A Systematic Review of XR-Enabled Remote Human-Robot Interaction Systems

XIAN WANG, The Hong Kong Polytechnic University, Hong Kong, Hong Kong
LUYAO SHEN, The Hong Kong University of Science and Technology - Guangzhou Campus, Guangzhou, China
LIK-HANG LEE*, The Hong Kong Polytechnic University, Hong Kong, Hong Kong

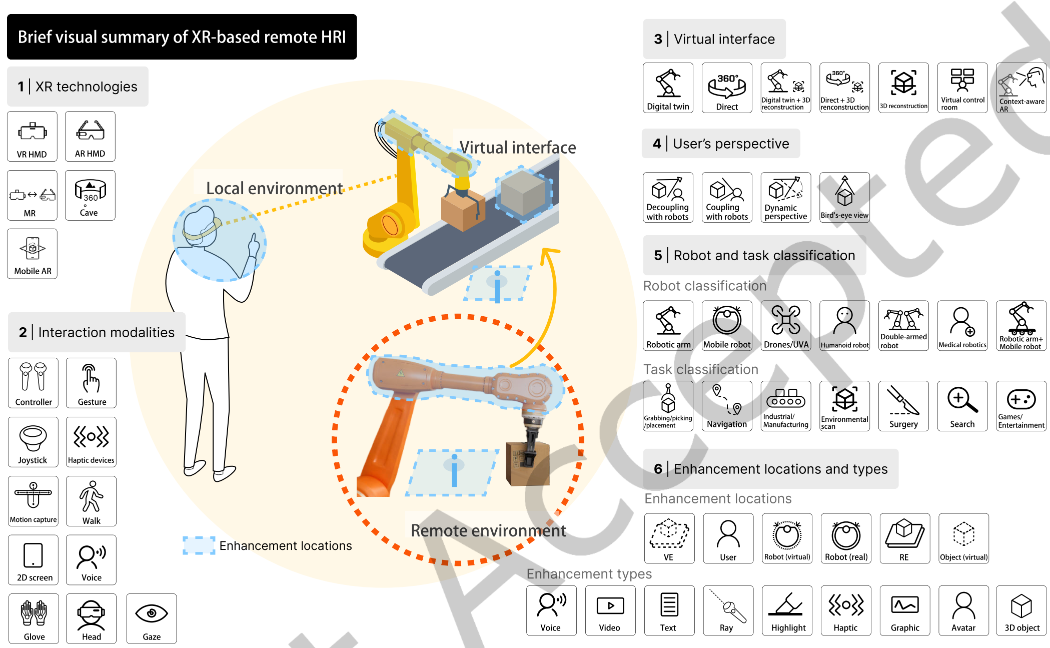

Fig. 1. Brief visual summary of the current state of the domain concerning XR-based remote HRI, summarizes the six key dimensions of current system design.

The rising interest in creating versatile robots to handle multiple tasks in various environments, with humans interacting through immersive interfaces. This survey provides a comprehensive review of extended reality (XR) applications in remote

*L.-H. Lee is the corresponding author.

Authors' Contact Information: Xian Wang, The Hong Kong Polytechnic University, Hong Kong, Hong Kong; e-mail: xiann.wang@connect.polyu.hk; Luyao Shen, The Hong Kong University of Science and Technology - Guangzhou Campus, Guangzhou, Guangdong, China; e-mail: lshen595@connect.hkust-gz.edu.cn; Lik-Hang Lee, The Hong Kong Polytechnic University, Hong Kong, Hong Kong; e-mail: lik-hang.lee@polyu.edu.hk.

ACM 1557-7341/2025/4-ART
https://doi.org/10.1145/3730574

human-robot interaction (HRI). We developed a systematic search strategy based on the PRISMA methodology, focusing on peer-reviewed publications that demonstrate practical implementations of XR in remote robot control, real robot system deployment, and HRI applications, we analyzed research published between January 2013 and December 2023. From the initial 2,561 articles, 100 met our inclusion criteria were included. We categorized and summarized the domain in detail, delving into the methods used in these articles to achieve intuitive and effective remote HRI, highlighting user experience enhancement and interaction designs. This survey identifies research opportunities, particularly emphasizes that future researchers should explore the potential of XR, such as exploring multimodal enhancement techniques that seamlessly integrate visual, haptic, and auditory feedback for more intuitive teleoperation. Our analysis reveals that while XR shows promising potential in remote HRI, there are significant gaps, such as user-centered design. This survey provides a framework for understanding the current state of XR-based remote HRI, establishing a foundation for future research.

CCS Concepts: • **Human-centered computing** → **Virtual reality**; **Mixed / augmented reality**; **Collaborative interaction**; *Ubiquitous and mobile devices.*

Additional Key Words and Phrases: Human-robot interaction, Extended Reality, Virtual Reality, Augmented Reality, Teleoperation, Remote collaboration

1 Introduction

The field of Human-Robot Interaction (HRI), originally perceived as the study of human engagement with robots, has evolved to investigate the design of robots for socially meaningful interactions with humans and the enhancement of these interactions [128]. HRI aims to engineer robots capable of effectively collaborating with humans across diverse contexts, such as industrial manufacturing [42, 91], domestic settings [56], and educational institutions [31]. In our increasingly globalized and digitalized world, where advances in communication and mobile technologies are driving the rapid development of collaboration concepts, traditional face-to-face interactions are being complemented, and, in certain instances, replaced by remote collaboration mechanisms that overcome geographical and temporal constraints. Remote human-robot collaboration holds considerable potential to revolutionize various fields. For instance, it can facilitate tasks in hazardous environments by allowing humans to control robots from a safe distance [86], or it can enhance complex tasks by enabling robots to relay information from remote locations to humans, such as remote robotic surgery [163], or aerospace [114]. The implications of this form of collaboration include safer workplaces, improved efficiency, and greater accessibility.

Despite videoconferencing and teleconferencing serving as effective tools for remote collaboration, these technologies exhibit limitations when applied to complex tasks within the context of remote human-robot collaboration. For instance, studies have shown that traditional video interfaces often lack the spatial and contextual awareness necessary for intuitive robotic control in dynamic or intricate environments [62]. Such limitations reduce the ability of users to perform and monitor tasks, and are especially pronounced in scenarios involving robotic teleoperation or remote instruction. To overcome these limitations, Extended Reality (XR), encompassing Augmented Reality (AR), Virtual Reality (VR) and Mixed Reality (MR) [116], offers a promising solution. AR overlays digital information onto the real world, VR immerses users in a completely digital environment, and MR blends real and virtual worlds. These technologies provide a fusion of the digital and physical worlds, allowing physical and digital objects to coexist and interact in real time [82]. XR technology is becoming increasingly accessible and portable such as commercial devices like Meta Quest 3 [1] and Microsoft HoloLens 2 [2]. With the help of these XR devices, novices can perform risky tasks (e.g., welding [160]) in a safe space and in a more intuitive way (e.g., with the first view of the robot [126]). In addition, operators can use XR to switch between different locations to control the robot without having to move in physical space [69]. XR empowers remote HRI to be more immersive, intuitive, and effective.

[1]https://www.meta.com/sg/quest/quest-3
[2]https://learn.microsoft.com/en-us/hololens

However, numerous challenges need to be addressed to unlock XR's full potential in the context of remote human-robot interaction. These challenges include designing intuitive interaction techniques, reducing remote control latency, and evaluating remote HRI systems. Given that this remains an under-researched topic, this article seeks to contribute to this emerging field of study. We provide a systematic literature review of XR technology-based remote human-robot collaboration applications, highlight key advancements, and identify areas that deserve to be studied in further research. The contributions of this survey are as follows: **(1)** Provide a comprehensive review of HRI system design based on XR technology for remote control scenarios. **(2)** Identify general convergences and divergences in system design within the existing literature. **(3)** Propose a research agenda for future XR-based remote HRI.

1.1 Overview of XR-based Remote Human-Robot Interaction

Figure 1 provides a brief pictorial summary of the current state of the domain concerning XR-based remote HRI. We establish two distinct spaces: the '**local space**' and the '**remote space**', the survey scope is highly relevant to these two spaces. Thus, we give their definitions as follows.

Local Space. 'Local space' refers to the physical environment where the user is located, often equipped with XR technologies like AR or VR headsets to overlay or immerse the user in a virtual environment (see Fig.1 left of the center schematic). This space is crucial for the user's interaction with the robot, as it hosts the virtual interface for controlling the robot. For example, in the local space, a user wearing a VR headset may see a robotic digital twin (see Fig.1 top right of the center schematic) in the virtual space. This virtual interface allows the user to issue commands to the robot located in another space through various interaction modalities, such as gesture control. The local space is designed to be intuitive and user-friendly, allowing the user to manipulate the robot to perform complex tasks without having to be physically present in the same space as the robot.

Remote Space. Conversely, 'remote space' is the physical environment where robots operate and perform tasks. This could be a factory handling hazardous materials, a distant planet, or a complex surgical field where human presence is impossible or undesirable. The robot acts as an agent for the user, performing tasks from the local space via a virtual interface. For example, in a manufacturing facility, a robotic arm is responsible for handling hazardous chemicals on a conveyor belt under the guidance of an operator in local space (see Fig.1 lower right of the center schematic). The remote space is characterized by its task-oriented property, utilizing the physical capabilities of the robot to perform specific actions that benefit from or require teleoperation.

To provide a comprehensive understanding of the two space types, we synthesized literature across several dimensions of system design: 1) **XR technologies** in remote HRI, exploring VR, AR, and MR (VAM) for bridging local and remote spaces with intuitive interfaces. 2) **Interaction modalities** between the user and the virtual interface, such as gesture control, affects the efficiency and ability of the user to operate the virtual interface and perform tasks remotely. 3) Design of **virtual interfaces** to present remote information in the local space. 4) **User perspectives** for observing robots, focus on how users perceive and understand the motion and remote environment of the robot through a virtual interface. 5) **Robot and specific tasks classification**, crucial for customizing XR interfaces. 6) **Enhancement locations and types** of the multimodal elements, integrating visual, auditory, and haptic feedback to enhance control and perception.

1.2 Existing Surveys

Previous research has conducted separate investigations into XR technology, remote collaboration, and HRI. However, the intersection of these three dimensions, particularly remote HRI, has remained largely unexplored. Notably, Schafer et al. [121], and Wang et al. [157] have conducted reviews of remote collaboration systems using XR technology. Schafer et al. emphasized synchronous remote collaboration systems, whereas Wang et al. concentrated on physical tasks. However, their main focus is human-to-human remote collaboration instead of human-robot interaction. Moreover, when analyzing HRI systems or collaborative robots, the majority of researchers have primarily explored the application of AR technology [10, 32, 36, 57, 89, 109, 136]. The study by Dianatfar et al. [37] encompasses VR technology but only synthesizes VR simulation applications for surgical robots and does not adequately consider interaction scenarios between humans and tangible robots. Walker et al. [153] proposed a taxonomy for HRI systems using XR technology, but their primary focus was not HRI in remote contexts.

In reviewing the existing literature, we notice that none of the existing survey articles systematically categorize and synthesize the usage of XR technologies in remote HRI settings. In contrast, our article addresses the gaps in the current research, including all XR technologies, and dives into the issue of HRI in remote contexts. This systematic review can serve as the first comprehensive guide for researchers to situate their work within a broader framework and explore innovative systems for XR-based remote HRI.

1.3 Motivation and Research Questions (RQs)

This study aims to provide a comprehensive understanding of XR-based remote HRI by analyzing relevant literature. The investigation categorizes existing system designs based on several dimensions (see section 1.1), objective is to explore how these dimensions can be utilized to create immersive, efficient, and user-centered XR-based remote HRI systems. Specifically, the survey addresses the following RQs:

- *RQ1:* What types of XR devices are used in remote HRI systems and what interaction modalities do users use? (*DE4-5*)
- *RQ2:* What robots have been used and what are their capabilities and functions? (*DE8-9*)
- *RQ3:* How are XR-based remote HRI systems evaluated and what gaps in current evaluation methodologies need further research? (*DE10*)
- *RQ4:* How does XR technology augment Remote HRI on virtual content? (*DE6-7, DE13-14*)
- *RQ5:* Does the existing remote HRI system support multi-player/multi-robot interaction? (*DE12*)
- *RQ6:* What are the pending issues with the current Remote HRI system? (*DE11*)

To address these RQs, we perform rigorous data extraction (DE, see section 2.3 for details) and, for ease of understanding, we labeled the data extraction identifiers after the RQs first.

1.4 Structure of the Survey

The remainder of this article is structured as follows: Section 2 thoroughly explains the methodology employed for this survey. Section 3 offers an in-depth discussion and analysis of the included articles, specifically focusing on the techniques, types, and tasks utilized by remote robots, task evaluation, the role of XR techniques, multiplayer/robot support, and system latency. This analysis is based on our developed taxonomy and data extraction rules. In Section 4, a detailed discussion and analysis are presented on the development and recent advances in remote HRI based on XR. Subsequently, Section 5 discusses the grand challenges and suggests potential future research directions. Finally, Section 6 provides a comprehensive summary of the entire survey article.

2 Methodology

To ensure methodological robustness and transparency in our literature review process, we used the preferred reporting element for systematic reviews and meta-analysis (PRISMA) framework, as recommended by Takkouche et al. [139]. The review used an online tool named **Covidence**[3], an collaborative tool for systematic literature reviews that can help streamline the process. It can automatically merge duplicates after reference import and support both title/abstract and full-text screening by multiple reviewers. In our study, the first and second authors conducted the review together and resolved any conflicting screening results through discussion. Complete

[3]https://www.covidence.org/

PRISMA results can be found in Figure 2. Upon completing our search, we began a structural process of filtering through 2,588 articles. We initially removed 27 duplicate articles, leaving us with a pool of 2,561 articles to examine. We screened these articles based on their titles and abstracts, leading to the exclusion of 2,216 articles that did not meet our criteria. Following a full-text evaluation, an additional 245 articles were further excluded. The specific inclusion and exclusion criteria are described in Section 2.2. Our literature review process was conducted in two phases to capture the most current research. The first round of data extraction took place in May 2022, followed by a second round in December 2023. Ultimately, we selected 100 articles for data extraction and further analysis in our survey.

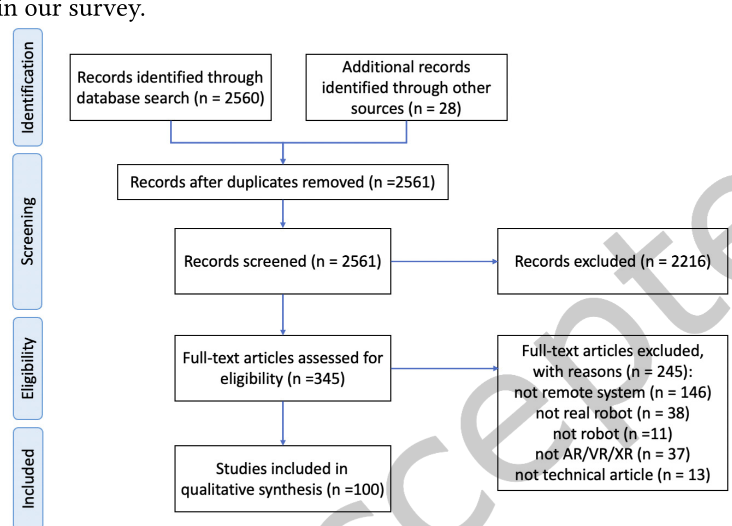

Fig. 2. Systematic review process using PRISMA.

2.1 Search Strategy

Our search strategy was iterative. Initially, the first author proposed a set of search terms based on relevant literature, which were refined with the second and third authors. We then invited two robotics researchers to verify comprehensive coverage. Throughout the review, we systematically documented keywords from identified articles, adding any missing ones and re-running searches until no new articles were found. A similar iterative process was applied to the inclusion and exclusion criteria. The first author developed these criteria from the relevant literature and the purpose of the survey. The first and second authors then independently screened articles on *Covidence*, reviewing each item in duplicate. In cases of ambiguity, screening paused, and all authors worked together to clarify and refine the criteria until consensus was reached. Any modifications to the criteria triggered a re-screening of previously reviewed articles. Finally, all authors resolved conflicts in the screening results through discussion.

2.1.1 Keywords. In developing our literature search strategy, we focused on three core dimensions of XR-enabled Remote HRI System: **remote operations**, **HRI**, and **immersive VAM technology**. Our research questions are focused on understanding how these dimensions intersect to influence the design, implementation, and evaluation of remote HRI systems. Consequently, we selected keywords that comprehensively capture each dimension. For **remote operations**, taking into account the terminological expressions in different contexts, we selected the keywords “distributed”, “remote”, “teleoperation”, “telerobotics”, “telepresence” and “spatial”, which reflect users interact with robots at a distance. To cover the **HRI** aspect, we included “robotic”, “robot”, “machine”, “human-robot interaction”, “human-robot collaboration”, “interaction”, “collaboration”, “cooperation”, “collaborate” as well as the abbreviations “HRI” and “HRC”, Lastly, to address **immersive VAM technology**, we searched for "virtual reality", "augmented reality", "mixed reality", "vr", "ar" and "mr". In addition to that, in each selected database (see Section 2.1.2), we employed the Advanced Search functionality to search the *Title* and *Abstract* fields, applying a combination of Boolean operators (e.g., (collaboration OR cooperation OR interaction) AND (virtual OR VR OR augmented OR AR) AND (robot OR robotics OR machine) AND (remote OR teleoperation OR telepresence OR "long-distance")).

2.1.2 Databases. We searched for the most relevant articles available on several publication databases, including ACM Digital Library, IEEE Xplore, and ScienceDirect (Elsevier). To supplement this search, we also conducted a snowball search on Google Scholar, which involved reviewing the reference lists of relevant articles to identify additional publications. Additionally, we utilized the **Connected Papers**[4] online tool to identify related articles and broaden our search until no new relevant articles appeared. This approach allowed us to thoroughly explore the relevant literature and identify key publications related to our research question.

2.2 Inclusion and Exclusion Criteria

To address RQs, we focused our search on studies that (a) used VAM technology to control real-world robots, virtual or simulated robots are not included, besides that, (b) the user and the robot need to be in two spaces, Local and Remote (see Section 1.1 for details), to ensure that the study is applied to remote control. We carefully developed the appropriate inclusion and exclusion criteria to guide our selection process and summarized in Table 1. Our survey focused on articles that met all inclusion criteria, while excluded any articles that met at least one of the exclusion criteria.

Table 1. Inclusion and Exclusion Criteria

Criterion	Description
I_1	The research proposes the design, development or system for remote control of robots based on VAM technology
I_2	The study used VAM technology as a remote control method for the robots
E_1	The system proposed in the study does not support remote control (i.e., based on our definition above, the user and the robot are not in two spaces)
E_2	The study does not involve real robots (e.g, robots were simulated in virtual environments or tests were not explicitly made on a realistic robot)
E_3	VAM technology is not used as a control input for the robots (e.g., use of VAM technology is to facilitate the robot programming, which is not included in this survey scope)
E_4	The study did not use VAM technology
E_5	This is not a technical article (e.g., literature review, survey, book chapters)
E_6	Studies reported in a language other than English

In addition, we provide additional clarification here for the criteria E_1 and E_3, which may cause ambiguity. For critera E_1, consider a study that uses a VAM interface to control a UAV. If the system relies exclusively on the spatial reference of the physical UAV, any line-of-sight occlusion between the user and the UAV will make control impossible. This situation would then conflict with our scope, which requires the user and the robot to operate in physically different spaces. For example, the study by Walker et al. [152] was excluded because they used AR to visualize the robot's intentions, but the study could only rely on UAV entities to provide spatial references. For critera E_3, As an illustrative example, consider the study by Cao et al. [21] Although it incorporates VAM technology, VAM is used primarily to facilitate robot programming. In this case, the user uses AR to annotate or leave programming references rather than directly control the robot's movement or behavior. Consequently, this type of article does not align with our focus on remote HRI control and was therefore excluded from the survey.

[4]https://www.connectedpapers.com/

2.3 Data Extraction and Analysis

To extract relevant information from the included articles, we developed a data extraction rubric. Initially, the first author of this survey selected 10 articles in a pseudo-random manner ([8, 13, 59, 69, 78, 144, 154, 156, 171, 185]), and developed items based on the relevant aspects identified in the articles. The initial data extraction rubric was then evaluated by all authors and refined into the final version described in Table 2. The first and second authors independently extracted data from each article using this finalised rubric. In cases of conflicting data, consensus was reached through discussions among the authors.

Data extraction items **DE1-DE3** pertain to general descriptors of the article, including study number (author and date), title, and keywords. XR technologies (**DE4**) encompass common types of augmented, virtual, and mixed reality technologies, such as VR HMD, AR HMD. Interaction modalities (**DE5**) highlight the variety of hardware and their applications in virtual environments. The virtual interface (**DE6**) allows users to operate the robot more intuitively with a virtual interface. It may include a direct interface or a digital twin of the remote robot. The user's perspective (**DE7**) refers to the user's viewpoint, such as observing the robot from a detached third view or a top-down "bird's-eye view". Generic types of robots are described under **DE8**, while specific tasks (**DE9**) that can be performed by the human-robot collaboration system are of interest, as different tasks might require various robots or telepresence designs. The system may have been enhanced at different locations (**DE13**) with multimodal enhancements (**DE14**), e.g., haptic, video, 2D or 3D overlays, etc. In addition, HRI can include multiplayer or robot collaboration (**DE12**), necessitating distinctions between one-to-one, one-to-multi, or multi-to-multi collaboration, which could influence the study design. The evaluation method of the study (**DE10**) is also essential; it could be quantitative or qualitative, including the potential presence of delays (**DE11**) in remote controls.

Table 2. Data Extraction Rubric for the Selected 100 Articles

ID	Data Extraction	Type
DE1	Study ID	Open text
DE2	Title	Open text
DE3	Keywords	Open text
DE4	Used XR technologies	VR HMD, AR HMD, MR, Mobile AR, CAVE, Other
DE5	Interaction modalities	Gesture, Controller, Joystick, Gaze, Head, Haptic devices, Motion capture, Walk, 2D screen, Voice, Glove, Other
DE6	Virtual interfaces	Direct, Digital twin, Virtual control room, Digital twin+3D reconstruction, Direct+3D reconstruction, 3D reconstruction, Context-aware AR, Multiple, Other
DE7	User perspective	Coupling with robots, Decoupling from robots, Dynamic perspective, Bird's-eye view, Other
DE8	Type of robots	Mobile robot, Drones/UAV, Humanoid Robot, Robotic arm, Mobile robot+robotic arm, Double-armed robot, Medical Robotics, Other
DE9	Specific tasks involved	Navigation, Grabbing/Picking/Placement, Surgery, Game/Entertainment, Industrial/Manufacturing, Search, Environment scan, No, Multiple, Other
DE10	How was it measured or evaluated?	Time/accuracy of the task, Interviews, Questionnaire, AR/VR performance, Comparison, N/A, Other
DE11	Was there a discussion about delays?	No, Yes(Times), Other
DE12	Support multiplayer collaboration?	No, Multi-user - one robot, One user - multi-robot, Multi - Multi, Other
DE13	Where are the enhancements located?	User, Robot (real), Robot (virtual), Object (virtual), Real environment(RE), Virtual environment(VE)
DE14	Enhancement types	Haptic, Voice, Graphic, Text, 3D Object, Highlight, Ray, Avatar

Due to the substantial number of included studies and the diversity of their study designs, conducting a meta-analysis or a unified data synthesis to determine the effect measures was not feasible. Therefore, we adopted a narrative synthesis approach, which allowed us to summarize and interpret the findings in relation to each research question. In addition to this, we used data charts to present the data results for each research question, and reported the results with a detailed list of the percentage of categories coded in order to provide the most accurate representation of the current state of the literature in the field.

3 Results and Descriptive Statistics

3.1 Overview of Included Articles

We meticulously extracted pertinent information from the 100 articles identified during our screening. The selected articles span from 2013 to 2023, and the annual publication count is illustrated in Figure 3. These articles are from well-known venues for HRI and human-computer interaction (HCI), including the IEEE International Conference on Intelligent Robots and Systems (IROS), ACM/IEEE International Conference on Human-Robot Interaction (HRI), IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), and the ACM Symposium on User Interface Software and Technology (UIST). These data were subsequently analyzed and summarized both statistically and graphically, with additional qualitative insights emerging during the iterative analysis. A comprehensive list of data extracts for the included articles can be found in Appendix A. Over the past decade, there has been a general increase in the number of articles published on the investigated topics, indicating growing interest and significance in recent years. This trend could be attributed to the landscape of XR technology maturing, and it is worth noting that the volume of publications in this field reached its zenith in 2023.

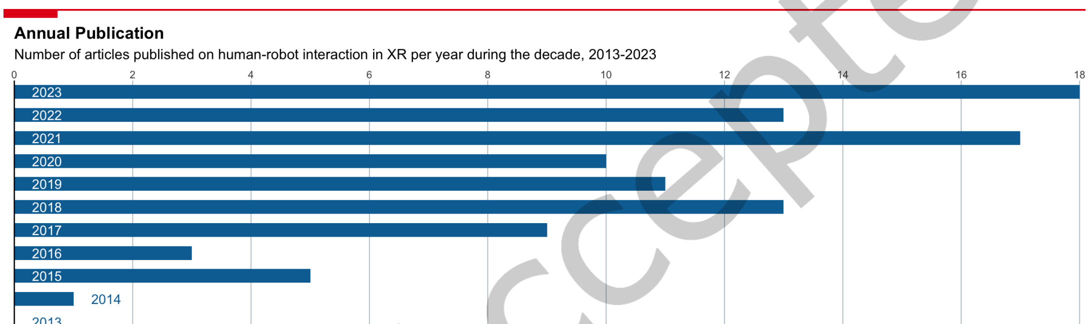

Fig. 3. Articles per year (N=100)

The articles originate primarily from the robotics and manufacturing domains, and our uniqueness is to extend the domains to remote interaction. Keywords such as "reality" and "virtual" frequently appear, with "augmented" being a high-frequency term, although less so than "virtual". This observation may suggest a preference for virtual reality over augmented reality in this research area, but a more detailed analysis is needed to confirm this. Keywords such as "robot", "robots", and "robotics" are also prevalent in statistics. The high frequency of the term "teleoperation" indicates that these articles predominantly focus on teleoperation, while the keyword "human-robot" often appears in the context of collaboration between humans and robots. These two keywords co-exist in several articles, suggesting that they explore the intersection of these two themes, i.e., teleoperated control in HRI. Some articles refer to this cooperation as "collaboration", while others use the term "interaction". Figure 4 presents the top 100 keywords in a word cloud, providing a more precise visualization of the frequency distribution of these terms within the included articles.

3.2 Technologies (RQ1)

Figure 5 shows the use of various XR technologies and various interaction modalities in the analyzed studies. It is evident that Virtual Reality Head-Mounted Display (VR HMD) dominates the landscape, constituting 67% of adoption, and significantly surpassing other apparatuses. Augmented Reality Head-Mounted Displays (AR HMDs) follow with 19% prevalence, while a limited number of studies utilize MR (5%), CAVE (2%), and Mobile AR (1%) technologies.

Most studies favor VR technology, which can likely be attributed to the need for remote user operations to receive information about the remote robotic system or environment. VR technology delivers these data in a more

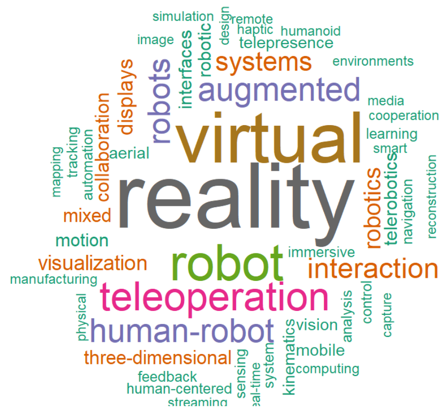

Fig. 4. Frequently-used keywords in the included articles

immersive and intuitive manner, fostering a heightened sense of presence, a key advantage of VR technology. In contrast, AR technology excels at superimposing virtual information onto real environments. Nonetheless, AR faces challenges in satisfying the demand for information overlay on the real robot and its workspace (where the user is absent), possibly accounting for the higher prevalence of VR technology in research. Another contributing factor could be the lower cost of commercial VR HMDs compared to AR HMDs [174]. Furthermore, VR HMDs can integrate supplementary depth cameras to achieve functionality similar to AR HMDs, as exemplified by the study conducted by Yew et al. [178]. Their research prototype used the attached camera to track the pose of the Oculus Rift HMD and the robotic arm to generate and display the AR environment in the HMD. This factor may explain the increased usage of VR HMDs in research. Additionally, a handful of studies employ a combination of devices, such as AR and VR [7], often within the context of multi-person remote collaboration. In these scenarios, a local operator utilizes an AR HMD to manipulate the robot, while a remote expert wears a VR HMD to obtain immersive three-dimensional guidance of the local robot and work environment. Subsequently, this guidance information is transmitted to the local worker's AR HMD.

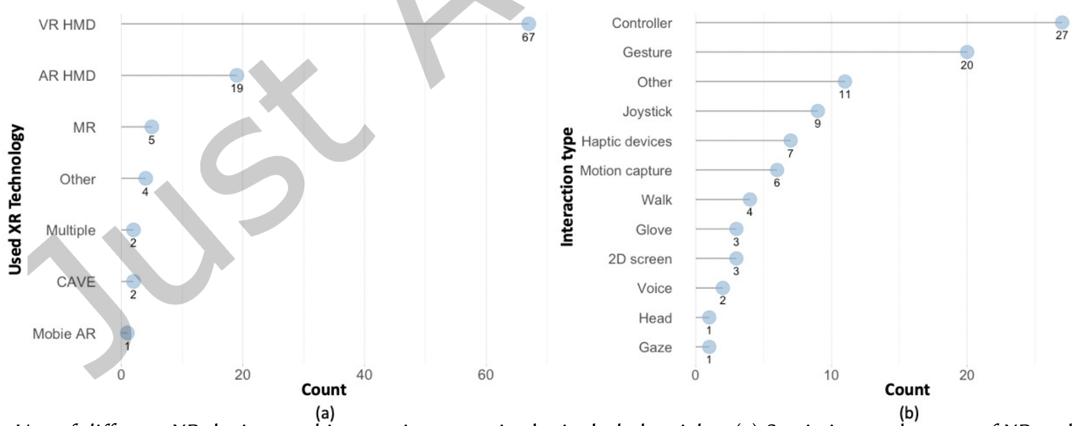

Fig. 5. Use of different XR devices and interaction types in the included articles. (a) Statistics on the type of XR technology used in the studies; (b) Statistics on the interaction modality used in the studies.

The most prevalent interaction type involves using built-in controllers provided with the devices (27%). This trend is reasonable, as the most commonly utilized XR devices are VR HMDs and commercial VR HMDs generally include their proprietary controllers, regardless of form factors. Gestural interaction is another frequent method (20%), primarily because commercial AR HMDs predominantly use hand or head gestures for interaction with virtual content. Additionally, many studies employ VR HMDs with supplementary depth cameras, such as Leap Motion [5], mounted on the headset to detect user gestures, as remote robot operation through gestures is often more intuitive than using controllers. The joystick interaction, typically associated with gamepads like Xbox [6], features prominently in the research (9%) and is often considered when the subject robots are drones or mobile robots [3, 12, 62, 132, 154, 167, 181], only the study by Vu et al. [150] used the joystick to manipulate the robotic arm. The use of virtual fixture haptic devices is also relatively high (7%). Haptic devices can overlay enhanced sensory information on users' perception of the real environment to improve human performance in both direct and teleoperated tasks [45]. Motion capture interactions often require users to wear sensors (6%) and map robotic arms to arm or shoulder coordinates, making operations more intuitive and reducing learning costs. The remaining interaction types, such as actual walking in remote environments (4%), 2D screen (3%), voice (2%), glove (3%), head movement (1%), and gaze (1%) constitute a minor percentage overall. These less common interaction options are frequently linked with specific robot operation tasks. For example, Moniri et al. studied user visual attention in HRI, and gaze was chosen as the interaction method since the human eye gaze is an important indicator of the direction of visual attention focus [94].

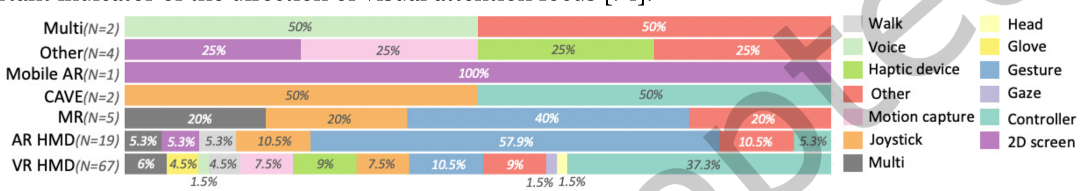

Fig. 6. Percentage bar stacking chart for different XR technologies corresponding to the interaction types.

Building on the previously discussed XR techniques and interaction modalities, Figure 6 provides a comprehensive overview of these elements in the examined research. Interaction refers to how users manipulate virtual environments or objects through their actions. Our findings indicate that the choice of XR technology partially determines the interaction paradigm. VR emerges as the most popular option for remotely controlling robots, with most studies opting for controllers or gestures as interaction methods. VR devices appear compatible with a diverse range of control techniques, except for 2D screens (i.e., touchscreen devices such as tablets or smartphones), typically not employed by VR HMDs. This exclusion is logical, given that VR devices obstruct the user's line of sight to the real world, making it impossible for users to view content on a 2D screen while wearing a VR HMD [20]. In contrast, most systems utilizing AR HMDs rely on gestures for interaction [49]. This preference may stem from the nature of consumer-grade AR HMDs, in general, not including proprietary controllers, making gestures a convenient, self-contained interaction solution.

3.3 Robot Types and Tasks (RQ2)

Our analysis offers a summary of the various robot types (see Fig. 8 with examples) featured in the included studies, as well as the specific tasks discussed or tested in the articles. This information is illustrated in Figure 7. As shown in Figure 7(a), the robotic arm is the type of robot most extensively researched, accounting for 41% of the studies. Mobile robots and drones follow, with respective shares of 20% and 10%. We categorize a special type of robot – robotic arm + mobile robot. This type of robot has the characteristics of both robotic arms and mobile robots, and its main structure is a movable base on which there is a robotic arm. It can be interpreted as a movable robotic arm. Such a robot type was 3% of the included articles. Other types of robots, such as humanoid robots (with human facial features, 8%), two-armed robots (7%), and medical robots (3%), constitute a smaller portion of the overall robot types. Regarding the specific tasks performed by the robots, Figure 7(b) reveals that the three most dominant task types are object grasping/picking/placement, robot navigation, and specialized

[5] https://www.ultraleap.com/product/leap-motion-controller/
[6] https://www.xbox.com/en-SG/accessories/controllers/xbox-wireless-controller

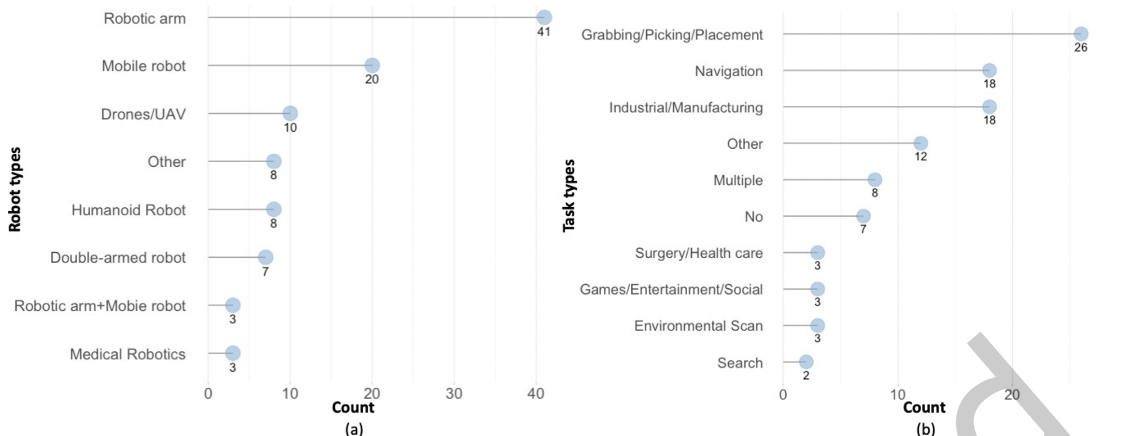

Fig. 7. Different robots and different types of tasks were included in the study. (a) Statistics of the various robot types used in the reviewed studies; (b) Statistics on tasks performed by robots.

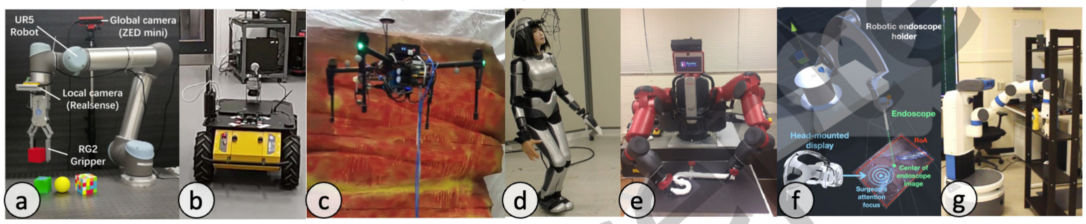

Fig. 8. Examples of different robot types: a) Robotic arm [164]; b) Mobile robot [129]; c) Drones/UAV [149]; d) Humanoid robot[24]; e) Double-armed robot [186]; f) Medical robot[188]; f) Robotic arm + mobile robot [63].

operations in industry/manufacturing, representing 26%, 18%, and 18% of the task types, respectively. A notable fraction of task types is multiple (8%), while some studies do not specify the exact tasks that the robots in the research can execute (7%). The remaining types of tasks, such as remote environmental scanning (3%), surgery/ health care (3%), search (2%), and gaming/ entertainment (3%), comprise a minimal percentage.

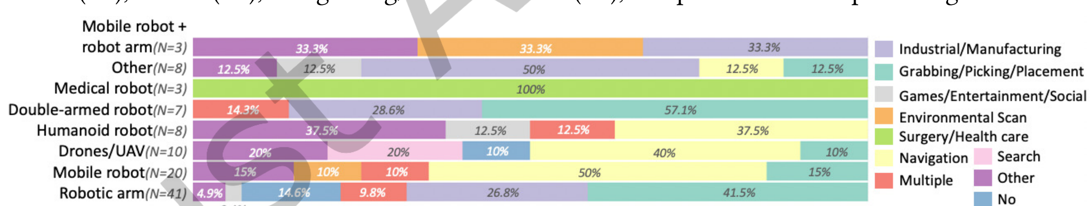

Fig. 9. Percentage stacked bar charts for different robots and corresponding task types.

We also observed potential correlations between various robot types and task types, as depicted in Figure 9. Our analysis indicates that robotic arms, two-armed robots, and other specific robot types, such as industrial machines [96, 97], maintenance robots [178], and mining robots [171], are predominantly employed in industrial or production tasks. These robots typically lack mobility capabilities (i.e., they do not possess a chassis that enables movement), so their task types do not involve navigation. Conversely, mobile robots, drones, and a subset of humanoid and other more mobile robots are responsible for tasks such as navigation, search, and environmental scanning that rely on mobility. Robotic arms and two-armed robots primarily perform the grasping, picking, and placement tasks, as these functions connect to the services or fundamental functions of robotic arms. The tasks are also a key component of the industrial or production chain for which robotic arms are originally responsible. To some extent, two-armed robots can be considered a combination of two robotic arms. Medical robots represent the most homogeneous robot type, as their sole responsibility is to assist in surgical procedures [144, 188].

Furthermore, we discovered that humanoid robots appear to possess unique social characteristics. The tasks they are assigned, such as intervening with children diagnosed with autism spectrum disorders (ASD) [77], engaging in chess games with remote users [126], assisting remote users with dressing [126], expressing emotions [138], receiving and guiding users [59], and maintaining road traffic security [55], are inherently linked to human or social activities. This suggests that humanoid robots, due to their anthropomorphic form and capabilities, are particularly suited for roles that require social interaction or human-like tasks.

3.4 Evaluation of Tasks (RQ3)

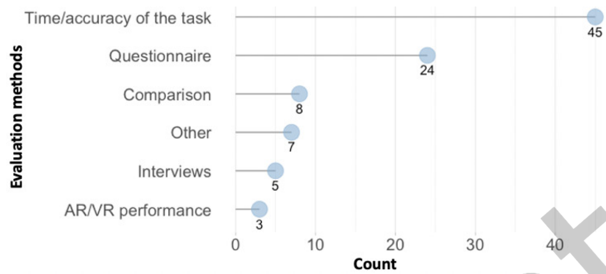

Fig. 10. Statistics of different evaluation methods.

We examined the evaluation methods used in the included studies (see Figure 10). We discovered that 38% of the articles did not conduct any form of evaluation, while 36% employed a single evaluation method, and 36% utilized a hybrid approach. Among these methods, 69% of the articles employed a quantitative evaluation, which included assessing time/accuracy to complete the task (48.91%), administering standardized questionnaires such as the NASA-TLX [61], or employing task-based design questionnaires (26.09%). A mere 5.43% of the articles focused on qualitative user evaluations, such as interviews. Additionally, a small subset of articles, amounting to 8.70%, concentrated on comparing the performance of robots and their digital twins. This type of study is typically evaluated by comparing the trajectory coordinates of the system's input and the robot's output. For instance, studies by Yun et al. [179], Cousins et al. [33], and Bian et al. [13] evaluated the coordinates of the user's hand input and the robot arm's output. Betancourt et al.'s study compared the 3D spatial coordinates of a virtual drone and a real flying vehicle [12]. A few studies, constituting 3.26%, evaluated the performance or impact of AR/VR itself. For example, Kuo et al. compared the accuracy of manipulating objects through VR, video, and the real world [78]. Similarly, Chen et al. evaluated different methods of 3D reconstruction in VR [25].

3.5 XR Technologies Facilitate Effective Remote Robot Collaboration (RQ4)

3.5.1 Virtual Interface Design. We adopt the taxonomy proposed by Walker et al. [153] to analyze the user interface design in XR, dividing it into two components: user perspective and user interface (see Figure 11). The user perspective comprises five categories: robot-coupled, robot-decoupled, Bird's-eye view, dynamic perspective, and others. Robot-coupled perspectives involve users viewing the scene through the "eyes" of the robot. This approach was exemplified in the works of Vempati et al. [149], Chacko et al. [22], and Brizzi et al. [18], who conducted their studies in the remote operating systems of UAVs, humanoid robots, and double-armed robots, respectively, with perspectives coupled to the robots. In this approach, the user's viewpoint is linked to the robot and changes as the robot moves. Conversely, robot-decoupled perspectives enable users to observe the robot's actions from a detached viewpoint, unbound from the robot's movements. This perspective was demonstrated in the work of Kuo et al. [78] and Zinchenko et al. [188], who developed systems that manipulate a remote robotic arm in VR with a perspective decoupled from the robot. Similarly, Stedman et al. [129] employed a decoupled perspective in their work with a remote mobile robot. The Bird's-eye view offers an overhead, top-down view, as illustrated by Jang et al. [70], who utilized this perspective to control swarm robots. while the dynamic perspective allows users to switch between different viewpoints, this perspective was exemplified in the works of Wei et al. [164] and Xu et al. [172], who both employed a combination of two perspectives in their studies.

We categorize the user interface into several types (see examples in Fig. 14). One of these is the direct interface, where the camera on the remote robot transmits to a 360-degree video interface in the virtual environment. This interface, as exemplified by the work of Zhao et al. [185], allows users to directly observe the remote workspace. Another variant is augmented by a 3D reconstruction of the remote environment, as demonstrated in the work of Chen et al. [24]. This approach enhances the user's perception of the remote workspace by providing a more immersive and spatially accurate representation. A different approach involves the use of a digital twin of the robot. In this setup, a digital replica of the remote robot exists in the virtual environment, and the user controls the remote robot by manipulating the digital twin, as illustrated in the work of Zinchenko et al. [188]. Other interface options include the digital twin combined with a 3D reconstruction of the remote environment as an example by Kuo et al. [78], virtual control room, rebuilding a virtual console in the virtual environment, as demonstrated by Kalinov et al. [73], 3D reconstruction of the remote environment as shown in the work of Zein et al. [181], or a combination of the aforementioned interfaces. These various interface designs offer different levels of immersion, control precision, and spatial awareness, catering to different user needs and task requirements.

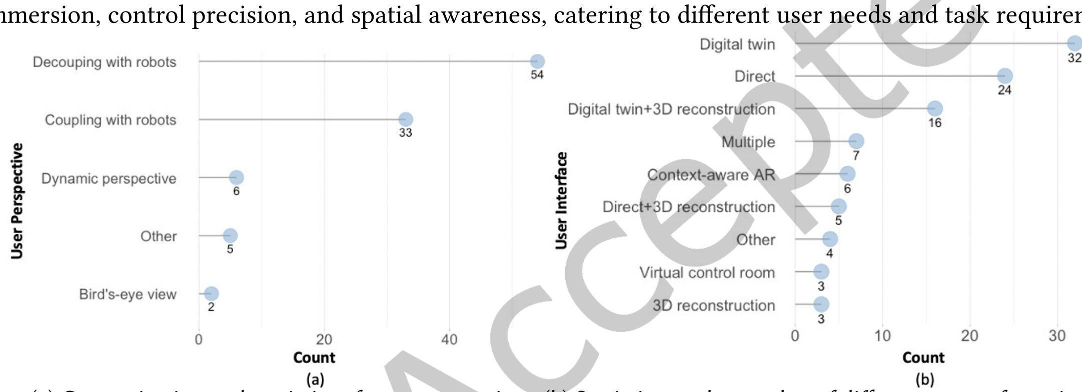

Fig. 11. (a) Categorization and statistics of user perspectives; (b) Statistics on the number of different types of user interfaces in the review.

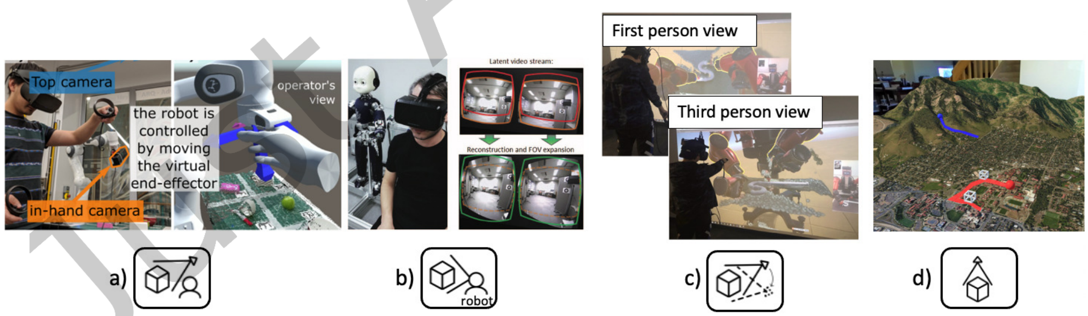

Fig. 12. Examples of different user perspectives: a) Decoupling with robot [102]; b) Coupling with robot [141]; c) Dynamic perspective [186]; d) Bird's-eye view [151].

We discovered that the user's perspective (see Fig. 12 with examples) predominantly involves robot-coupled (33%) or robot-decoupled (54%) views, while other perspectives, such as dynamic (6%), bird's-eye view (2%), and others (5%), constitute a small proportion. Regarding user interfaces, the digital twin (32%) and direct interfaces (24%) are most prevalent, followed by a considerable share of digital twin combined with 3D reconstruction of the remote environment (16%). The remaining interfaces, including combinations of multiple interfaces (7%), overlaying virtual interfaces on real environments (i.e., context-aware AR interface, 6%), direct interfaces augmented with 3D reconstructions of the remote environment (5%), standalone 3D reconstruction of the remote environment (3%), virtual control rooms (3%), and other interfaces (4%), represent a minor portion.

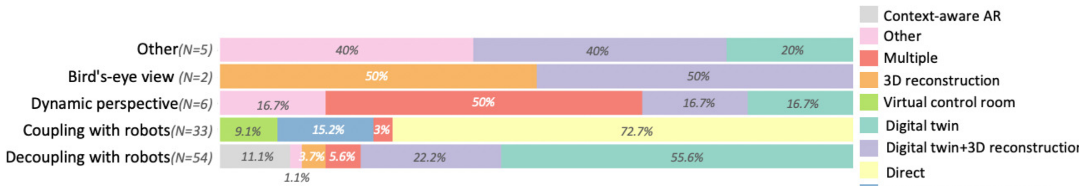

Fig. 13. Percentage stacked bar chart of the relationship between the user perspective and the user interface.

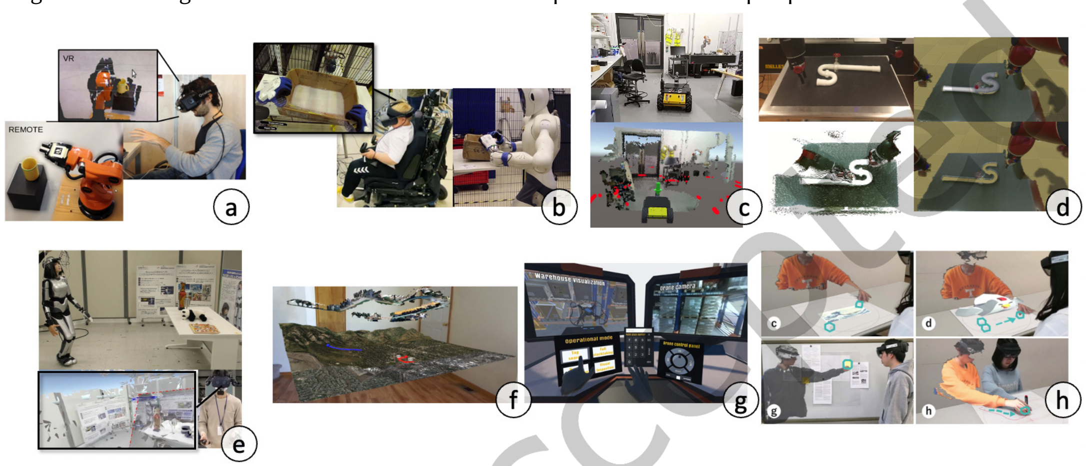

Fig. 14. Examples of different virtual user interfaces: a) Digital twin[107]; b) Direct interface[120]; c) Digital twin + 3D reconstruction[129]; d) Multiple interface[186]; e) Direct + 3D reconstruction interface[24]; f) 3D reconstruction interface[151]; g) Virtual control room[73]; h) Context-aware AR interface[68].

Consistent with our previous analysis, we investigated the relationship between user perspective and user interface using a percentage stacked bar chart, as illustrated in Figure 13. The most significant observation is that the direct user view within the user interface tends to be robot-coupled. This is primarily because users need to observe the remote environment's workspace directly from the robot's viewpoint. Moreover, we found that when the user perspective is decoupled from the robot, it is frequently necessary to incorporate a digital twin of the robot within the XR environment. This is understandable since users need to be aware of the remote robot's motion state, necessitating the creation of a corresponding digital twin in XR to facilitate better comprehension of the robot's operational state. Notably, most studies have opted for the digital twin solution, while only a few, such as Xu et al. [175], have employed a camera positioned next to the remote robot to convey the robot's work status via live video. The virtual control room user interface also requires the user's perspective to be robot-coupled. Although only two studies employed Bird's-eye perspectives, we observed that both executed 3D reconstructions of the remote environment. Lastly, we also discovered that user interface designs for dynamic perspectives tend to be more intricate, such as Zhou et al. [186], often incorporating multiple user interfaces.

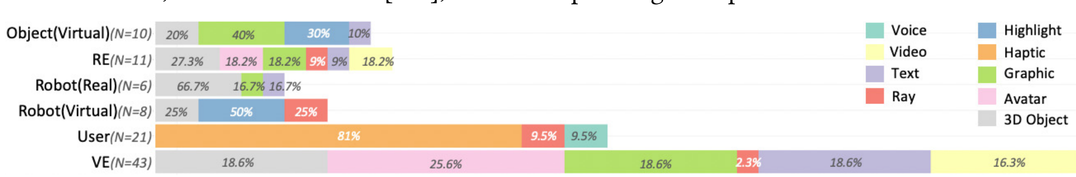

Fig. 15. Relationships between types and locations of multimodal enhancements.

3.5.2 Enhancement. In our analysis of the included studies, we evaluated the location and type of information enhanced by the multimodality of the systems. We found that 39% of the studies employ only a single modal approach to improve remote operations. Most of these systems only supported users to view remote robots, workspaces, or environments immersively via XR. Such systems do not have multimodal enhancements in any location, and the improvement to the user is simply the immersion of XR. On the other hand, 62% of the studies opted for multimodal enhancement during specific parts of remote robot operation (see Figure 15 for details). The primary areas of enhancement were the virtual environment (43.43%) and the user (21.21%). A smaller number of studies chose to enhance virtual objects manipulated within the virtual workspace (10.10%), virtual (8.08%) and real robots (6.06%) or in real environments (11.11%). This could be attributed to the fact that fewer remote operating systems are using AR technology compared to VR technology. AR technology is the predominant technology applicable to augmentation in real robots and real environments. Augmentation on top of real robots often necessitates collaboration among multiple individuals. For instance, the work of Mourtzis et al. [97] and Schwarz et al. [126] involved a cooperative system with multiple users, with one user at a remote location and another user with the robot. However, this type of system represents a very small percentage of our collected articles (for more details, see Section 3.6). Additionally, we observed that the highlight enhancement mode was only applied to the virtual representations of the robot and the objects being manipulated. This enhancement type accentuates specific parts of the digital twin or the virtual object for clearer interaction. Avatar enhancement also appeared exclusively in the virtual environment. On the user side, enhancement primarily involved haptic feedback, utilizing tools such as haptic gloves or virtual fixtures. The research conducted by Du et al. [40], Aschenbrenner et al. [7], and Hormaza et al. [65] also explored the use of voice enhancement on the user side. In the environment, whether real or virtual, a few studies employed live video enhancement, opening a live video window within the environment, such as Zinchenko et al.'s work [188].

3.6 Multi-player and Multi-Robot Interaction Support (RQ5)

Figure 16 illustrates some examples of multi-player and multi-robot interaction system. In our analysis, 81% of the included articles did not support multi-player or multi-robot interaction, and all of their collaboration involved one user or operator collaborating with a single robot. Only 19% of the interaction paradigms described in the articles supported multi-player or multi-robot interaction. One of the most common interaction paradigms involved one user interacting remotely with a single robot, while another user was situated next to the remote robot (N = 12). For instance, articles by Mourtzis et al. [96, 97], Black et al. [15], Fuchino et al. [48] and Moniri et al. [94] propose that a novice operator or worker on the robot side receives instruction from a remote expert or commander using XR for environmental awareness and communication. The collaboration model in both Jang et al. [69] and Gong et al.'s [55] studies involved one user collaborating with multiple robots (N = 2). The remaining modes include multiple users interacting with multiple robots, such as the work of Phan et al. [110] and Honing et al. [64] (N = 2), multiple users operating a single robot, as exemplified by Galambos et al. [51] (N = 1), and one user collaborating with multiple robots while multiple users are present on the robot side, as demonstrated by Aschenbrenner et al. [7] and Walker et al. [151] (N = 2).

3.7 Exploration of System Latency Issues (RQ6)

We discovered that system latency effects are not considered in nearly half of the studies (49%), despite the fact that remote control latency is a critical factor influencing user experience and task accuracy. A significant number of our included studies simply acknowledged the presence of system latency or claimed that their systems experienced delays, without providing specific measurements or quantifying the duration of their system latency (23%). Only 28% of the studies explicitly analyzed the latency of the system, and we compiled the reported study times in Figure 17. Figure 17 (a) displays the latency times reported by all studies, revealing that the majority of

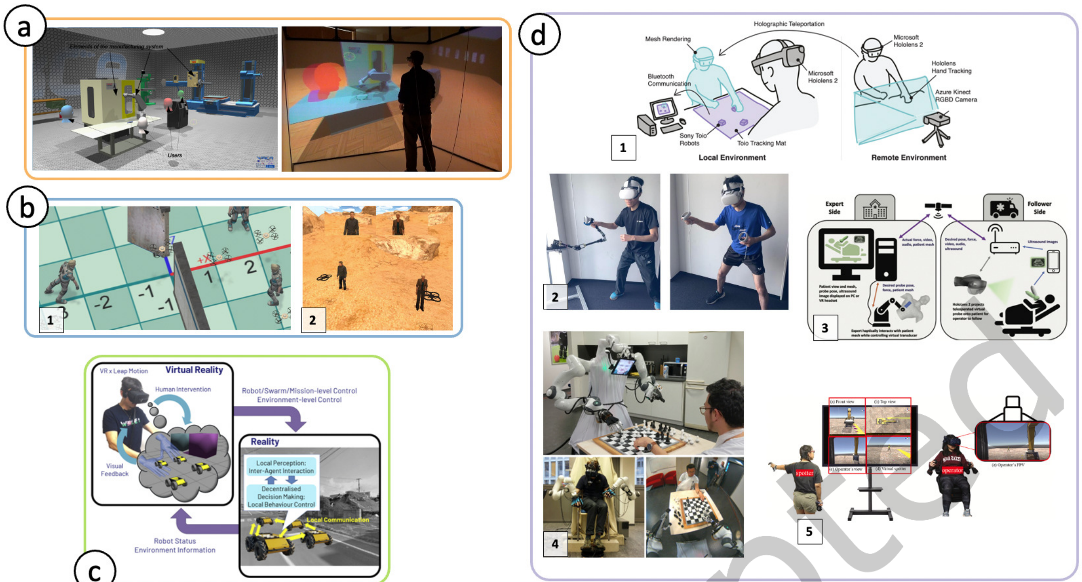

Fig. 16. Examples of systems that support remote multiplayer or multi-robot interaction: a) Multi-user remote operation of one robot [51]; b) Multi-user remote operate multi-robot, 1: Multi-user control of multiple drones [110], 2: Two users control two drones [64]; c) One user remotely operates multi-robot [69]; d) A local user and a remote user cooperate together to control a robot, 1: Local and remote users collaborate with desktop robots [68], 2: Coach and learner remotely exercise table tennis in VR via robotic arm [48], 3: Experts remotely guide novices in using medical equipment [15], 4: The local user controls the robot to play chess with the remote user [126], 5: The conductor remotely directs the user to maneuver the excavator [85].

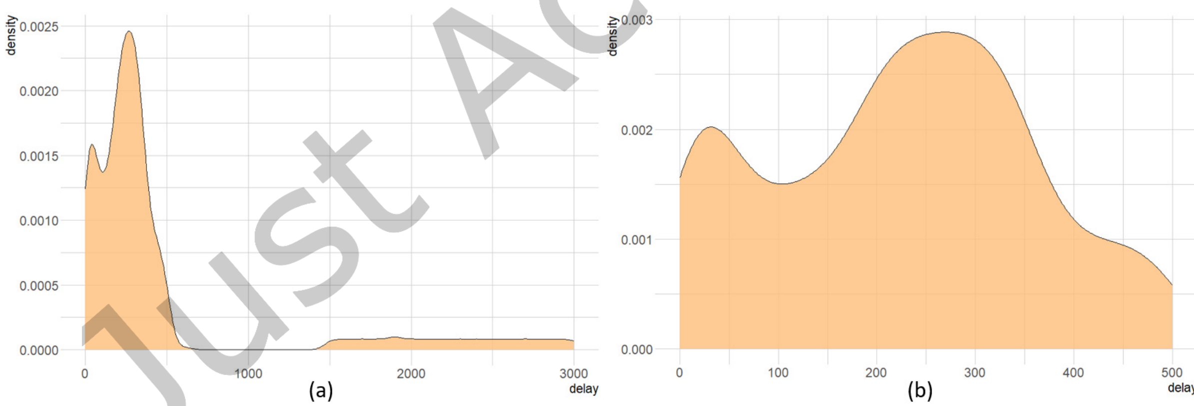

Fig. 17. Density distribution of system latency times (unit: ms) in the included studies. (a) Latency times were reported by all included studies; (b) Latency was scaled to 0 − 500ms to show more detail.

studies had latency times within the range of less than 500*ms*, and only a few studies had latency times in the range of greater than 1000*ms*. Focusing on latency times less than 500*ms*, as seen in Figure 17 (b), we found that most studies had delay times within the range of 200*ms* to 400*ms*. Among the included studies, several stand out for their unique approaches. Le et al.'s [80] research focused on controlling system latency and comparing the impact of different latencies on the user experience. In contrast, McHenry et al. [92] study employed an asynchronous system operation to overcome the high latency associated with Earth-Moon transmission. These studies highlight different strategies for addressing latency in extended reality systems for remote robotic control.

4 Discussion

Figure 18 provides a visual summary of the number of articles linked to each dimension (see Section 1.1), based on our data extraction. The Results section outlines the factual aspects of these articles, while this section interprets those findings and offers insights, focusing on three main areas: robotics, design, and XR technologies.

4.1 Impact of different robot types on remote HRI

4.1.1 How different robot types influence user interface and user perspective. Our analysis has identified specific associations between various types of robots and their corresponding tasks (Section 3.3), which are critical to shaping the user interface and perspective design (See Figure 19). As analyzed in Section 3.3, industrial robots, such as robotic arms and double-armed robots, are primarily deployed for production tasks. Given their limited mobility [6], these robots necessitate a user interface and perspective that focuses primarily on precision control and manipulation [35], rather than navigation. For instance, user interfaces designed for double-armed robots and robotic arms often facilitate multi-viewpoint together with dynamic perspective observation to assist users [173, 186]. A key distinguishing factor between these two robot types is their preferred control method in their user interface designs: two-armed robots typically favour direct operation, whereas robotic arms generally reflect a digital twin [104, 134, 147] (i.e., the data mapping between the virtuality and physical worlds). This distinction could stem from the fact that the double-armed configuration of the robots aligns closely with the human dual-arm anatomy [54]. This alignment facilitates direct control of the remote robot and couples the user's perspective with the robot, potentially reducing the user's learning curve while making the operation more intuitive.

On the contrary, mobile robots, drones, and certain humanoid robots with advanced mobility are designed for tasks that require navigation, search, and environmental scanning. These tasks require user interfaces and perspectives that enhance spatial awareness and promote dynamic movement within the remote environment [177]. For instance, mobile robot interfaces uniquely support a bird's-eye view, assisting users in comprehending the remote three-dimensional environment [7, 151]. In addition to the above interface designs related to the type of task the robot is assigned, it can be noticed through the bubble diagram that overall, direct interface have been widely used in all types of robots except for medical robots. Perhaps one reason is that the cost of such direct interfaces is minimal and does not require 3D reconstruction or the creation of a digital twin of the real robot [2, 66]. Moreover, we find that such direct interfaces are often coupled with the robot perspective, except for being more intuitive, probably also for cost reduction considerations. This design often requires only one or two cameras mounted on the robot [8, 50, 120, 183]. Finally, humanoid robots, due to their unique social characteristics, are usually engaged in tasks related to human or social activities [47]. In such scenarios, user interface and perspective designs should be geared towards intuitive control and interaction, allowing users to interact with the robot more human-like and socially.

Our analysis reveals a clear trend in interface design across different robot types. For stationary or industrial robots (e.g., robotic arms), interfaces prioritizing precise manipulation often predominate. Digital twin interfaces, which decouple the user's viewpoint from the robot's, are beneficial for monitoring the robot's state and position, while direct interfaces offer cost-effectiveness when the focus is strictly on task execution. Conversely, mobile robots (e.g., drones, mobile robots, humanoids) typically require broader spatial awareness and navigation aids, such as panoramic or bird's-eye views. In general, user perspective choices are driven by immediate concerns, and the interface should be tailored to support the robot's specific tasks.

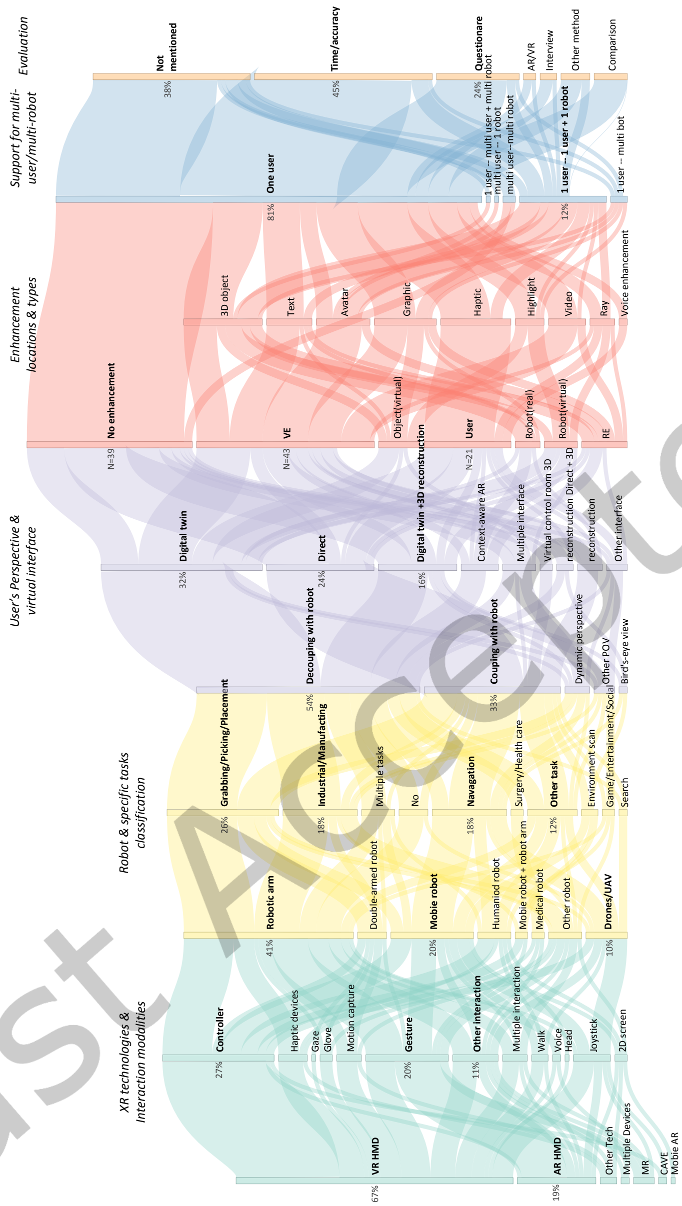

Fig. 18. Sankey Diagram, a visualization with overall counts of characteristics across all dimensions. Different stages (nodes) colors are used to distinguish different dimensions; ■: XR technologies and interaction modalities; ■: Robot and specifictasks classification; ■: User's Perspective and virtual interface; ■: Enhancement locations and types; ■: Support for multi-user/multi-robot; ■: Evaluation of tasks. Where the width of the flow (links) is proportional to the total number of articles we reviewed (N = 100).

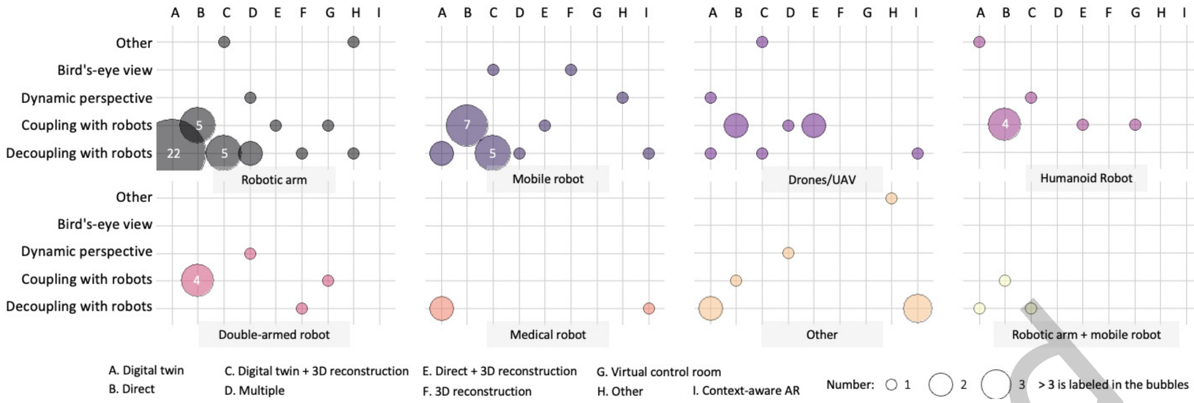

Fig. 19. Bubble diagram of the relationship between types of robots and user interface and perspective. Different colored bubbles represent different robot types, the vertical axis indicates different user perspectives, and the horizontal axes A to I indicate different user interfaces. The size of the bubbles is proportional to the number of articles; the bigger the bubble, the more articles in the corresponding category, or vice versa. e.g., among the 100 articles reviewed, one study utilized a robotic arm, adopting a perspective of coupling with the robot through a virtual control room interface, while 22 articles also focusing on robotic arms, employed a decoupling perspective using a digital twin user interface.

4.1.2 Types of robots affect virtual enhancements. In our review, medical robots extensively utilize virtual enhancement elements, with all examined medical robots incorporating such enhancements in their interactions [15, 144, 188]. We acknowledge that this observation may be influenced by the limited number of medical robot studies (only 3) included in our review, which may not provide a sufficient representation of the field. Alternatively, this high utilization of virtual enhancements could be attributed to the unique requirements of medical task scenarios, which require full use of the XR capabilities to assist operators [140], such as physicians.

Robotic arms, an industrial robot type characterized by their rising popularity, also demonstrated a high use of virtual augmentation elements, with 61% of the examined robotic arms incorporating one or more such elements. This prevalence may be due to the extensive research focus on XR-based remote control of robotic arms, leading to a greater exploration of XR's unique augmentation characteristics. In particular, robotic arms were the only robot type that augmented virtual objects, which could be associated with their common task of picking and placing objects [169]. This task may necessitate additional augmentation on objects to enhance user-manipulation capabilities.

Furthermore, robotic arms and mobile robots were the only types that utilized text overlays in the virtual environment. This feature may also be related to the tasks performed by robotic arms and mobile robots. Robotic arms are often used in professional contexts and industrial environments, where remote operators may benefit from text prompts or reminders for the next operational task, as exemplified in the design by Wang et al. [159–161]. The remote operator of the mobile robot can also get information about the orientation of the robot movement from the text prompts [74, 137].

The unique mobility properties of mobile robots also influenced the choice of virtual enhancements. Mobile robots make extensive use of enhanced design in their environments (64.29%), both virtual and real. Many designs used 3D object enhancements in the environment [34, 69, 151]. This feature may be necessary to provide spatial location cues for users navigating remote mobile robots using XR, a requirement that does not apply to other non-mobile robots [9]. Trinitatova et al.'s design uses a robotic arm but still uses 3D object enhancement in the environment [145]. Their purpose was to use 3D spheres to indicate the center of the manipulated part of the robotic arm, still to indicate positional information in space. This suggests that the enhancement of 3D objects in the environment is often associated with positional information.

4.2 Designing remote HRI system with users and scenarios

The influence of users and scenarios on remote HRI system design is a crucial aspect to consider when developing effective human-robot collaboration experiences. Depending on the users' expertise, background, and preferences, as well as specific task scenarios, the design of remote HRI systems may need to be adapted accordingly to ensure optimal performance and usability. For example, adjust the user perspective and virtual interface accordingly, or select different enhanced design elements appropriately (see Section 3.5).

Expertise level of users and its impact on interaction design. Expert users (e.g., engineers or technicians) have advanced skills, enabling them to handle complex interfaces and control mechanisms. In contrast, novice users (e.g., non-specialist workers or first-time users) typically need more intuitive, user-friendly interfaces that emphasize ease of use and learning over advanced functionality [131, 166]. Additionally, some beginner-oriented systems may consider support multiple operators, facilitating remote expert guidance [15]. **Designing adaptive systems to suit diversified scenarios.** Different tasks may require varying levels of detail and control in interface design. High-precision scenarios (e.g., remote surgery or delicate manipulation) demand fine-grained control and extensive feedback [184], while simpler tasks (e.g., basic navigation or object transport) benefit from streamlined designs prioritizing usability and efficiency [17]. Some tasks also require specialized interface features; for example, search-and-rescue missions may incorporate real-time mapping and tracking to assist robot navigation and localization [113].

4.3 The role of XR in facilitating remote HRI

Remote interaction via XR devices can be considered an instance of mediated reality [90], where the user's interaction with a remote space is augmented or altered through technological mediation. This experience reshapes the user's interaction with the robot, providing new possibilities for remote HRI and allowing the user to manipulate the robot and the remote space in ways that are not possible in the physical world.

4.3.1 Enhancing user perspective and understanding of remote environments. One key advantage of XR in remote HRI is creating highly realistic and accurate digital twins of both the remote environment and the robot itself (see Section 3.5.1, digital twins are the most commonly used user interface). Users can better understand the remote workspace by simulating the robot's movements and actions within a virtual environment, making planning and executing tasks easier. Furthermore, the digital twin approach allows for more intuitive control mechanisms, as users can directly interact with the virtual representation of the robot, which in turn is mapped to the physical robot's behavior [123]. Another essential aspect of XR in remote HRI is the provision of enhanced user perspectives (Section 3.5.1). XR technology offers users a variety of perspective options, allowing them to adjust views based on their preferences and specific task requirements, improving situational awareness and overall task performance [29].

4.3.2 Supporting multi-player or multi-robot interactions through XR.. One of the key advantages of XR in supporting multi-user or multi-robot interactions is its ability to create a shared virtual space, enabling more effective collaboration and communication [124]. This shared space allows users to visualize the actions and intentions of others, improving overall task performance and efficiency. Additionally, XR technology facilitates the development of advanced user interfaces for multi-user or multi-robot scenarios. For example, XR interfaces can provide real-time status updates, task assignments, and performance metrics for each user or robot, enhancing monitoring and decision-making. Moreover, XR interfaces support intuitive control mechanisms, allowing users to seamlessly switch between controlling multiple robots and collaborating with others. Despite the advantages of XR in multi-user and multi-robot interactions, our review indicates that a significant portion of the included studies do not support such interactions (Section 3.6). Most research focuses on one-user-one-robot collaboration, suggesting that XR's potential in this area remains underutilized. In the studies we reviewed, those actively exploring multi-user and multi-robot collaboration should receive more attention from future researchers, who can draw insights from their system designs and application scenarios. Examples include collaboration between local and remote users [15, 48, 68, 85, 126] and one-user-multiple-robot interaction [69].

In XR-enabled multi-user or multi-robot interaction systems, latency is a critical challenge, as emphasized by Jay et al. [71], latency can severely impact the effectiveness of collaboration and the user's presence in the virtual environment, affecting the user's experience. High latency can disrupt coordination between users and robots, leading to errors and decreased task performance. Minimizing latency to enhance user experience and ensure smooth real-time interactions is important. Our review was pleased to find that many studies have noted the problem of latency (Section 3.7); however, addressing latency is especially important for systems that require the participation of multiple users and robots. As the number of participating users and robots increases, the latency problem may become more pronounced (See Section 5.2.2 for further discussion).

4.3.3 The use of multimodal enhancement in XR to improve remote operations. XR technologies in remote HRI create a form of sensorimotor reality [148], where users can perceive and interact with remote environments through enhanced sensory inputs and motion outputs. For example, haptic feedback allows users to feel remote objects, and spatial audio enhances their situational awareness [81]. These capabilities extend users' abilities by providing sensory augmentation that directly supports robot control and task execution in remote spaces. Also, multimodal enhancements can significantly improve the user experience and task performance in remote operations [76]. Our results (Section 3.5.2) provide a detailed summary of the locations and kinds of multimodal enhancements. These enhancements provide users with more intuitive and immersive ways of perceiving and interacting with the remote environment and the robots involved. By leveraging multimodal feedback, XR can help bridge the gap between the user and the remote workspace, leading to more efficient and accurate task execution [88]. For example, visual enhancements, such as highlighting specific parts of a robot or virtual object, can draw the user's attention to important elements and provide context-aware information [14]. Auditory feedback, such as spatial audio or voice commands, can deliver crucial information to the user and facilitate natural communication with other users or the robotic system itself [118]. Haptic feedback, enabled through devices such as haptic gloves or virtual fixtures, can provide users with a more tangible sense of touch, enhancing their perception of the remote environment and improving their ability to perform complex tasks [103]. Our results indicate that a significant portion of the studies incorporated multimodal enhancements in their XR systems, with a focus on augmenting the virtual environment, user, and virtual objects. However, there remains room for further exploration in terms of augmenting real robots, real environments, and other aspects of remote operations. We will provide further recommendations for future researchers in Section 5.2.1.

5 Challenges and Future Directions

After reviewing our comprehensive survey, we found that XR-based remote HRI research has made significant progress but still faces many challenges and opportunities for development. The following sections discuss our insights in detail, offering guidance for future researchers and developers. We categorize these into three main areas: challenges in the selection of evaluation methods, Unleashing the Potential of XR in Remote HRI, and user-centered system design.

5.1 Challenges in the selection of evaluation methods

Selecting effective evaluation methods is a key challenge for future researchers of XR-based remote human-computer interaction. Our review found that 38% of relevant studies did not report any evaluation methods (see Section 3.4). By analyzing existing work, we classify evaluation methods into two main categories: (1) **system efficiency** and (2) **user experience**.

Evaluating system efficiency involves metrics such as system latency and task completion time, which should align with the intended applications. For example, industrial production tasks can emphasize latency, accuracy and completion time [60], while digital twin interfaces (XR as a teleoperation extension) can focus on ensuring that the virtual twin mirrors the actions of the real robot, often comparing trajectory data [87]. Because XR can significantly affect latency, future research should consider assessing its impact on teleoperation efficiency [4]. In contrast, evaluating user experience in XR-based HRI centers on more subjective factors such as satisfaction and ease of use. Qualitative methods are often used, including structured interviews and standardized scales. Researchers can use *NASA-TLX*[61] to measure workload, the *System Usability Scale*[19] for usability, or the *User Experience Questionnaire*[79] for broader impressions. Interviews can also yield deeper insights into user perceptions beyond what quantitative measures can capture[106].

5.2 Unleashing the Potential of XR in Remote HRI

Rapid development of XR will redefine the field of remote HRI by enabling more immersive, intuitive, and efficient interactions. In our reviewed studies, Chen et al. [24] used VR and controllers to intuitively control remote humanoid robots (immersion), Walker et al. [151] used AR to generate an indoor bird's-eye view to assist with remote robot planning and deployment (intuition), and Holobots [68] is an MR-based remote collaboration system (efficient interaction). However, much of the potential of XR remains unexplored under current research conditions, and future researchers could focus on the following directions.

5.2.1 Multi-modal Remote Enhancement in XR. In the visual domain, identifying the most effective visual augmentation is important for different operational contexts. For instance, one might compare 3D object overlays with highlighted cues in the user's field of view for mobile robot navigation. Jang et al. [69] explored both efficacy and subjective perceptions of two virtual enhancements: a Pick-and-Place interface (*Highlight*) and a Virtual Wall interface (*3D object* overlay). Similarly, object manipulation tasks can benefit from highlighting or pictorial markers (e.g., arrows) to expedite remote operators' responses. However, effective implementation of these virtual cues becomes more complex when the target object is outside the user's field of view [16, 58]. Although such strategies have been investigated in XR applications, they remain underexplored in robot teleoperation. Further research is needed because current findings may not generalize to various robot types and task requirements. Future studies could also examine visual enhancements across different robot categories and tasks.

Pinto et al. [111] showed that emphasizing sensory feedback related to unexpected events (e.g., items dropped on the robotic arm), is crucial for maintaining a positive user experience. Meanwhile, Rivera-Pinto et al. [117] employed spatial acoustic feedback to aid rapid localization of robotic targets. However, acoustic feedback can extend beyond localization or speech communication. With the growing prevalence of social robots [53], auditory feedback can convey emotions, enhancing human perception and collaboration in teleoperated settings [182]. Regarding haptic feedback, there may be requirements for additional haptic devices [15, 40, 44, 45, 75, 126, 135, 145, 170]. However, specialized haptic devices are not applicable across different HRI systems, particularly when operators need to switch seamlessly among multiple robots via XR. Future research could explore vibrotactile feedback in commercial XR controllers or pseudo-haptic [146], reducing hardware overhead while still providing essential haptic cues. In particular, additional haptic devices carry the risk of increasing the physical burden of operators [162].

Although our survey has identified several instances of using multimodal enhancement, few studies have examined how these multimodal cues enhance user understanding and control in teleoperation. Moreover, the integration of multimodal approaches is rarely discussed in current research. The future researcher should explore combining these elements to propose optimal strategies that reduce cognitive load, improve task performance, and increase user immersion.

5.2.2 Multi-player and multi-robot interactions and system latency. XR is increasingly popular in industrial, collaborative, and social domains, and the need for multi-user, multi-robot systems is likely to grow [187]. Researchers and developers must address system delays to cope with future trends (see Section 4.3.2). Waltemate et al. [155] found that latency above $75ms$ affects perceived motor performance and simultaneity, above $125ms$ reduces agency and body ownership, and worsens beyond $300ms$. Although our survey observed that most studies latency within a resonable range of $200ms$ to $400ms$, users may notice delays in the system, but not enough to completely disrupt their experience. As more people or robots join the system, the latency effect may worsen, making it an important issue for future researchers to address. However, Waltemate et al. [155] noted that whether participants notice latency in a virtual environment may depend on the motor task and its performance, rather than the physical latency. This suggests that a user-friendly interaction design can largely mitigate the negative effects of latency. However, for tasks requiring precise control, latency may need to be kept within $75ms$.

To mitigate latency in XR tele-robotics, predictive algorithms [133] and time warping [155] have shown promise. Foveated rendering [5, 105] or lowering VR image quality can also help but require further study. Especially in multiplayer and multi-robot contexts, these approaches may require trade-offs among usability, effectiveness, and user experience [43].

5.2.3 Navigating Complex Environments with XR in Remote HRI. Remote operation in complex environments presents significant challenges [83]. Robots may need to navigate various terrains, unpredictable obstacles, or dynamic conditions, each adding complexity for remote operators. The fidelity of reproducing these environments in XR can impact the effectiveness of human-robot interaction (HRI). Currently, most systems use a static third-person perspective or a robot-coupled view to transmit environmental data to remote users. However, whether relying on a single viewpoint or merely replicating the robot's perspective, this approach may not provide enough context for users to make optimal decisions. Advanced XR interfaces have the potential to offer a comprehensive environmental overview, such as dynamic or bird's-eye views, but our survey found that such designs remain limited.

Future research could focus on the integration of multiple perspectives [143], such as a primary view linked to the robot, a third-person perspective to observe the robot, and a top-down bird's-eye view. This multi-perspective approach could enhance remote users' comprehension of the remote space. It might even be feasible to establish a virtual environment camera to monitor the user avatar's operational state and the robot's digital twin from a third-person perspective in the virtual environment [11], potentially mitigating risks associated with certain tasks. Another promising avenue is the enhancement of the environment through multimodality, such as incorporating haptic feedback and auditory cues to enrich the user's perception of the remote environment. In addition, adaptive environment reconstruction presents a potential research direction. Different robots and tasks may necessitate varying degrees of environmental reconstruction fidelity. For instance, pick-and-place tasks may only require low-fidelity reconstruction of the operator's table, while geological exploration tasks may demand high-fidelity environmental reconstruction. Implementing adaptive environmental reconstruction tailored to specific tasks and robots could potentially reduce system latency and prevent bandwidth wastage [168]. Future research should focus on the development of advanced environmental reconstruction techniques to provide a more comprehensive and real-time depiction of complex environments.

5.2.4 Digital Twin in XR-based Remote Human-Robot Interaction. Digital twins also present a range of opportunities for future research. In reviewed studies, digital twins have emerged as an important component in enhancing the interaction and integration between physical and virtual environments. Digital twins can be used as intuitive interfaces to improve system efficiency and user experience, and widely used in industrial and social scenarios [123]. Future research should focus on improving the fidelity of digital twins and their accurate real-time synchronization with their physical counterparts, which is critical for applications in industries such as manufacturing, healthcare, and urban planning. Furthermore, the integration of advanced machine learning and artificial intelligence techniques can provide smarter and more autonomous digital twins capable of predictive maintenance, adaptive learning, and decision support [67].

Digital twins also provide possibilities for the study of scenarios that are impossible or impractical to test in reality. For example, the design by Su et al. [134] displays both the zoomed-in parts of the task's operational details and a scaled-down model of the robotic arm's digital twin from the user's perspective. Such a design allows the operator to observe both the local details and the overall motion of the robotic arm at the same time, which is not possible in reality. This design takes full advantage of the potential of XR and digital twins, and future designs could make efforts to explore XR and digital twin capabilities that cannot be realized in reality.

5.3 User-centered System Design

User-centered design is important for improving the user experience. A key research question is how to accommodate users with varying skill levels. Future research should focus on accessible and efficient systems by designing adaptive interfaces that adjust to user proficiency, reducing learning costs. For example, reinforcement learning can personalize interfaces based on individual skills and interests [142]. Different user groups have different needs: systems for professional engineers may prioritize precision and advanced features, while those for the general public may emphasize ease of use and intuitive controls [28]. Usage scenarios further shape design requirements; for example, remote surgical systems have different demands than remote assembly systems. Future research should clearly identify the needs of each user group and develop targeted designs accordingly.

In addition, accessibility requires that the system be designed to accommodate diverse user groups, such as people with disabilities, older adults, and children. For example, XR-based teleoperation may create new employment opportunities for people with mobility impairments, provided that the interfaces support alternative input methods (e.g., voice commands or eye tracking) [95]. Utilizing XR to remotely operate anthropomorphic robots inside the domestic setting (e.g., re-imagining Robot Design [136]) may facilitate remote engagement with family members and alleviate loneliness in both the elderly and children [165]. To make these systems intuitive for the elderly, designers might implement larger text displays, voice guidance, or simplified controls [72]. Future researchers may consider participatory design methods [125] that actively involve end users in the design process to ensure that the system meets their needs and preferences.

Lastly, the innovative potential of XR should be considered in system design. For example, with the increasing popularity of home robots, users who operate a home robot using XR might interact with an avatar that could be a small animal or even a human, rather than a traditional robot [38, 122]. Future research should probe innovative methods of harnessing XR's potential to elevate the user experience. This could involve the exploration of novel interaction techniques, the development of immersive feedback systems, or the invention of unique visualization methods.

6 Conclusion

This survey reviews 100 related studies from six dimensions to explore the application of XR in remote HRI, addressing the key research question of how to create immersive, efficient, and user-centered XR-based remote HRI systems. We analyze the impact of different robot types on system design, how to tailor systems to various user needs and scenarios, and how to fully leverage XR's potential. Additionally, we identify key research gaps, including appropriate system evaluation methods, multi-user or multi-robot interactions, and issues related to latency and real-time performance. Current system designs often fail to fully utilize XR's capabilities, highlighting the need for further research on user-friendly remote HRI XR systems. Our insights provide valuable resources for researchers, practitioners, and system designers aiming to optimize remote HRI. Stakeholders can use our findings to customize XR interfaces for specific robot types, user groups, and XR devices. By adapting our recommendations to their individual environments, from industrial automation to home or social robotics, researchers and system designers can improve the usability, safety, and efficiency of XR-supported remote HRI applications.

Acknowledgments

This research was supported by the Hong Kong Polytechnic University's Start-up Fund for New Recruits (Project ID: P0046056). Xian Wang was supported by a grant from the PolyU Research Committee (code RMHD).

References

[1] Mahmoud Abdulsalam and Nabil Aouf. 2023. VitRob Pipeline: A Seamless Teleoperation Pipeline for Advanced Virtual Reality-Robot Interface Applied for Precision Agriculture. In *2023 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE, 1–6.

[2] Dmytro Adamenko, Steffen Kunnen, Robin Pluhnau, André Loibl, and Arun Nagarajah. 2020. Review and comparison of the methods of designing the Digital Twin. *Procedia CIRP* 91 (2020), 27–32.

[3] Zhuming Ai, Mark A Livingston, and Ira S Moskowitz. 2016. Real-time unmanned aerial vehicle 3D environment exploration in a mixed reality environment. In *2016 International Conference on Unmanned Aircraft Systems (ICUAS)*. IEEE, 664–670.

[4] Ian F Akyildiz and Hongzhi Guo. 2022. Wireless communication research challenges for extended reality (XR). *ITU Journal on Future and Evolving Technologies* 3, 1 (2022), 1–15.

[5] Rachel Albert, Anjul Patney, David Luebke, and Joohwan Kim. 2017. Latency requirements for foveated rendering in virtual reality. *ACM Transactions on Applied Perception (TAP)* 14, 4 (2017), 1–13.

[6] Haider AF Almurib, Haidar Fadhil Al-Qrimli, and Nandha Kumar. 2012. A review of application industrial robotic design. In *2011 Ninth International Conference on ICT and Knowledge Engineering*. IEEE, 105–112.

[7] Doris Aschenbrenner, Meng Li, Radoslaw Dukalski, Jouke Verlinden, and Stephan Lukosch. 2018. Collaborative production line planning with augmented fabrication. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 509–510.

[8] Xue Bai, Changqiang Li, Keyan Chen, Yongjie Feng, Zhaowei Yu, and Ming Xu. 2018. Kinect-based hand tracking for first-person-perspective robotic arm teleoperation. In *2018 IEEE International Conference on Information and Automation (ICIA)*. IEEE, 684–691.

[9] Karlin Bark, Cuong Tran, Kikuo Fujimura, and Victor Ng-Thow-Hing. 2014. Personal navi: Benefits of an augmented reality navigational aid using a see-thru 3d volumetric hud. In *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. 1–8.

[10] Zahraa Bassyouni and Imad H Elhajj. 2021. Augmented reality meets artificial intelligence in robotics: A systematic review. *Frontiers in Robotics and AI* (2021), 296.

[11] Steve Benford, Chris Greenhalgh, Tom Rodden, and James Pycock. 2001. Collaborative virtual environments. *Commun. ACM* 44, 7 (2001), 79–85.

[12] Julio Betancourt, Baptiste Wojtkowski, Pedro Castillo, and Indira Thouvenin. 2022. Exocentric control scheme for robot applications: An immersive virtual reality approach. *IEEE Transactions on Visualization and Computer Graphics* 29, 7 (2022), 3392–3404.

[13] Feifei Bian, Ruifeng Li, Lijun Zhao, Yihuan Liu, and Peidong Liang. 2018. Interface design of a human-robot interaction system for dual-manipulators teleoperation based on virtual reality. In *2018 IEEE International Conference on Information and Automation (ICIA)*. IEEE, 1361–1366.

[14] Frank Biocca, Arthur Tang, Charles Owen, and Fan Xiao. 2006. Attention funnel: omnidirectional 3D cursor for mobile augmented reality platforms. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. 1115–1122.

[15] David Black, Yas Oloumi Yazdi, Amir Hossein Hadi Hosseinabadi, and Septimiu Salcudean. 2023. Human teleoperation-a haptically enabled mixed reality system for teleultrasound. *Human–Computer Interaction* (2023), 1–24.

[16] Felix Bork, Christian Schnelzer, Ulrich Eck, and Nassir Navab. 2018. Towards Efficient Visual Guidance in Limited Field-of-View Head-Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics* 24, 11 (2018), 2983–2992.

[17] Doug A Bowman, Joseph L Gabbard, and Deborah Hix. 2002. A survey of usability evaluation in virtual environments: classification and comparison of methods. *Presence: Teleoperators & Virtual Environments* 11, 4 (2002), 404–424.

[18] Filippo Brizzi, Lorenzo Peppoloni, Alessandro Graziano, Erika Di Stefano, Carlo Alberto Avizzano, and Emanuele Ruffaldi. 2017. Effects of augmented reality on the performance of teleoperated industrial assembly tasks in a robotic embodiment. *IEEE Transactions on Human-Machine Systems* 48, 2 (2017), 197–206.

[19] John Brooke. 1996. Sus: a "quick and dirty'usability. *Usability evaluation in industry* 189, 3 (1996), 189–194.

[20] Grigore C Burdea and Philippe Coiffet. 2003. *Virtual reality technology*. John Wiley & Sons.

[21] Yuanzhi Cao, Tianyi Wang, Xun Qian, Pawan S. Rao, Manav Wadhawan, Ke Huo, and Karthik Ramani. 2019. GhostAR: A Time-space Editor for Embodied Authoring of Human-Robot Collaborative Task with Augmented Reality. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 521–534.

[22] Sonia Mary Chacko, Armando Granado, Ashwin RajKumar, and Vikram Kapila. 2020. An augmented reality spatial referencing system for mobile robots. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 4446–4452.

[23] Junshen Chen, Marc Glover, Chenguang Yang, Chunxu Li, Zhijun Li, and Angelo Cangelosi. 2017. Development of an immersive interface for robot teleoperation. In *Towards Autonomous Robotic Systems: 18th Annual Conference, TAROS 2017, Guildford, UK, July 19–21, 2017, Proceedings 18*. Springer, 1–15.

[24] Yang Chen, Leyuan Sun, Mehdi Benallegue, Rafael Cisneros-Limón, Rohan P Singh, Kenji Kaneko, Arnaud Tanguy, Guillaume Caron, Kenji Suzuki, Abderrahmane Kheddar, et al. 2022. Enhanced visual feedback with decoupled viewpoint control in immersive humanoid robot teleoperation using slam. In *2022 IEEE-RAS 21st International Conference on Humanoid Robots (Humanoids)*. IEEE, 306–313.

[25] Yi Chen, Baohua Zhang, Jun Zhou, and Kai Wang. 2020. Real-time 3D unstructured environment reconstruction utilizing VR and Kinect-based immersive teleoperation for agricultural field robots. *Computers and Electronics in Agriculture* 175 (2020), 105579.

[26] Sung Ho Choi, Kyeong-Beom Park, Dong Hyeon Roh, Jae Yeol Lee, Yalda Ghasemi, and Heejin Jeong. 2022. An xr-based approach to safe human-robot collaboration. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 481–482.

[27] Torben Cichon and Jürgen Roßmann. 2018. Digital twins: assisting and supporting cooperation in human-robot teams. In *2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV)*. IEEE, 486–491.

[28] Andy Cockburn, Carl Gutwin, Joey Scarr, and Sylvain Malacria. 2014. Supporting novice to expert transitions in user interfaces. *ACM Computing Surveys (CSUR)* 47, 2 (2014), 1–36.

[29] Andy Cockburn, Amy Karlson, and Benjamin B Bederson. 2009. A review of overview+ detail, zooming, and focus+ context interfaces. *ACM Computing Surveys (CSUR)* 41, 1 (2009), 1–31.

[30] Robert Codd-Downey, P Mojiri Forooshani, Andrew Speers, Hui Wang, and Michael Jenkin. 2014. From ROS to unity: Leveraging robot and virtual environment middleware for immersive teleoperation. In *2014 IEEE International Conference on Information and Automation (ICIA)*. IEEE, 932–936.

[31] A. Coninx, P. Baxter, E. Oleari, S. Bellini, B. Bierman, O. B. Henkemans, L. Cañamero, P. Cosi, V. Enescu, R. R. Espinoza, A. Hiolle, R. Humbert, B. Kiefer, I. Kruijff-Korbayová, R. Looije, M. Mosconi, M. A. Neerincx, G. Paci, G. Patsis, C. Pozzi, F. Sacchitelli, H. Sahli, A. Sanna, G. Sommavilla, F. Tesser, Y. Demiris, and T. Belpaeme. 2016. Towards long-term social child-robot interaction: using multi-activity switching to engage young users. *Journal of Human-Robot Interaction* 5 (2016), 32. Issue 1.

[32] Gabriel de Moura Costa, Marcelo Roberto Petry, and António Paulo Moreira. 2022. Augmented Reality for Human–Robot Collaboration and Cooperation in Industrial Applications: A Systematic Literature Review. *Sensors* 22, 7 (2022), 2725.

[33] Matthew Cousins, Chenguang Yang, Junshen Chen, Wei He, and Zhaojie Ju. 2017. Development of a mixed reality based interface for human robot interaciotn. In *2017 International Conference on Machine Learning and Cybernetics (ICMLC)*, Vol. 1. IEEE, 27–34.

[34] Christyan Cruz Ulloa, David Domínguez, Jaime Del Cerro, and Antonio Barrientos. 2022. A Mixed-Reality Tele-Operation Method for High-Level Control of a Legged-Manipulator Robot. *Sensors* 22, 21 (2022), 8146.

[35] Mark R Cutkosky. 2012. *Robotic grasping and fine manipulation*. Vol. 6. Springer Science & Business Media.

[36] Francesco De Pace, Federico Manuri, Andrea Sanna, and Claudio Fornaro. 2020. A systematic review of Augmented Reality interfaces for collaborative industrial robots. *Computers & Industrial Engineering* 149 (2020), 106806.

[37] Morteza Dianatfar, Jyrki Latokartano, and Minna Lanz. 2021. Review on existing VR/AR solutions in human–robot collaboration. *Procedia CIRP* 97 (2021), 407–411.

[38] Mauro Dragone, Thomas Holz, and G.M.P. O'Hare. 2007. Using Mixed Reality Agents as Social Interfaces for Robots. In *RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication*. 1161–1166.

[39] Guanglong Du, Yongda Deng, Wing WY Ng, and Di Li. 2022. An Intelligent Interaction Framework for Teleoperation Based on Human-Machine Cooperation. *IEEE Transactions on Human-Machine Systems* 52, 5 (2022), 963–972.

[40] Guanglong Du, Ruiguang Han, Gengcheng Yao, Wing WY Ng, and Di Li. 2021. A gesture-and speech-guided robot teleoperation method based on mobile interaction with unrestricted force feedback. *IEEE/ASME Transactions on Mechatronics* 27, 1 (2021), 360–371.

[41] Guanglong Du, Bo Zhang, Chunquan Li, and Hua Yuan. 2019. A novel natural mobile human-machine interaction method with augmented reality. *IEEE Access* 7 (2019), 154317–154330.

[42] Shirine El Zaatari, Mohamed Marei, Weidong Li, and Zahid Usman. 2019. Cobot programming for collaborative industrial tasks: An overview. *Robotics and Autonomous Systems* 116 (2019), 162–180.

[43] Mohammed S Elbamby, Cristina Perfecto, Mehdi Bennis, and Klaus Doppler. 2018. Toward low-latency and ultra-reliable virtual reality. *IEEE network* 32, 2 (2018), 78–84.

[44] Wen Fan, Xiaoqing Guo, Enyang Feng, Jialin Lin, Yuanyi Wang, Jiaming Liang, Martin Garrad, Jonathan Rossiter, Zhengyou Zhang, Nathan Lepora, et al. 2023. Digital Twin-Driven Mixed Reality Framework for Immersive Teleoperation With Haptic Rendering. *IEEE Robotics and Automation Letters* (2023).

[45] Simone Fani, Simone Ciotti, Manuel G Catalano, Giorgio Grioli, Alessandro Tognetti, Gaetano Valenza, Arash Ajoudani, and Matteo Bianchi. 2018. Simplifying telerobotics: Wearability and teleimpedance improves human-robot interactions in teleoperation. *IEEE Robotics & Automation Magazine* 25, 1 (2018), 77–88.

[46] Simone Fani, Simone Ciotti, Manuel G Catalano, Giorgio Grioli, Alessandro Tognetti, Gaetano Valenza, Arash Ajoudani, and Matteo Bianchi. 2018. Simplifying telerobotics: Wearability and teleimpedance improves human-robot interactions in teleoperation. *IEEE Robotics & Automation Magazine* 25, 1 (2018), 77–88.

[47] Jesse Fox and Andrew Gambino. 2021. Relationship development with humanoid social robots: Applying interpersonal theories to human–robot interaction. *Cyberpsychology, Behavior, and Social Networking* 24, 5 (2021), 294–299.

[48] Kodai Fuchino, Mohammed Al-Sada, and Tatsuo Nakajima. 2023. T2Remoter: a Remote Table Tennis Coaching System Combining VR and Robotics. In *2023 IEEE 29th International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA)*. IEEE, 275–276.

[49] Markus Funk, Mareike Kritzler, and Florian Michahelles. 2017. HoloLens is more than air Tap: natural and intuitive interaction with holograms. In *Proceedings of the seventh international conference on the internet of things*. 1–2.

[50] Wei Gai, Huiyu Li, Yanshuai Zhao, Maiwang Shi, Wenfei Wang, Yunchuan Sun, and Chenglei Yang. 2020. A New Navigation Method for VR-based Telerobotic System via Supervisor's Real Walking in a Limited Physical Space. In *2020 International Wireless Communications and Mobile Computing (IWCMC)*. IEEE, 494–498.

[51] Péter Galambos, Ádám Csapó, Péter Zentay, István Marcell Fülöp, Tamás Haidegger, Péter Baranyi, and Imre J Rudas. 2015. Design, programming and orchestration of heterogeneous manufacturing systems through VR-powered remote collaboration. *Robotics and Computer-Integrated Manufacturing* 33 (2015), 68–77.

[52] Luigi Gammieri, Marco Schumann, Luigi Pelliccia, Giuseppe Di Gironimo, and Philipp Klimant. 2017. Coupling of a redundant manipulator with a virtual reality environment to enhance human-robot cooperation. *Procedia Cirp* 62 (2017), 618–623.

[53] Moojan Ghafurian, Shruti Chandra, Rebecca Hutchinson, Angelica Lim, Ishan Baliyan, Jimin Rhim, Garima Gupta, Alexander M. Aroyo, Samira Rasouli, and Kerstin Dautenhahn. 2024. Systematic Review of Social Robots for Health and Wellbeing: A Personal Healthcare Journey Lens. *J. Hum.-Robot Interact.* 14, 1, Article 19 (Dec. 2024), 48 pages.

[54] Vicent Girbes-Juan, Vinicius Schettino, Yiannis Demiris, and Josep Tornero. 2020. Haptic and visual feedback assistance for dual-arm robot teleoperation in surface conditioning tasks. *IEEE Transactions on Haptics* 14, 1 (2020), 44–56.

[55] Liang Gong, Changyang Gong, Zhao Ma, Lujie Zhao, Zhenyu Wang, Xudong Li, Xiaolong Jing, Haozhe Yang, and Chengliang Liu. 2017. Real-time human-in-the-loop remote control for a life-size traffic police robot with multiple augmented reality aided display terminals. In *2017 2nd International conference on advanced robotics and mechatronics (ICARM)*. IEEE, 420–425.

[56] Florenz Graf, Çağatay Odabaşi, Theo Jacobs, Birgit Graf, and Thomas Födisch. 2019. Mobika-low-cost mobile robot for human-robot interaction. In *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 1–6.

[57] Scott A Green, Mark Billinghurst, XiaoQi Chen, and J Geoffrey Chase. 2007. Human robot collaboration: An augmented reality approach—a literature review and analysis. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Vol. 48051. 117–126.

[58] Uwe Gruenefeld, Daniel Lange, Lasse Hammer, Susanne Boll, and Wilko Heuten. 2018. FlyingARrow: Pointing Towards Out-of-View Objects on Augmented Reality Devices. In *Proceedings of the 7th ACM International Symposium on Pervasive Displays* (Munich, Germany) *(PerDis '18)*. Association for Computing Machinery, New York, NY, USA, Article 20, 6 pages.

[59] Fabien Grzeskowiak, Marie Babel, Julien Bruneau, and Julien Pettré. 2020. Toward virtual reality-based evaluation of robot navigation among people. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 766–774.

[60] Vehbi C Gungor and Gerhard P Hancke. 2009. Industrial wireless sensor networks: Challenges, design principles, and technical approaches. *IEEE Transactions on industrial electronics* 56, 10 (2009), 4258–4265.

[61] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.

[62] Hooman Hedayati, Michael Walker, and Daniel Szafir. 2018. Improving collocated robot teleoperation with augmented reality. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. 78–86.

[63] Juan David Hernández, Shlok Sobti, Anthony Sciola, Mark Moll, and Lydia E Kavraki. 2020. Increasing robot autonomy via motion planning and an augmented reality interface. *IEEE Robotics and Automation Letters* 5, 2 (2020), 1017–1023.

[64] Wolfgang Hoenig, Christina Milanes, Lisa Scaria, Thai Phan, Mark Bolas, and Nora Ayanian. 2015. Mixed reality for robotics. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 5382–5387.

[65] Leire Amezua Hormaza, Wael M Mohammed, Borja Ramis Ferrer, Ronal Bejarano, and Jose L Martinez Lastra. 2019. On-line training and monitoring of robot tasks through virtual reality. In *2019 IEEE 17th International Conference on Industrial Informatics (INDIN)*, Vol. 1. IEEE, 841–846.

[66] Ji Hou, Benjamin Graham, Matthias Nießner, and Saining Xie. 2021. Exploring data-efficient 3d scene understanding with contrastive scene contexts. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 15587–15597.

[67] Ziqi Huang, Yang Shen, Jiayi Li, Marcel Fey, and Christian Brecher. 2021. A survey on AI-driven digital twins in industry 4.0: Smart manufacturing and advanced robotics. *Sensors* 21, 19 (2021), 6340.

[68] Keiichi Ihara, Mehrad Faridan, Ayumi Ichikawa, Ikkaku Kawaguchi, and Ryo Suzuki. 2023. HoloBots: Augmenting Holographic Telepresence with Mobile Robots for Tangible Remote Collaboration in Mixed Reality. In *Proceedings of the 36th Annual ACM*

Symposium on User Interface Software and Technology. 1–12.
[69] Inmo Jang, Junyan Hu, Farshad Arvin, Joaquin Carrasco, and Barry Lennox. 2021. Omnipotent virtual giant for remote human–swarm interaction. In *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*. IEEE, 488–494.
[70] Inmo Jang, Hanlin Niu, Emily C Collins, Andrew Weightman, Joaquin Carrasco, and Barry Lennox. 2021. Virtual kinesthetic teaching for bimanual telemanipulation. In *2021 IEEE/SICE International Symposium on System Integration (SII)*. IEEE, 120–125.
[71] Caroline Jay, Mashhuda Glencross, and Roger Hubbold. 2007. Modeling the effects of delayed haptic and visual feedback in a collaborative virtual environment. *ACM Transactions on Computer-Human Interaction (TOCHI)* 14, 2 (2007), 8–es.
[72] Xiaofu Jin, Wai Tong, Xiaoying Wei, Xian Wang, Emily Kuang, Xiaoyu Mo, Huamin Qu, and Mingming Fan. 2024. Exploring the Opportunity of Augmented Reality (AR) in Supporting Older Adults to Explore and Learn Smartphone Applications. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 21, 18 pages.
[73] Ivan Kalinov, Daria Trinitatova, and Dzmitry Tsetserukou. 2021. Warevr: Virtual reality interface for supervision of autonomous robotic system aimed at warehouse stocktaking. In *2021 ieee international conference on systems, man, and cybernetics (smc)*. IEEE, 2139–2145.
[74] Kotaro Kanazawa, Noritaka Sato, and Yoshifumi Morita. 2023. Considerations on Interaction with Manipulator in Virtual Reality Teleoperation Interface for Rescue Robots. In *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 386–391.
[75] Yaesol Kim, Myrna Citlali Castillo Silva, Sara Anastasi, and Nikhil Deshpande. 2023. Towards Immersive Bilateral Teleoperation Using Encountered-Type Haptic Interface. In *2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 1354–1359.
[76] Akemi Kobayashi, Ryosuke Aoki, Norimichi Kitagawa, Toshitaka Kimura, Youichi Takashima, and Tomohiro Yamada. 2016. Towards enhancing force-input interaction by visual-auditory feedback as an introduction of first use. In *Human-Computer Interaction. Interaction Platforms and Techniques: 18th International Conference, HCI International 2016, Toronto, ON, Canada, July 17-22, 2016. Proceedings, Part II 18*. Springer, 180–191.
[77] Roman Kulikovskiy, Megan Sochanski, Matteson Eaton, Jessica Korneder, Wing-Yue Geoffrey Louie, et al. 2021. Can therapists design robot-mediated interventions and teleoperate robots using VR to deliver interventions for ASD?. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 3669–3676.
[78] Chen-Yu Kuo, Chun-Chi Huang, Chih-Hsuan Tsai, Yun-Shuo Shi, and Shana Smith. 2021. Development of an immersive SLAM-based VR system for teleoperation of a mobile manipulator in an unknown environment. *Computers in Industry* 132 (2021), 103502.
[79] Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and evaluation of a user experience questionnaire. In *HCI and Usability for Education and Work: 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society, USAB 2008, Graz, Austria, November 20-21, 2008. Proceedings 4*. Springer, 63–76.
[80] Dinh Tung Le, Sheila Sutjipto, Yujun Lai, and Gavin Paul. 2020. Intuitive virtual reality based control of a real-world mobile manipulator. In *2020 16th International Conference on Control, Automation, Robotics and Vision (ICARCV)*. IEEE, 767–772.
[81] Ju-Hwan Lee and Charles Spence. 2008. Assessing the benefits of multimodal feedback on dual-task performance under demanding conditions. *People and Computers XXII Culture, Creativity, Interaction 22* (2008), 185–192.
[82] Lik-Hang Lee, Tristan Braud, Peng Yuan Zhou, Lin Wang, Dianlei Xu, Zijun Lin, Abhishek Kumar, Carlos Bermejo, Pan Hui, et al. 2024. All one needs to know about metaverse: A complete survey on technological singularity, virtual ecosystem, and research agenda. *Foundations and trends® in human-computer interaction* 18, 2–3 (2024), 100–337.
[83] Zhijun Li, Cuichao Xu, Qiang Wei, Chao Shi, and Chun-Yi Su. 2018. Human-inspired control of dual-arm exoskeleton robots with force and impedance adaptation. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 50, 12 (2018), 5296–5305.
[84] Jeffrey I. Lipton, Aidan J. Fay, and Daniela Rus. 2017. Baxter's homunculus: Virtual reality spaces for teleoperation in manufacturing. *IEEE Robotics and Automation Letters* 3, 1 (2017), 179–186.
[85] Di Liu, Jeonghee Kim, and Youngjib Ham. 2023. Multi-user immersive environment for excavator teleoperation in construction. *Automation in Construction* 156 (2023), 105143.
[86] Hongyi Liu and Lihui Wang. 2020. Remote human–robot collaboration: A cyber–physical system application for hazard manufacturing environment. *Journal of manufacturing systems* 54 (2020), 24–34.
[87] Xin Liu, Du Jiang, Bo Tao, Guozhang Jiang, Ying Sun, Jianyi Kong, Xiliang Tong, Guojun Zhao, and Baojia Chen. 2022. Genetic algorithm-based trajectory optimization for digital twin robots. *Frontiers in Bioengineering and Biotechnology* 9 (2022), 793782.
[88] Praveen Kumar Reddy Maddikunta, Quoc-Viet Pham, B Prabadevi, Natarajan Deepa, Kapal Dev, Thippa Reddy Gadekallu, Rukhsana Ruby, and Madhusanka Liyanage. 2022. Industry 5.0: A survey on enabling technologies and potential applications. *Journal of Industrial Information Integration* 26 (2022), 100257.
[89] Zhanat Makhataeva and Huseyin Atakan Varol. 2020. Augmented reality for robotics: A review. *Robotics* 9, 2 (2020), 21.
[90] Steve Mann. 1999. Mediated Reality. *Linux J.* 1999, 59es (March 1999), 5–es.
[91] Jeremy A. Marvel, Shelly Bagchi, Megan Zimmerman, and Brian Antonishek. 2020. Towards Effective Interface Designs for Collaborative HRI in Manufacturing: Metrics and Measures. *J. Hum.-Robot Interact.* 9, 4, Article 25 (May 2020), 55 pages.

[92] Neil McHenry, Jason Spencer, Patrick Zhong, Jeremy Cox, Michael Amiscaray, KC Wong, and Gregory Chamitoff. 2021. Predictive xr telepresence for robotic operations in space. In *2021 IEEE Aerospace Conference (50100)*. IEEE, 1–10.

[93] Lingxiao Meng, Jiangshan Liu, Wei Chai, Jiankun Wang, and Max Q-H Meng. 2023. Virtual Reality Based Robot Teleoperation via Human-Scene Interaction. *Procedia Computer Science* 226 (2023), 141–148.

[94] Mohammad Mehdi Moniri, Fabio Andres Espinosa Valcarcel, Dieter Merkel, and Daniel Sonntag. 2016. Human gaze and focus-of-attention in dual reality human-robot collaboration. In *2016 12th International Conference on Intelligent Environments (IE)*. IEEE, 238–241.

[95] Martez Mott, John Tang, Shaun Kane, Edward Cutrell, and Meredith Ringel Morris. 2020. "I just went into it assuming that I wouldn't be able to have the full experience": Understanding the Accessibility of Virtual Reality for People with Limited Mobility. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility* (Virtual Event, Greece) *(ASSETS '20)*. Association for Computing Machinery, New York, NY, USA, Article 43, 13 pages.

[96] Dimitris Mourtzis, Vasilios Zogopoulos, and E Vlachou. 2017. Augmented reality application to support remote maintenance as a service in the robotics industry. *Procedia Cirp* 63 (2017), 46–51.

[97] Dimitris Mourtzis, Vasilios Zogopoulos, and Fotini Xanthi. 2019. Augmented reality application to support the assembly of highly customized products and to adapt to production re-scheduling. *The International Journal of Advanced Manufacturing Technology* 105 (2019), 3899–3910.

[98] Hengyang Mu, Yifei Li, Diansheng Chen, Jiting Li, and Min Wang. 2021. Design of Tank Inspection Robot Navigation System Based on Virtual Reality. In *2021 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE, 1773–1778.

[99] Bálint György Nagy, János Dóka, Sáandor Rácz, Géza Szabó, István Pelle, János Czentye, László Toka, and Balázs Sonkoly. 2019. Towards human-robot collaboration: An industry 4.0 vr platform with clouds under the hood. In *2019 IEEE 27th International Conference on Network Protocols (ICNP)*. IEEE, 1–2.

[100] P Nandhini, P Chellammal, JS Jaslin, S Harthy Ruby Priya, M Uma, and R Kaviyaraj. 2023. Teleoperation in the Age of Mixed Reality: VR, AR, and ROS Integration for Human-Robot Direct Interaction. In *2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC)*. IEEE, 240–245.

[101] Federica Nenna, Davide Zanardi, and Luciano Gamberini. 2023. Enhanced Interactivity in VR-based Telerobotics: An Eye-tracking Investigation of Human Performance and Workload. *International Journal of Human-Computer Studies* 177 (2023), 103079.

[102] Bukeikhan Omarali, Brice Denoun, Kaspar Althoefer, Lorenzo Jamone, Maurizio Valle, and Ildar Farkhatdinov. 2020. Virtual reality based telerobotics framework with depth cameras. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 1217–1222.

[103] Claudio Pacchierotti, Asad Tirmizi, and Domenico Prattichizzo. 2014. Improving transparency in teleoperation by means of cutaneous tactile force feedback. *ACM Transactions on Applied Perception (TAP)* 11, 1 (2014), 1–16.

[104] Sungman Park, Yeongtae Jung, and Joonbum Bae. 2018. An interactive and intuitive control interface for a tele-operated robot (AVATAR) system. *Mechatronics* 55 (2018), 54–62.

[105] Anjul Patney, Marco Salvi, Joohwan Kim, Anton Kaplanyan, Chris Wyman, Nir Benty, David Luebke, and Aaron Lefohn. 2016. Towards foveated rendering for gaze-tracked virtual reality. *ACM Transactions on Graphics (TOG)* 35, 6 (2016), 1–12.

[106] Michael Quinn Patton. 2002. *Qualitative research & evaluation methods*. sage.

[107] Lorenzo Peppoloni, Filippo Brizzi, Carlo Alberto Avizzano, and Emanuele Ruffaldi. 2015. Immersive ROS-integrated framework for robot teleoperation. In *2015 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 177–178.

[108] Lorenzo Peppoloni, Filippo Brizzi, Emanuele Ruffaldi, and Carlo Alberto Avizzano. 2015. Augmented reality-aided tele-presence system for robot manipulation in industrial manufacturing. In *Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology*. 237–240.

[109] Ornnalin Phaijit, Mohammad Obaid, Claude Sammut, and Wafa Johal. 2022. A Taxonomy of Functional Augmented Reality for Human-Robot Interaction. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction*. 294–303.

[110] Thai Phan, Wolfgang Hönig, and Nora Ayanian. 2018. Mixed reality collaboration between human-agent teams. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 659–660.

[111] Andoni Rivera Pinto, Johan Kildal, and Elena Lazkano. 2020. Multimodal mixed reality impact on a hand guiding task with a holographic cobot. *Multimodal Technologies and Interaction* 4, 4 (2020), 78.

[112] Polina Ponomareva, Daria Trinitatova, Aleksey Fedoseev, Ivan Kalinov, and Dzmitry Tsetserukou. 2021. Grasplook: a vr-based telemanipulation system with r-cnn-driven augmentation of virtual environment. In *2021 20th International Conference on Advanced Robotics (ICAR)*. IEEE, 166–171.

[113] Jorge Pena Queralta, Jussi Taipalmaa, Bilge Can Pullinen, Victor Kathan Sarker, Tuan Nguyen Gia, Hannu Tenhunen, Moncef Gabbouj, Jenni Raitoharju, and Tomi Westerlund. 2020. Collaborative multi-robot systems for search and rescue: Coordination and perception. *arXiv preprint arXiv:2008.12610* (2020).

[114] Nicolaus A Radford, Philip Strawser, Kimberly Hambuchen, Joshua S Mehling, William K Verdeyen, A Stuart Donnan, James Holley, Jairo Sanchez, Vienny Nguyen, Lyndon Bridgwater, et al. 2015. Valkyrie: Nasa's first bipedal humanoid robot. *Journal of Field Robotics*

32, 3 (2015), 397–419.

[115] Naveen Rastogi and Amit Kumar Srivastava. 2019. Control system design for tokamak remote maintenance operations using assisted virtual reality and haptic feedback. *Fusion engineering and design* 139 (2019), 47–54.

[116] Philipp A Rauschnabel, Reto Felix, Chris Hinsch, Hamza Shahab, and Florian Alt. 2022. What is XR? Towards a framework for augmented and virtual reality. *Computers in human behavior* 133 (2022), 107289.

[117] Andoni Rivera-Pinto, Johan Kildal, and Elena Lazkano. 2023. Toward Programming a Collaborative Robot by Interacting with Its Digital Twin in a Mixed Reality Environment. *International Journal of Human–Computer Interaction* (2023), 1–13.

[118] Giulio Rosati, Antonio Rodà, Federico Avanzini, and Stefano Masiero. 2013. On the role of auditory feedback in robot-assisted movement training after stroke: review of the literature. *Computational intelligence and neuroscience* 2013 (2013), 11–11.

[119] Mose Sakashita, Hyunju Kim, Brandon Woodard, Ruidong Zhang, and François Guimbretière. 2023. VRoxy: Wide-Area Collaboration From an Office Using a VR-Driven Robotic Proxy. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–13.

[120] Filippo Sanfilippo, Jesper Smith, Sylvain Bertrand, and Tor Halvard Skarberg Svendsen. 2022. Mixed reality (MR) Enabled Proprio and Teleoperation of a Humanoid Robot for Paraplegic Patients. In *2022 5th International Conference on Information and Computer Technologies (ICICT)*. IEEE, 153–158.

[121] Alexander Schäfer, Gerd Reis, and Didier Stricker. 2021. A Survey on Synchronous Augmented, Virtual and Mixed Reality Remote Collaboration Systems. *ACM Computing Surveys (CSUR)* (2021).

[122] Jürgen Scheible, Achim Hoth, Julian Saal, and Haifeng Su. 2013. Displaydrone: a flying robot based interactive display. In *Proceedings of the 2nd ACM International Symposium on Pervasive Displays*. 49–54.

[123] Michael Schluse, Marc Priggemeyer, Linus Atorf, and Juergen Rossmann. 2018. Experimentable digital twins—Streamlining simulation-based systems engineering for industry 4.0. *IEEE Transactions on industrial informatics* 14, 4 (2018), 1722–1731.

[124] Ralph Schroeder. 2010. *Being There Together: Social interaction in shared virtual environments.* Oxford University Press.

[125] Douglas Schuler and Aki Namioka. 1993. *Participatory design: Principles and practices.* CRC press.

[126] Max Schwarz, Christian Lenz, Andre Rochow, Michael Schreiber, and Sven Behnke. 2021. Nimbro avatar: Interactive immersive telepresence with force-feedback telemanipulation. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 5312–5319.

[127] Chung Xue Er Shamaine, Yuansong Qiao, John Henry, Ken McNevin, and Niall Murray. 2020. RoSTAR: ROS-based telerobotic control via augmented reality. In *2020 IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP)*. IEEE, 1–6.

[128] Thomas B Sheridan. 2016. Human–robot interaction: status and challenges. *Human factors* 58, 4 (2016), 525–532.

[129] Harvey Stedman, Basaran Bahadir Kocer, Mirko Kovac, and Vijay M Pawar. 2022. VRTAB-map: A configurable immersive teleoperation framework with online 3D reconstruction. In *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 104–110.

[130] Jannek Steinke, Justus Rischke, Peter Sossalla, Johannes Hofer, Christian L Vielhaus, Nico Vom Hofe, and HP Frank Fitzek. 2023. The Future of Dog Walking–Four-Legged Robots and Augmented Reality. In *2023 IEEE 24th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*. IEEE, 352–354.

[131] Franz Steinmetz, Annika Wollschläger, and Roman Weitschat. 2018. Razer—a hri for visual task-level programming and intuitive skill parameterization. *IEEE Robotics and Automation Letters* 3, 3 (2018), 1362–1369.

[132] Patrick Stotko, Stefan Krumpen, Max Schwarz, Christian Lenz, Sven Behnke, Reinhard Klein, and Michael Weinmann. 2019. A VR system for immersive teleoperation and live exploration with a mobile robot. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 3630–3637.

[133] Yunpeng Su, Leo Lloyd, Xiaoqi Chen, and J Geoffrey Chase. 2023. Latency mitigation using applied HMMs for mixed reality-enhanced intuitive teleoperation in intelligent robotic welding. *The International Journal of Advanced Manufacturing Technology* 126, 5 (2023), 2233–2248.

[134] Yun-Peng Su, Xiao-Qi Chen, Tony Zhou, Christopher Pretty, and Geoffrey Chase. 2021. Mixed reality-enhanced intuitive teleoperation with hybrid virtual fixtures for intelligent robotic welding. *Applied Sciences* 11, 23 (2021), 11280.

[135] Zhongda Sun, Minglu Zhu, Zhaocong Chen, Xuechuan Shan, and Chengkuo Lee. 2021. Haptic-feedback ring enabled human-machine interface (HMI) aiming at immersive virtual reality experience. In *2021 21st International Conference on Solid-State Sensors, Actuators and Microsystems (Transducers)*. IEEE, 333–336.

[136] Ryo Suzuki, Adnan Karim, Tian Xia, Hooman Hedayati, and Nicolai Marquardt. 2022. Augmented Reality and Robotics: A Survey and Taxonomy for AR-enhanced Human-Robot Interaction and Robotic Interfaces. In *CHI Conference on Human Factors in Computing Systems*. 1–33.

[137] Krzysztof Adam Szczurek, Raul Marin Prades, Eloise Matheson, Jose Rodriguez-Nogueira, and Mario Di Castro. 2023. Multimodal multi-user mixed reality human–robot interface for remote operations in hazardous environments. *IEEE Access* 11 (2023), 17305–17333.

[138] Barnabas Takacs, Gergely Richter, Klara Csizinszky, Daniele Mazzei, and Lajos Simon. 2015. Towards a unified control framework for humanoid robots and their virtual avatars in physical and virtual reality-based interactions. In *2015 15th International Conference on*

Control, Automation and Systems (ICCAS). IEEE, 1905–1909.
[139] Bahi Takkouche and Guy Norman. 2011. PRISMA statement. *Epidemiology* 22, 1 (2011), 128.
[140] Russell H Taylor, Arianna Menciassi, Gabor Fichtinger, Paolo Fiorini, and Paolo Dario. 2016. Medical robotics and computer-integrated surgery. *Springer handbook of robotics* (2016), 1657–1684.
[141] Konstantinos Theofilis, Jason Orlosky, Yukie Nagai, and Kiyoshi Kiyokawa. 2016. Panoramic view reconstruction for stereoscopic teleoperation of a humanoid robot. In *2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)*. IEEE, 242–248.
[142] Kashyap Todi, Gilles Bailly, Luis Leiva, and Antti Oulasvirta. 2021. Adapting user interfaces with model-based reinforcement learning. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–13.
[143] J Gregory Trafton, Nicholas L Cassimatis, Magdalena D Bugajska, Derek P Brock, Farilee E Mintz, and Alan C Schultz. 2005. Enabling effective human-robot interaction using perspective-taking in robots. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 35, 4 (2005), 460–470.
[144] Fernando Trejo and Yaoping Hu. 2018. User performance of VR-based dissection: direct mapping and motion coupling of a surgical tool. In *2018 IEEE international conference on systems, man, and cybernetics (SMC)*. IEEE, 3039–3044.
[145] Daria Trinitatova and Dzmitry Tsetserukou. 2023. Study of the Effectiveness of a Wearable Haptic Interface With Cutaneous and Vibrotactile Feedback for VR-Based Teleoperation. *IEEE Transactions on Haptics* (2023).
[146] Yusuke Ujitoko and Yuki Ban. 2021. Survey of pseudo-haptics: Haptic feedback design and application proposals. *IEEE Transactions on Haptics* 14, 4 (2021), 699–711.
[147] Balazs P. Vagvolgyi, Will Pryor, Ryan Reedy, Wenlong Niu, Anton Deguet, Louis L. Whitcomb, Simon Leonard, and Peter Kazanzides. 2018. Scene modeling and augmented virtuality interface for telerobotic satellite servicing. *IEEE Robotics and Automation Letters* 3, 4 (2018), 4241–4248.
[148] Radu-Daniel Vatavu. 2022. Sensorimotor Realities: Formalizing Ability-Mediating Design for Computer-Mediated Reality Environments. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 685–694.
[149] Anurag Sai Vempati, Harshit Khurana, Vojtech Kabelka, Simon Flueckiger, Roland Siegwart, and Paul Beardsley. 2019. A virtual reality interface for an autonomous spray painting UAV. *IEEE Robotics and Automation Letters* 4, 3 (2019), 2870–2877.
[150] Thanh Long Vu, Dac Dang Khoa Nguyen, Sheila Sutjipto, Dinh Tung Le, and Gavin Paul. 2022. Investigation of User Performance in Virtual Reality-based Annotation-assisted Remote Robot Control. In *Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology*. 1–2.
[151] Michael Walker, Zhaozhong Chen, Matthew Whitlock, David Blair, Danielle Albers Szafir, Christoffer Heckman, and Daniel Szafir. 2021. A mixed reality supervision and telepresence interface for outdoor field robotics. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2345–2352.
[152] Michael Walker, Hooman Hedayati, Jennifer Lee, and Daniel Szafir. 2018. Communicating Robot Motion Intent with Augmented Reality. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (Chicago, IL, USA) *(HRI '18)*. Association for Computing Machinery, New York, NY, USA, 316–324.
[153] Michael Walker, Thao Phung, Tathagata Chakraborti, Tom Williams, and Daniel Szafir. 2022. Virtual, augmented, and mixed reality for human-robot interaction: A survey and virtual design element taxonomy. *arXiv preprint arXiv:2202.11249* (2022).
[154] Michael E Walker, Hooman Hedayati, and Daniel Szafir. 2019. Robot teleoperation with augmented reality virtual surrogates. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 202–210.
[155] Thomas Waltemate, Irene Senna, Felix Hülsmann, Marieke Rohde, Stefan Kopp, Marc Ernst, and Mario Botsch. 2016. The impact of latency on perceptual judgments and motor performance in closed-loop interaction in virtual reality. In *Proceedings of the 22nd ACM conference on virtual reality software and technology*. 27–35.
[156] Ker-Jiun Wang, Caroline Yan Zheng, and Zhi-Hong Mao. 2019. Human-centered, ergonomic wearable device with computer vision augmented intelligence for VR multimodal human-smart home object interaction. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 767–768.
[157] Peng Wang, Xiaoliang Bai, Mark Billinghurst, Shusheng Zhang, Xiangyu Zhang, Shuxia Wang, Weiping He, Yuxiang Yan, and Hongyu Ji. 2021. AR/MR remote collaboration on physical tasks: A review. *Robotics and Computer-Integrated Manufacturing* 72 (2021), 102071.
[158] Pan Wang, Junhao Xiao, Huimin Lu, Hui Zhang, Ruoyi Yan, and Shaozun Hong. 2017. A novel human-robot interaction system based on 3D mapping and virtual reality. In *2017 Chinese Automation Congress (CAC)*. IEEE, 5888–5894.
[159] Qiyue Wang, Yongchao Cheng, Wenhua Jiao, Michael T Johnson, and YuMing Zhang. 2019. Virtual reality human-robot collaborative welding: A case study of weaving gas tungsten arc welding. *Journal of Manufacturing Processes* 48 (2019), 210–217.
[160] Qiyue Wang, Wenhua Jiao, Peng Wang, and YuMing Zhang. 2020. Digital twin for human-robot interactive welding and welder behavior analysis. *IEEE/CAA Journal of Automatica Sinica* 8, 2 (2020), 334–343.
[161] Qiyue Wang, Wenhua Jiao, Rui Yu, Michael T Johnson, and YuMing Zhang. 2019. Modeling of human welders' operations in virtual reality human–robot interaction. *IEEE Robotics and automation letters* 4, 3 (2019), 2958–2964.
[162] Xian Wang, Diego Monteiro, Lik-Hang Lee, Pan Hui, and Hai-Ning Liang. 2022. Vibroweight: Simulating weight and center of gravity changes of objects in virtual reality for enhanced realism. In *2022 IEEE haptics symposium (HAPTICS)*. IEEE, 1–7.

[163] Ziheng Wang, Isabella Reed, and Ann Majewicz Fey. 2018. Toward intuitive teleoperation in surgery: Human-centric evaluation of teleoperation algorithms for robotic needle steering. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 5799–5806.
[164] Dong Wei, Bidan Huang, and Qiang Li. 2021. Multi-view merging for robot teleoperation with virtual reality. *IEEE Robotics and Automation Letters* 6, 4 (2021), 8537–8544.
[165] Xiaoying Wei, Yizheng Gu, Emily Kuang, Xian Wang, Beiyan Cao, Xiaofu Jin, and Mingming Fan. 2023. Bridging the Generational Gap: Exploring How Virtual Reality Supports Remote Communication Between Grandparents and Grandchildren. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 444, 15 pages.
[166] David Weintrop, Afsoon Afzal, Jean Salac, Patrick Francis, Boyang Li, David C Shepherd, and Diana Franklin. 2018. Evaluating CoBlox: A comparative study of robotics programming environments for adult novices. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–12.
[167] S Wibowo, I Siradjuddin, F Ronilaya, and MN Hidayat. 2021. Improving teleoperation robots performance by eliminating view limit using 360 camera and enhancing the immersive experience utilizing VR headset. In *IOP Conference Series: Materials Science and Engineering*, Vol. 1073. IOP Publishing, 012037.
[168] Mathias Wien, Renaud Cazoulat, Andreas Graffunder, Andreas Hutter, and Peter Amon. 2007. Real-time system for adaptive video streaming based on SVC. *IEEE Transactions on Circuits and Systems for Video Technology* 17, 9 (2007), 1227–1237.
[169] Mahisorn Wongphati, Yushi Matsuda, Hirokata Osawa, and Michita Imai. 2012. Where do you want to use a robotic arm? And what do you want from the robot?. In *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 322–327.
[170] Pengxiang Xia, Hengxu You, and Jing Du. 2023. Visual-haptic feedback for ROV subsea navigation control. *Automation in Construction* 154 (2023), 104987.
[171] Jiacheng Xie, Shuguang Liu, and Xuewen Wang. 2022. Framework for a closed-loop cooperative human Cyber-Physical System for the mining industry driven by VR and AR: MHCPS. *Computers & Industrial Engineering* 168 (2022), 108050.
[172] Ruishuo Xu, Weijun Wang, Wei Feng, Zhaokun Zhou, Boyoung An, Ruizhen Gao, and Kaichen Zhou. 2022. Design of a human-robot interaction system for robot teleoperation based on digital twinning. In *2022 IEEE Conference on Telecommunications, Optics and Computer Science (TOCS)*. IEEE, 720–726.
[173] Shiyu Xu, Scott Moore, and Akansel Cosgun. 2022. Shared-control robotic manipulation in virtual reality. In *2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*. IEEE, 1–6.
[174] Xuanhui Xu, Eleni Mangina, and Abraham G Campbell. 2021. Hmd-based virtual and augmented reality in medical education: A systematic review. *Frontiers in Virtual Reality* 2 (2021), 692103.
[175] Yang Xu, Chenguang Yang, Xiaofeng Liu, and Zhijun Li. 2018. A Novel Robot Teaching System Based on Mixed Reality. In *2018 3rd International Conference on Advanced Robotics and Mechatronics (ICARM)*. IEEE, 250–255.
[176] Chung Xue, Yuansong Qiao, and Niall Murray. 2020. Enabling human-robot-interaction for remote robotic operation via augmented reality. In *2020 IEEE 21st International Symposium on" A World of Wireless, Mobile and Multimedia Networks"(WoWMoM)*. IEEE, 194–196.
[177] Yuri DV Yasuda, Luiz Eduardo G Martins, and Fabio AM Cappabianco. 2020. Autonomous visual navigation for mobile robots: A systematic literature review. *ACM Computing Surveys (CSUR)* 53, 1 (2020), 1–34.
[178] AWW Yew, SK Ong, and AYC Nee. 2017. Immersive augmented reality environment for the teleoperation of maintenance robots. *Procedia Cirp* 61 (2017), 305–310.
[179] Huitaek Yun and Martin BG Jun. 2022. Immersive and interactive cyber-physical system (I2CPS) and virtual reality interface for human involved robotic manufacturing. *Journal of Manufacturing Systems* 62 (2022), 234–248.
[180] Faisal Zaman, Craig Anslow, and Taehyun James Rhee. 2023. Vicarious: Context-aware Viewpoints Selection for Mixed Reality Collaboration. In *Proceedings of the 29th ACM Symposium on Virtual Reality Software and Technology*. 1–11.
[181] Mohammad Kassem Zein, Majd Al Aawar, Daniel Asmar, and Imad H Elhajj. 2021. Deep learning and mixed reality to autocomplete teleoperation. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 4523–4529.
[182] Brian J. Zhang and Naomi T. Fitter. 2023. Nonverbal Sound in Human-Robot Interaction: A Systematic Review. *J. Hum.-Robot Interact.* 12, 4, Article 46 (Dec. 2023), 46 pages.
[183] Jingxin Zhang. 2018. Natural human-robot interaction in virtual reality telepresence systems. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 812–813.
[184] Jiachen Zhang, Onaizah Onaizah, Kevin Middleton, Lidan You, and Eric Diller. 2017. Reliable grasping of three-dimensional untethered mobile magnetic microgripper for autonomous pick-and-place. *IEEE Robotics and Automation Letters* 2, 2 (2017), 835–840.
[185] Lijun Zhao, Xiaoyu Li, Zhenye Sun, Ke Wang, and Chenguang Yang. 2017. A robot navigation method based on human-robot interaction for 3D environment mapping. In *2017 IEEE International Conference on Real-time Computing and Robotics (RCAR)*. IEEE, 409–414.
[186] Tianyu Zhou, Qi Zhu, and Jing Du. 2020. Intuitive robot teleoperation for civil engineering operations with virtual reality and deep learning scene reconstruction. *Advanced Engineering Informatics* 46 (2020), 101170.

[187] Cindy Ziker, Barbara Truman, and Heather Dodds. 2021. Cross reality (XR): Challenges and opportunities across the spectrum. *Innovative learning environments in STEM higher education: Opportunities, challenges, and looking forward* (2021), 55–77.

[188] Kateryna Zinchenko and Kai-Tai Song. 2021. Autonomous endoscope robot positioning using instrument segmentation with virtual reality visualization. *IEEE Access* 9 (2021), 72614–72623.

A Data Extraction List for Included Publication

Table 3. XR Technologies and Interaction Modalities (DE4 and DE5)

Category	Citations
XR technologies	
VR HMD	[24][172][112][164][150][84][134][151][167][181][147][77][186][23][104][132][108][159][25][51][78][179][171][126][27][149][46][188][30][8][107][138][110][102][173][69][70][156][73][144][141][185][94][175][80][135][33][160][65][18][59][158][13][161][183][99][92][7][119][100][75][44][1][74][48][145][101][170][85][120]
AR HMD	[97][62][96][171][63][50][40][154][127][41][26][55][176][7][137][130][93][117][15][39]
MR	[34][3][64][68][180]
CAVE	[51][52][12]
Mobie AR	[22]
Other	[129][115][178][98]
Interaction modalities	
Controller	[24][164][84][151][77][186][159][25][52][78][149][30][138][102][173][73][63][80][158][99][100][1][74][48][101][120][179][180][44][85]
Gesture	[34][96][171][108][69][70][185][175][33][127][13][26][176][68][137][93][15][40][119][180][117]
Joystick	[150][167][181][62][132][3][12][154][170]
Haptic devices	[112][23][134][115][126][144][75][179][44]
Motion capture	[172][108][46][8][18]
Walk	[110][50][59][41][119]
2D screen	[98][22][55]
Voice	[65][7][117]
Glove	[104][135][145]
Head	[141][119]
Gaze	[94]
Other	[129][97][147][51][27][188][156][64][160][41][92][85]

Table 4. Virtual Interface and the User's Perspective (DE6 and DE7)

Category	Citations
XR technologies	
Digital twin	[34][134][147][159][178][52][179][171][188][108][138][70][144][12][80][160][40][154][18][127][59][161][99][26][92][176][100][44][1][145][93][101][117]
Direct	[172][167][62][62][23][104][132][108][25][3][126][46][8][156][98][141][185][50][33][22][18][13][183][55][68][180][170][15][120]
Digital twin+3D reconstruction	[112][129][115][51][78][27][30][110][102][69][64][158][137][40][7]
Multiple	[150][186][173][41][74][85]
Direct+3D reconstruction	[24][181][149][119][75]
3D reconstruction	[164][151][94]
Virtual control room	[84][77][130]
Other	[97][96][63][135][175][65][48]
User's perspective	
Decoupling with robots	[112][150][129][34][97][134][62][147][159][115][178][51][52][78][179][171][27][188][30][107][102][70][144][94][175][80][64][160][40][65][154][127][158][161][41][99][26][92][176][68][180][100][137][44][1][74][93][101][117][39]
Coupling with robots	[172][84][167][181][77][23][104][132][108][96][25][3][126][149][46][8][156][98][73][141][185][50][135][33][22][18][13][183][55][119][75][130][170][120]
Dynamic perspective	[186][173][63][12][85]
God perspective	[151][69]
Other	[24][150][138][110][59][7][48][15]

Received 15 March 2024; revised 6 February 2025; accepted 9 April 2025

Table 5. Robot Types and Specific Tasks Classification (DE8 and DE9)

Category	Citations
Robot types	
Robotic arm	[172][112][164][150][134][147][23][159][115][51][52][78][179][46][8][107][102][173][70][175][80][33][160][65][127][161][41][99][26][92][176][7][100][75][44][130][48][145][93][101][117]
Mobile robot	[129][34][151][167][132][27][30][69][63][185][50][40][22][158][183][68][119][137][74][170]
Drones/UAV	[181][62][3][149][138][156][73][12][64][154]
Humanoid robot	[24][77][126][138][141][59][55][120]
Double-armed robot	[84][186][104][108][94][18][13]
Medical Robotics	[188][144][15]
Other	[97][96][178][171][98][135][180][85]
Robotic arm+mobile robot	[25][1][39]
Specific tasks	
Grabbing/Picking/Placement	[172][112][164][150][34][23][108][78][179][8][102][173][70][63][94][64][18][13][92][176][180][44][130][74][93][101][120]
Navigation	[24][129][167][132][3][30][110][69][98][141][12][50][22][154][59][158][183][170]
Industrial/Manufacturing	[97][84][134][147][186][96][159][178][51][171][46][175][160][161][99][7][117][39]
Multiple	[104][115][27][107][40][41]
No	[52][156][33][127][26][100][145]
Environment scan	[151][25][185]
Surgery/Healthcare	[188][144][15]
Search	[73]
Game/Entertainment/Social	[126][135][48]
Other	[181][77][149][138][80][55][68][119][75][137][1][85]

Table 6. Enhancement locations and types (DE13 and DE14)

Category	Citations
Enhancement location	
VE(Virtual environment)	[34][151][181][159][51][149][138][144][33][64][160][59][161][55][130][74][93][85][78][188][173][69][94][127][18][26][119][180][137][44][48][145][117][15]
User	[112][23][104][115][126][46][135][40][75][170][134][179][94][26][44][48][145][117][15]
Robot(Virtual)	[12][99][92][78][173][69][18][137]
Robot(Real)	[97][96][62][26][119]
RE(Real environment)	[63][22][154][68][62][180]
Object(Virtual)	[150][147][52][176][134][179][188][127][18]
No	[24][172][164][129][84][167][77][186][132][108][25][178][171][3][27][30][8][107][110][102][70][156][98][73][141][185][50][175][80][65][158][13][41][183][7][100][1][101][39][120]
Enhancement types	
Voice	[40][117]
Video	[130][93][62][188][173][180][137][44][74]
Text	[159][160][161][97][78][179][173][22][127][137][74]
Ray	[78][94][18][26][180]
Highlight	[12][176][173][69][127][18][137]
Haptic	[112][23][104][115][126][46][135][75][170][134][179][40][44][48][145][15]
Graphic	[147][52][149][144][55][150][181][188][127][18][68][119][180][145]
Avatar	[51][138][33][64][59][85][69][94][26][68][119][180]
3D Object	[34][151][96][63][154][99][92][150][97][134][181][62][69][22][26][48][145][117][15]
No	[24][172][164][129][84][167][77][186][132][108][25][178][171][3][27][30][8][107][110][102][70][156][98][73][141][185][50][175][80][65][158][13][41][183][7][100][1][101][39][120]

Table 7. Systems evaluation methods (DE10)

Category	Citations
N/A	[129][151][186][96][115][178][51][52][171][27][188][30][8][138][110][173][70][156][98][141][63][185][94][175][135][64][65][127][158][183][99][26][55][92][176][7][100][130]
Time/accuracy of the task	[172][112][34][97][181][62][132][46][102][144][80][154][180][75][74][145][93][101][170][85][15][164][150][84][134][167][147][77][23][104][108][159][149][107][160][40][22][18][59][161][41][44][39]
Questionnaire	[112][34][97][181][62][132][126][46][102][73][144][80][154][68][180][75][74][145][93][101][170][85][117]
Comparison	[172][179][12][33][1][145][170]
Interview	[62][154][68][120]
AR/VR	[25][15]
Other	[25][179][3][50]

Table 8. Whether the system supports multiplayer/multi-bot (only included supported categories, DE12)

Category	Citations
Multi-user - One robot	[51]
One user - Multi-robot	[69][55]
Multi-user - Multi-robot	[110][64][7]
One-user - One-user+One robot	[97][96][126][94][119][180][137][48][85][15]
One-user - Multi-user+One robot	[151]

Included articles that are not listed on the table i.e. do not support multiplayer/multi-bot operations.