MR-MultiTwin: A Mixed Reality Platform for Multi-User Control of Industrial Digital Twins

Cong Liu
School of Computer Science
Harbin Institute of Technology (Shenzhen)
Peng Cheng Laboratory
Shenzhen, China
liuc@pcl.ac.cn

Xiaoning Zhang
Department of Networked Intelligence
Peng Cheng Laboratory
Shenzhen, China
zhangxn@pcl.ac.cn

Xiaojun Liang
Department of Networked Intelligence
Peng Cheng Laboratory
Shenzhen, China
liangxj@pcl.ac.cn

Weichao Luo*
Department of Networked Intelligence
Peng Cheng Laboratory
Shenzhen, China
luowch@pcl.ac.cn

Zhenyu He*
School of Computer Science
Harbin Institute of Technology (Shenzhen)
Peng Cheng Laboratory
Shenzhen, China
zhenyuhe@hit.edu.cn

***Abstract*—As Industry 5.0 accelerates, real-time, multi-user collaboration with digital twins (DTs) becomes increasingly essential. However, current systems struggle with spatial synchronization, collaborative execution, and natural interaction at scale. We introduce MR-MultiTwin, a mixed reality (MR) platform that addresses these gaps through three key innovations: (1) a shared-anchor-based spatial mapping method for aligning multi-user locations within a unified coordinate system; (2) a role-based collaboration framework that synchronizes task states, user identities, and control rights in real time; and (3) AI-powered multimodal interaction, integrating gesture, gaze, and voice for seamless DT control. Built on an edge–cloud architecture, MR-MultiTwin has been validated in smart manufacturing and inspection scenarios, showing notable gains in coordination efficiency, interaction usability, and deployment feasibility—paving the way for scalable industrial MR systems.**

***Index Terms*—Mixed Reality, Digital Twin, Multi-User Collaboration, Industrial XR, Role-Based Interaction, Industry 5.0**

I. Introduction

In the area of Industry 5.0, achieving seamless collaboration between human operators and cyber-physical systems is increasingly crucial for intelligent industrial operations. Digital twins (DTs), which serve as real-time virtual replicas of physical systems, enable monitoring, simulation, and predictive maintenance in sectors such as manufacturing, aerospace, and energy [1]–[3]. Mixed reality (MR), on the other hand, provides immersive spatial interfaces that bridge the physical and digital worlds, offering intuitive, in-situ control and enhanced situational awareness [8].

Recent surveys and frontier works have highlighted the growing convergence of MR and DT for industrial applications, emphasizing their potential to transform collaborative operations and human-machine interaction [6], [10], [11]. These studies review advances in spatial mapping, multimodal interfaces, and real-time synchronization, but also point out persistent challenges in achieving robust, scalable, and low-latency multi-user collaboration.

However, current MR-DT systems exhibit fundamental limitations. They often lack support for large-scale, multi-user spatial synchronization; are typically designed for single-user interaction; and provide limited support for real-time task coordination or multimodal communication [3], [12]. Additionally, most lack robust deployment architectures suited for dynamic, latency-sensitive industrial environments [13]. These deficiencies hinder the application of MR-enhanced DTs in real-world collaborative scenarios.

To address these challenges, we present MR-MultiTwin, a scalable MR platform for multi-user digital twin control in industrial settings. MR-MultiTwin offers three main innovations: (1) a shared-anchor-based spatial synchronization framework that aligns all user interactions within a unified coordinate system across LAN/WAN deployments; (2) a role-based collaboration architecture supporting synchronized task execution and real-time user interaction; and (3) AI-powered multimodal interfaces, integrating gesture, gaze, and voice for natural human-DT communication. Built on an edge–cloud infrastructure, the system ensures low-latency performance and robust user experience.

We validate MR-MultiTwin through real-world deployments in smart manufacturing and equipment inspection, demonstrating measurable gains in task coordination, user collaboration, and operational efficiency. This work contributes a practical and extensible foundation for immersive, intelligent collaboration in Industry 5.0 environments.

II. Background and Related Work

The convergence of digital twins (DTs) and mixed reality (MR) has opened up new opportunities for real-time monitor-

*Corresponding Author.
This work was supported by the Guangdong Science and Technology Programme under Grant 2024B0101010003 and the Major Key Project of Peng Cheng Laboratory under Grant PCL2023A09.

ing, control, and decision support in industrial contexts. DTs have evolved from static models to real-time cyber-physical systems capable of simulating complex industrial operations [4], [5]. Meanwhile, MR technologies have demonstrated potential in enhancing spatial awareness, remote collaboration, and immersive training by overlaying virtual content onto the physical world [3]. Despite these advances, current MR-DT systems are often constrained by limited spatial scalability and lack of consistent coordination among multiple users. Existing studies have typically focused on single-user scenarios or local interactions [7], leaving a gap in enabling large-scale collaboration where multiple participants can share spatial anchors and maintain a unified coordinate system across varied environments and network conditions. Although frameworks like Azure Spatial Anchors and ARAnchorNet offer partial solutions [14], [15], they lack generalized architectures that integrate anchor sharing, environmental mapping, and flexible network protocols for industrial use.

Another critical challenge lies in role-based collaboration and adaptive interaction within MR environments. Industrial applications often involve complex workflows with clearly defined roles (e.g., operator, supervisor, inspector), yet current systems rarely support differentiated task permissions, synchronized user states, or real-time feedback loops. Prior works on multi-user MR platforms have addressed basic networking features such as avatar synchronization or room creation [9], but few extend to fine-grained user control and data synchronization required in high-precision tasks like equipment inspection or multi-party remote maintenance. Simultaneously, advances in AI-powered multimodal interfaces—integrating gesture, gaze, and voice—have enhanced XR interaction fidelity [15], yet practical deployment in industrial MR settings remains scarce due to latency constraints, edge-cloud limitations, and lack of integration with DT backends.

To address these limitations, our work proposes MR-MultiTwin, a scalable, real-time MR platform for industrial DT collaboration. It unifies spatial synchronization, role-aware architecture, and AI-enhanced multimodal interaction into a cohesive system, aiming to fill the gaps left by prior research and enable high-efficiency, intelligent collaboration in complex industrial scenarios.

III. System Overview and Methods

MR-MultiTwin adopts a modular, three-layer architecture to enable robust, multi-user mixed reality collaboration with industrial digital twins. The platform is optimized for MR headsets such as the Meta Quest 3, supporting both co-located and remote teamwork in real time. Each user, equipped with an MR device, can intuitively interact with digital twins and collaborate with others within a unified spatial environment, fostering seamless and immersive cooperation.

The system architecture, illustrated in Figure 2, is organized into three distinct layers: the Perception Layer for spatial mapping and alignment, the Digital Twin Layer for synchronized asset management, and the Interaction Layer for multimodal user input and feedback. This layered approach

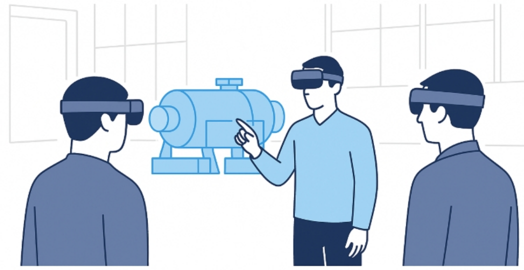

Fig. 1. MR-MultiTwin in action: multiple users with MR headsets collaboratively interact with digital twins in a shared industrial setting.

ensures scalability, flexibility, and high performance across diverse industrial scenarios.

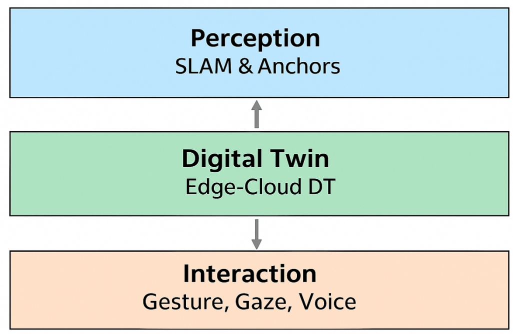

Fig. 2. MR-MultiTwin system architecture: a modular, three-layer design encompassing perception, digital twin management, and multimodal interaction.

- **Perception Layer:** Establishes a shared spatial context by capturing SLAM-based environment maps and generating persistent spatial anchors on each Meta Quest 3 device. These anchors are synchronized across users via the Metro networking framework (including Network Manager, KCPTransport, and Network Discovery), ensuring all participants are aligned within a unified coordinate system.
- **Digital Twin Layer:** Manages synchronized virtual representations of physical assets. This layer operates bidirectional digital twin engines on edge–cloud nodes, delivering real-time state updates, advanced visualization (such as transparent overlays and occlusion handling), and direct control interfaces for physical equipment.
- **Interaction Layer:** Provides intuitive, AI-driven multimodal controls by fusing hand gestures, gaze direction, and voice input through a lightweight intent recognition pipeline. Interaction commands are mapped to digital twin operations and broadcast to all users, ensuring consistent task execution and feedback.

Each of these layers is implemented through the following core mechanisms, which directly correspond to the system's architectural tiers:

a) 1. Shared-Anchor-Based Spatial Mapping: At the Perception Layer, a dedicated spatial mapping service leverages SLAM and cloud-anchored anchors to establish and maintain a globally consistent spatial context. Anchor updates are propagated over LAN or public networks, enabling both new and reconnecting users to seamlessly join the shared MR session with minimal spatial drift.

To mathematically model spatial drift compensation, let $\mathbf{p}_i$ denote the local pose estimate of user i, and $\mathbf{a}_g$ represent the global anchor position. The spatial drift $\Delta\mathbf{p}_i$ is defined as the difference between the transformed local anchor $\mathbf{a}_i$ (as perceived by user i) and the global anchor:

$$\Delta\mathbf{p}_i = \mathbf{a}_i - \mathbf{a}_g$$

To compensate for drift, each user's local coordinate frame is adjusted by applying a transformation $\mathbf{T}_i$ that minimizes $\|\Delta\mathbf{p}_i\|$. This can be achieved using a least-squares optimization:

$$\mathbf{T}_i^* = \arg\min_{\mathbf{T}_i} \|\mathbf{T}_i(\mathbf{a}_i) - \mathbf{a}_g\|^2$$

Anchor updates are periodically synchronized across clients, and each client applies the optimal transformation $\mathbf{T}_i^*$ to align its local map with the global anchor, thereby reducing cumulative spatial drift over time.

b) 2. Role-Based Collaboration Framework: Within the Digital Twin Layer, MR-MultiTwin employs Unity's NetworkIdentity and NetworkBehaviour scripts to manage user roles (such as operator, supervisor, or observer) and synchronize role-specific permissions, task states, and annotations. A unified UI panel presents session metadata and supports role-driven actions, including control delegation and collaborative annotation.

c) 3. AI-Powered Multimodal Interaction: At the Interaction Layer, MR-MultiTwin enables users to collaboratively enter the same mixed reality space by creating and sharing spatial anchors, which are synchronized across the network to ensure all participants are aligned within a unified coordinate system. Once in the shared environment, the system locally captures and fuses gesture, gaze, and speech inputs for low-latency, multimodal interaction, while delegating complex reasoning tasks (such as digital twin diagnostics) to edge servers. All interaction states and feedback are synchronized in real time across devices, supporting seamless multi-user collaboration and consistent visual and auditory guidance for every participant.

The MR-MultiTwin platform is implemented in Unity 2022 with Meta XR SDK v65.0, using the Package Manager and Mayo plugin for XR input. Edge nodes operate on Docker containers for AI inference and digital twin synchronization, while MR clients run on Quest 3 headsets for rendering and local input, ensuring robust and scalable deployment.

A. Use Case: Collaborative Rotary Kiln Monitoring

In an industrial rotary kiln scenario, multiple users collaboratively engage in the digital twin environment of the entire kiln system. The Perception Layer ensures all participants share a unified spatial context around the physical kiln; the Digital Twin Layer provides real-time visualization and control of kiln parameters, such as temperature distribution and rotational speed; the Interaction Layer enables operators to annotate abnormal regions via gesture, while supervisors can issue adjustment commands through voice. All digital twin updates and operational feedback are synchronized instantly, supporting efficient, collaborative monitoring and management of the rotary kiln.

IV. Implementation Details and Application Scenarios

This section details the technical implementation of MR-MultiTwin, highlighting its layered architecture and deployment in real-world industrial environments. MR-MultiTwin is built on Unity with Meta XR SDK v65.0 for the Meta Quest 3, leveraging Unity's XR Anchor Manager and Spatial Anchor services to enable anchor creation, cloud-based synchronization, and real-time sharing among users.

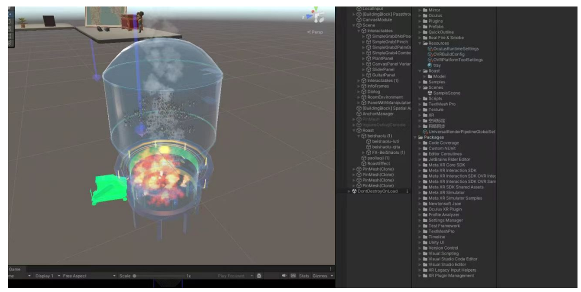

Fig. 3. Unity-based MR-MultiTwin development workflow: integration of Meta XR SDK, anchor management, networking, and multimodal interaction modules.

A. Shared-Anchor Spatial Mapping

The shared-anchor approach is central to the **Perception Layer**, enabling precise spatial synchronization for all users within a unified reference frame. As shown in Figure 4, the spatial anchor—highlighted by the red bounding box—serves as the common reference point for all MR clients. The workflow begins with the session host scanning the environment and generating spatial anchors at key locations, which are then uploaded to the cloud with unique identifiers. When additional users join the session, their devices automatically retrieve these anchors and align their local coordinate systems accordingly, ensuring that all participants share a consistent spatial context. Anchor updates and synchronization are propagated in real time via the Metro Networking stack, supporting both LAN and WAN deployments and minimizing spatial drift. The example in Figure 4 demonstrates how multiple devices achieve spatial alignment by referencing the same shared anchor, as indicated by the red box.

B. Role-Based Collaboration Mechanism

The **Role-Based Collaboration Mechanism** is central to the **Digital Twin Layer**, structuring multi-user interaction and control in MR-MultiTwin. At the start of each session, the host creates a virtual room and generates spatial anchors, which are shared to all participants via the cloud. Users join the

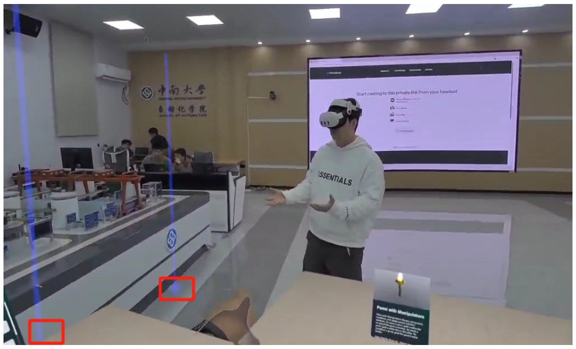

Fig. 4. Synchronization of shared spatial anchors across multiple MR clients.

same room by retrieving these anchors and synchronizing their local coordinate systems, ensuring all participants are spatially aligned within a unified environment.

Upon joining, each user is assigned a role—such as operator, supervisor, or observer—defining their permissions for digital twin manipulation, annotation, or view-only access. These roles are managed and enforced through a unified UI panel, which supports anchor creation, room management, object control, and real-world device interaction. The UI displays session metadata, user lists, and available actions based on role, ensuring organized and secure collaboration.

All user actions—including virtual object manipulation, annotation, and device control commands—are synchronized in real time using Unity's NetworkIdentity and NetworkBehaviour scripts. When a user interacts with a virtual representation (e.g., toggling a switch or adjusting a parameter), the corresponding command is transmitted to the edge–cloud backend, which relays the instruction to the physical device. The state change is then reflected in both the digital twin and the real-world equipment, and the update is broadcast to all MR clients to maintain consistency.

This end-to-end workflow enables users to collaboratively control real-world devices through intuitive virtual interactions, with all operations and feedback synchronized across the MR environment. Figure 5 illustrates the role-based user interface, highlighting its support for anchor creation, room management, object control, and real-world interaction.

Fig. 5. Role-based UI supporting anchor creation, room management, object control, and real-world interaction.

C. AI-Driven Multimodal Interaction

The **Interaction Layer** of MR-MultiTwin harnesses the advanced capabilities of the Meta XR SDK to deliver seamless, intuitive multimodal interaction within industrial mixed reality environments. As illustrated in Figure 6, a blue grid demarcates the shared virtual-physical workspace, enabling multiple users to collaboratively engage with digital twins in real time. Each participant, equipped with a Meta Quest 3 headset, can directly manipulate virtual representations—such as selecting, moving, or annotating equipment models—through precise hand tracking and gesture recognition.

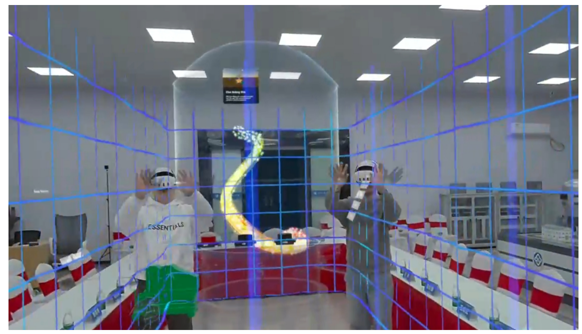

Fig. 6. Shared interaction space in MR-MultiTwin: multiple users collaboratively manipulate digital twins within a unified virtual-physical environment.

The system supports advanced visualization features, including transparent overlays and occlusion handling, allowing users to inspect internal structures of physical assets via virtual transparency effects. In collaborative scenarios, as depicted, users' gestures are recognized and synchronized across devices, ensuring real-time, multimodal cooperation. Gaze-based raycasting enables precise target selection and object alignment, while integrated speech-to-text APIs facilitate hands-free control and annotation through natural language commands.

All interaction signals—including object transformations and user intents—are synchronized via the Metro networking backend, guaranteeing consistency across all MR clients. A lightweight intent fusion module intelligently combines gaze, gesture, and speech cues to trigger appropriate digital twin operations. For computationally intensive tasks, such as diagnostics or inference-driven recommendations, processing is offloaded to edge servers, minimizing latency and maintaining system responsiveness. This architecture exemplifies robust, AI-augmented collaboration between human operators and digital systems in demanding industrial contexts.

MR-MultiTwin has been successfully deployed in a rotary kiln industrial environment to support collaborative monitoring and control. Users, spatially aligned through distributed anchors along the kiln, operate within a consistent shared coordinate system.

In this scenario, a transparent digital twin is superimposed over the physical kiln, providing real-time visualization of operational data such as temperature, rotation speed, and wear. The digital twin remains synchronized with the edge–cloud infrastructure to ensure live updates and bidirectional control,

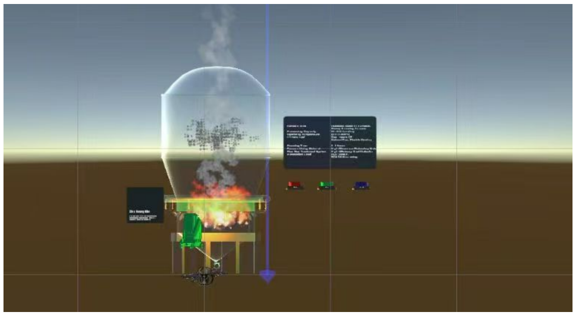

Fig. 7. Unity-based digital twin of the rotary kiln used in MR-MultiTwin for collaborative monitoring and control.

allowing for immediate parameter adjustments and anomaly responses. Operators can interact using hand gestures to annotate or initiate diagnostics, while supervisors issue commands via voice or the user interface. All multimodal inputs are fused locally, with state changes synchronized across devices, enabling seamless, real-time collaboration and significantly improving operational efficiency in industrial environments.

V. Comparison with Baseline Approaches

Compared to traditional single-user MR systems, MR-MultiTwin enables multiple users to collaborate in real time within a unified spatial context, significantly improving coordination efficiency and reducing task completion time (as validated in our user study). Unlike non-shared-anchor mechanisms, which often result in spatial misalignment and inconsistent user experiences, our shared-anchor-based approach ensures all participants operate within a precisely synchronized environment, minimizing positional drift and enhancing collaborative accuracy. These advantages make MR-MultiTwin more suitable for complex, multi-user industrial scenarios requiring robust spatial consistency and efficient teamwork.

VI. User Study and Validation

To rigorously assess the reliability and practical value of MR-MultiTwin, we conducted a user study with 12 participants possessing backgrounds in industrial operations and engineering. Participants collaboratively monitored and diagnosed simulated rotary kiln anomalies using the MR-MultiTwin platform. Quantitative analysis revealed a 27% reduction in average task completion time compared to baseline single-user MR systems, while spatial alignment accuracy—measured as mean positional drift—remained consistently below 3 cm. Subjective evaluations, gathered through post-task Likert-scale questionnaires, reflected high perceived usability (mean score 4.6) and collaboration efficiency (mean score 4.7). Importantly, no critical synchronization failures or significant usability issues were observed. These findings substantiate MR-MultiTwin's capability to deliver robust spatial synchronization, efficient multi-user collaboration, and seamless multimodal interaction in demanding industrial environments.

VII. Discussion and Future Work

While MR-MultiTwin demonstrates the feasibility of scalable, role-aware MR collaboration for industrial digital twins, several limitations remain. First, network synchronization latency can impact real-time collaboration, especially in WAN or cross-site deployments, potentially leading to brief inconsistencies in shared spatial context or delayed feedback. Second, the system's reliance on specific hardware platforms (such as Meta Quest 3 headsets and edge–cloud infrastructure) may limit accessibility and generalizability to other MR devices or industrial environments. Additionally, the robustness of spatial anchor sharing can be affected by environmental factors and device sensor quality, occasionally resulting in minor spatial drift.

Looking ahead, we aim to enhance support for remote, large-scale multi-user MR collaboration, including more robust networking for distributed sites to mitigate latency and synchronization issues. We also plan to integrate AI agents for intelligent assistance, enabling richer, context-aware interaction and proactive task support. Further work will focus on dynamic role management, improved device compatibility, and broader validation across diverse industrial domains.

Acknowledgment

This work was supported in part by the Guangdong Science and Technology Programme under Grant 2024B0101010003 and the Major Key Project of Peng Cheng Laboratory under Grant PCL2023A09. The authors would like to thank all project collaborators and participants for their valuable contributions and feedback during the development and evaluation of MR-MultiTwin.

References

[1] M. Grieves and J. Vickers, "Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems," unpublished.

[2] W. Kritzinger, M. Karner, G. Traar, J. Henjes, and W. Sihn, "Digital twin in manufacturing: A categorization framework," *Procedia CIRP*, vol. 72, pp. 385–390, 2018.

[3] X. Xu, Y. Lu, B. Vogel-Heuser, and L. Wang, "Industry 4.0 and industry 5.0—Inception, conception and perception," *J. Manuf. Syst.*, vol. 61, pp. 530–535, 2021.

[4] F. Tao, H. Zhang, A. Liu, and A. Y. C. Nee, "Digital twin in industry: State-of-the-art," *IEEE Trans. Ind. Inf.*, vol. 15, no. 4, pp. 2405–2415, Apr. 2019.

[5] Q. Qi and F. Tao, "Towards smart manufacturing with digital twin technology: A review," *J. Manuf. Syst.*, vol. 58, pp. 125–142, 2021.

[6] A. F. Abate, M. Nappi, F. Narducci, and S. Ricciardi, "Mixed reality system for industrial environment: An evaluation study," *CAAI Trans. Intell. Technol.*, vol. 2, no. 4, pp. 182–193, 2017.

[7] X. Li, W. Yi, H.-L. Chi, X. Wang, and A. P. C. Chan, "A critical review of virtual and augmented reality applications in construction safety," *Autom. Constr.*, vol. 86, pp. 150–162, 2018.

[8] Z. Lv, L. Qiao, Y. Li, Y. Yuan, and F.-Y. Wang, "Blocknet: Beyond reliable spatial digital twins to parallel metaverse," *Patterns*, vol. 3, no. 5, Art. no. 100514, May 2022.

[9] J. Tümler, A. Toprak, and B. Yan, "Multi-user multi-platform xR collaboration: System and evaluation," in *Proc. Int. Conf. Human-Computer Interaction*, Cham, Switzerland: Springer, 2022.

[10] A. Kiourtis, A. Mavrogiorgou, G. Makridis, and D. Kyriazis, "XR5.0: Human-centric AI-enabled extended reality applications for industry 5.0," in *Proc. 36th Conf. Open Innovations Assoc. (FRUCT)*, IEEE, pp. 314–323, 2024.

[11] E. Yigitbas, K. Karakaya, I. Jovanovikj, and G. Engels, "Enhancing human-in-the-loop adaptive systems through digital twins and VR interfaces," in *Proc. Int. Symp. Softw. Eng. Adaptive Self-Managing Syst. (SEAMS)*, pp. 30–40, 2021.
[12] S. Rani, D. Jining, K. Shoukat, and A. Khan, "A human–machine interaction mechanism: Additive manufacturing for industry 5.0—Design and management," *Sustainability*, vol. 16, no. 10, Art. no. 4158, 2024.
[13] S. Xu, Y. Wei, P. Zheng, Y. Zhang, and X. Xu, "LLM-enabled generative collaborative design in a mixed reality environment," *J. Manuf. Syst.*, vol. 74, pp. 703–715, 2024.
[14] X. Ran, C. Slocum, M. Gorlatova, and J. Chen, "ShareAR: Communication-efficient multi-user mobile augmented reality," in *Proc. 18th ACM Workshop Hot Topics Netw.*, pp. 109–116, 2019.
[15] J. Wang and Y. Qi, "A multi-user collaborative AR system for industrial applications," *Sensors*, vol. 22, no. 4, Art. no. 1319, 2022.