace Conference, Big Sky*
- Baldwin C Y, Clark K B (2000). *Design rules: The power of modularity*. Cambridge, MA: MIT Press
- Bavelos A C, Anastasiou E, Dimitropoulos N, Oikonomou G, Makris S (2024). Augmented reality-based method for road maintenance operators in human–robot collaborative interventions. *Computer-Aided Civil and Infrastructure Engineering*, 39(7): 1077–1095
- Bloss R (2016). Collaborative robots are rapidly providing major improvements in productivity, safety, programming ease, portability and cost while addressing many new applications. *Industrial Robot. International Journal*. 43(5): 463–468
- Bock T (2015). The future of construction automation: Technological disruption and the upcoming ubiquity of robotics. *Automation in Construction*, 59: 113–121
- Boschetti G, Faccio M, Granata I (2022). Human-centered design for productivity and safety in collaborative robots cells: A new methodological approach. *Electronics*, 12(1): 167
- Bouchlaghem D, Shang H, Whyte J, Ganah A (2005). Visualisation in architecture, engineering and construction (AEC). *Automation in Construction*, 14(3): 287–295
- Brosque C, Galbally E, Khatib O, Fischer M (2020). human–robot collaboration in construction: Opportunities and challenges. In: *2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*. IEEE, 1–8
- Burden A G, Caldwell G A, Guertler M R (2022). Towards human–robot collaboration in construction: Current cobot trends and forecasts. *Construction Robotics*, 6(3–4): 209–220
- Buyruk Y, Çağdaş G (2022). Interactive parametric design and robotic fabrication within mixed reality environment. *Applied Sciences*, 12(24): 12797
- Carroll C, Booth A, Leaviss J, Rick J (2013). “Best fit” framework synthesis: Refining the method. *BMC Medical Research Methodology*, 13(1): 37
- Chen J, Fu Y, Lu W, Pan Y (2023). Augmented reality-enabled human–robot collaboration to balance construction waste sorting efficiency and occupational safety and health. *Journal of Environmental Management*, 348: 119341
- Chen K, Xue F (2022). The renaissance of augmented reality in construction: History, present status and future directions. *Smart and Sustainable Built Environment*, 11(3): 575–592
- Chi H L, Chen Y C, Kang S C, Hsieh S H (2012). Development of user interface for tele-operated cranes. *Advanced Engineering Informatics*, 26(3): 641–652
- Chiang Y H, Tao L, Wong F K (2015). Causal relationship between construction activities, employment and GDP: The case of Hong Kong. *Habitat International*, 46: 1–12
- Choi J, Kim G J (2013). Usability of one-handed interaction methods for handheld projection-based augmented reality. *Personal and Ubiquitous Computing*, 17(2): 399–409

- Davila Delgado J M, Oyedele L, Beach T, Demian P (2020a). Augmented and virtual reality in construction: Drivers and limitations for industry adoption. *Journal of Construction Engineering and Management*, 146(7): 04020079
- Davila Delgado J M, Oyedele L, Demian P, Beach T (2020b). A research agenda for augmented and virtual reality in architecture, engineering and construction. *Advanced Engineering Informatics*, 45: 101122
- Dixon-Woods M (2011). Using framework-based synthesis for conducting reviews of qualitative studies. *BMC Medicine*, 9(1): 39
- Fazel A, Adel A (2024). Enhancing construction accuracy, productivity, and safety with augmented reality for timber fastening. *Automation in Construction*, 166: 105596
- Forrester J W (1997). Industrial dynamics. *Journal of the Operational Research Society*, 48(10): 1037–1041
- Frijns H A, Schürer O, Koeszegi S T (2023). Communication models in human–robot interaction: An asymmetric MODEL of ALterity in human–robot interaction (AMODAL-HRI). *International Journal of Social Robotics*, 15(3): 473–500
- Fu Y, Chen J, Lu W (2024). human–robot collaboration for modular construction manufacturing: Review of academic research. *Automation in Construction*, 158: 105196
- Furht B, ed (2011). *Handbook of augmented reality*. Springer Science & Business Media
- Galin R R, Meshcheryakov R V (2020). human–robot interaction efficiency and human–robot collaboration. In *Robotics: Industry 4.0 issues & new intelligent control paradigms* (pp. 55–63). Springer
- Graser K, Kahlert A, Hall D M (2021). DFAB HOUSE: implications of a building-scale demonstrator for adoption of digital fabrication in AEC. *Construction Management and Economics*, 39(10): 853–873
- Halder S, Afsari K, Serdakowski J, DeVito S, Ensafi M, Thabet W (2022). Real-time and remote construction progress monitoring with a quadruped robot using augmented reality. *Buildings*, 12(11): 2027
- Han I X, Meggers F, Parascho S (2021). Bridging the collectives: A review of collective human–robot construction. *International Journal of Architectural Computing*, 19(4): 512–531
- Hassenzahl M (2013). User experience and experience design. In: *The Encyclopedia of Human–Computer Interaction* (2nd ed.). Copenhagen: Interaction Design Foundation
- Hertel J, Karaosmanoglu S, Schmidt S, Bräker J, Semmann M, Steinicke F (2021). A taxonomy of interaction techniques for immersive augmented reality based on an iterative literature review. In: *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 431–440
- Ho A G, Siu K W M G (2012). Emotion design, emotional design, emotionalize design: A review on their relationships from a new perspective. *Design Journal*, 15(1): 9–32
- Hollan J, Hutchins E, Kirsh D (2000). Distributed cognition: toward a new foundation for human–computer interaction research. *ACM Transactions on Computer–Human Interaction*, 7(2): 174–196
- Hutchins E (2020). The distributed cognition perspective on human interaction. In *Roots of human sociality*. Routledge, 375–398
- Jahn G, Newnham C, Berg N (2024). Mixed reality carpentry. In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, NY, Article 843, 1–15
- Johansen S S, Burden A, Schneiders E, Walzer A N (2024). Designing for people in human–robot collaboration. *Interaction Design and Architecture*, 61(61): 5–10
- Johns R L (2014). Augmented materiality: Modelling with material indeterminacy. *Fabricate*, 2014: 216–223
- Johns R L, Kilian A, Foley N (2014). Design approaches through augmented materiality and embodied computation. In *Robotic Fabrication in Architecture, Art and Design 2014* (pp. 319–332). Springer
- Kaufmann T, Holz E M, Kübler A (2013). Comparison of tactile, auditory, and visual modality for brain–computer interface use: A case study with a patient in the locked-in state. *Frontiers in Neuroscience*, 7: 129
- Khan A, Halimi S M M, Saad S, Rasheed K, Ammad S (2025). Robotics in construction: Transforming the built environment. In *Applications of Digital Twins and Robotics in the Construction Sector* (pp. 23–48). CRC Press
- Kopp T, Baumgartner M, Kinkel S (2021). Success factors for introducing industrial human–robot interaction in practice: An empirically driven framework. *International Journal of Advanced Manufacturing Technology*, 112(3–4): 685–704
- Kostavelis I, Nalpantidis L, Detry R, Bruyninckx H, Billard A, Christian S, Bosch M, Andronikidis K, Lund-Nielsen H, Yosefipor P, Wajid U, Tomar R, Martínez F L L, Fugaroli F, Papargyriou D, Mehandjiev N, Bhullar G, Gonçalves E, Bentzen J, Essenbæk M, Cremona C, Wong M, Sanchez M, Giakoumis D, Tzovaras D (2024). RoBétArmé Project: human–robot collaborative construction system for shotcrete digitization and automation through advanced perception, cognition, mobility and additive manufacturing skills. *Open Research Europe*, 4(4): 4
- Krupke D, Starke S, Einig L, Zhang J, Steinicke F (2018). Prototyping of immersive HRI scenarios. In: *2018 27th IEEE International Symposium on Robot and Human Interactive Communication*. 1035–1040
- Kyaw A H, Spencer L, Lok L (2024). Human–machine collaboration using gesture recognition in mixed reality and robotic fabrication. *Architectural Intelligence*, 3(1): 11
- Li Y, Hu Y, Tan T, Yu B, Li J, Fingrut A (2023). AR-assisted assembly in self-build construction with discrete components. In: *30th EG-ICE International Workshop on Intelligent Computing in Engineering*
- Liang C J, Wang X, Kamat V R, Menassa C C (2021). Human–robot collaboration in construction: Classification and research trends. *Journal of Construction Engineering and Management*, 147(10): 03121006
- Liu S, Wei Z, Wang S (2022). On-site holographic building construction: A case study of Aurora. In: *Proceedings of the 27th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA)*. Vol. 2, 405–414
- Luhmann N, Baecker D, Gilgen P (2013). *Introduction to systems theory*. Polity Press
- Lunghi G, Marin R, Di Castro M, Masi A, Sanz P J (2019). Multimodal human–robot interface for accessible remote robotic interventions in hazardous environments. *IEEE Access: Practical Innovations, Open Solutions*, 7: 127290–127319
- Makransky G, Terkildsen T S, Mayer R E (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction*, 60: 225–236

- Mitterberger D, Ercan Jenny S, Vasey L, Lloret-Fritschi E, Aejmelaeus-Lindström P, Gramazio F, Kohler M (2022). Interactive robotic plastering: Augmented interactive design and fabrication for on-site robotic plastering. *CHI Conference on Human Factors in Computing Systems*, 1–18
- Ootsubo K, Kato D, Kawamura T, Yamada H (2016). Support system for slope shaping based on a teleoperated construction robot. *Journal of Robotics and Mechatronics*, 28(2): 149–157
- Ootsubo K, Kawamura T, Yamada H (2013). Construction tele-robotics system with AR presentation. *Journal of Physics: Conference Series*, 433: 012029
- Palmarini R, Erkoyuncu J A, Roy R, Torabmostaedi H (2018). A systematic review of augmented reality applications in maintenance. *Robotics and Computer-integrated Manufacturing*, 49: 215–228
- Pan M, Wong M O, Lam C C, Pan W (2024). Integrating extended reality and robotics in construction: A critical review. *Advanced Engineering Informatics*, 62: 102795
- Pan N H, Isnaeni N N (2024). Integration of augmented reality and building information modeling for enhanced construction inspection—A case study. *Buildings*, 14(3): 612
- Pedersen J, Neythalath N, Hesslink J, Søndergaard A, Reinhardt D (2020). Augmented drawn construction symbols: A method for ad hoc robotic fabrication. *International Journal of Architectural Computing*, 18(3): 254–269
- Pérez L, Rodríguez-Jiménez S, Rodríguez N, Usamentiaga R, García D F, Wang L (2020). Symbiotic human–robot collaborative approach for increased productivity and enhanced safety in the aerospace manufacturing industry. *International Journal of Advanced Manufacturing Technology*, 106(3-4): 851–863
- Rauschnabel P A, Felix R, Hinsch C, Shahab H, Alt F (2022). What is XR? Towards a framework for augmented and virtual reality. *Computers in Human Behavior*, 133: 107289
- Rogers Y, Sharp H, Preece J (2011). *Interaction design: Beyond human-computer interaction 3rd ed.* Wiley
- Sandagomika H, Shringi A, Mohandes S R, Kineber A F, Bazli M, Arashpour M (2024). Mixed reality-based approach for minimizing time uncertainty in prefabrication. *Journal of Construction Engineering and Management*, 150(11): 04024147
- Song Y, Agkathidis A, Koeck R (2022). Augmented bricks an onsite AR immersive design to fabrication framework for masonry structures. In: *Proceedings of the 2022 ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*. Association for Computing Machinery, NY, 14: 1–9
- Song Y, Koeck R, Luo S (2021). Review and analysis of augmented reality (AR) literature for digital fabrication in architecture. *Automation in Construction*, 128: 103762
- Speicher M, Hall B D, Nebeling M (2019). What is mixed reality? In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–15
- Suzuki R, Karim A, Xia T, Hedayati H, Marquardt N (2022). Augmented reality and robotics: A survey and taxonomy for AR-enhanced human–robot interaction and robotic interfaces. *CHI Conference on Human Factors in Computing Systems*, 1–33
- Tan T, Hall D, Papadonikolaki E, Mills G, Graser K (2024). Project delivery methods to digital fabrication in architecture: A comparative case study from a modularity perspective. In: *Routledge Handbook of Smart Built Environment*. Routledge, 46–63
- Tan T, Ng M S, Hall D (2023). Demystifying barriers to digital fabrication in architecture. *Engineering Project Organization Conference*
- Tandon A, Stührenberg J, Dragos K, Mohite I, Smarsly K (2025). BIM-based human–robot collaboration for building inspections using mixed reality
- Von Bertalanffy L (1968). *General system theory: Foundations, development, applications*. George Braziller
- Walzer A N, Kahlert A, Baumann M, Uhlmann M, Vasey L, Hall D M (2022). Beyond googly eyes: Stakeholder perceptions of robots in construction. *Construction Robotics*, 6(3–4): 221–237
- Walzer A N, Tan T, Graser K, Hall D M (2025). Bug or feature? Institutional misalignments between construction technology and venture capital. *Construction Management and Economics*, 43(2): 130–152
- Wang S, Lin D, Sun L (2023). Human-cyber-physical system for post-digital design and construction of lightweight timber structures. *Automation in Construction*, 154: 105033
- Wang X, Liang C J, Menassa C C, Kamat V R (2021). Interactive and immersive process-level digital twin for collaborative human–robot construction work. *Journal of Computing in Civil Engineering*, 35(6): 04021023
- Wei H H, Zhang Y, Sun X, Chen J, Li S (2023). Intelligent robots and human–robot collaboration in the construction industry: A review. *Journal of Intelligent Construction*, 1(1): 1–12
- Wu S, Hou L, Chen H, Zhang G, Zou Y, Tushar Q (2023). Cognitive ergonomics-based augmented reality application for construction performance. *Automation in Construction*, 149: 104802
- Wu S, Walzer A N, Kahlert A, Dillenburger B, Hall D M (2024). Understanding stakeholders’ intention to use construction robots: A fuzzy-set qualitative comparative analysis. *Construction Robotics*, 8(1): 5
- Xiang S, Wang R, Feng C (2021). Mobile projective augmented reality for collaborative robots in construction. *Automation in Construction*, 127: 103704
- Xiao Y, Watson M (2019). Guidance on conducting a systematic literature review. *Journal of Planning Education and Research*, 39(1): 93–112
- Yuen S C Y, Yaoyuneyong G, Johnson E (2011). Augmented reality: An overview and five directions for AR in education. *Journal of Educational Technology Development and Exchange*, 4(1): 11
- Zari G, Condino S, Cutolo F, Ferrari V (2023). Magic leap 1 versus microsoft hololens 2 for the visualization of 3d content obtained from radiological images. *Sensors*, 23: (6): 3040
- Zhang M, Xu R, Wu H, Pan J, Luo X (2023). Human–robot collaboration for on-site construction. *Automation in Construction*, 150: 104812
- Zulu S L, Saad A M, Gledson B (2023). Individual characteristics as enablers of construction employees’ digital literacy: An exploration of leaders’ opinions. *Sustainability*, 15: (2): 1531