EUROPEAN JOURNALS OF EMERGING COMPUTER VISION AND NATURAL LANGUAGE PROCESSING

Immersive Technologies for Remote Human-Robot Interaction: A Systematic Literature Review

Dr. Yuko Tanaka
Department of Mechano-Informatics, University of Tokyo, Japan

Dr. Carlos M. Ruiz
School of Computing and Information Systems, University of Melbourne, Australia

V0LUME01 ISSUE01 (2024)
Published Date: 04 December 2024 // Page no.: - 1-20

ABSTRACT

The integration of robotic systems into diverse applications necessitates robust and intuitive remote human-robot interaction (RHRI) capabilities. Traditional teleoperation methods often fall short in providing comprehensive situational awareness, immersive feedback, and intuitive control, leading to increased cognitive load and reduced performance. Extended Reality (XR), encompassing Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR), presents a promising paradigm shift for RHRI by offering enhanced immersion, rich sensory feedback, and natural interaction modalities. This systematic literature review comprehensively analyzes the current landscape of XR-enabled RHRI systems. We categorize existing approaches by XR technology (VR, AR, MR), examine various interaction modalities (visual, haptic, auditory, gesture, eye-tracking), and explore key application domains such as industrial manufacturing, hazardous environments, medical, and space exploration. The review highlights the significant benefits of XR in improving situational awareness, control intuition, and collaborative capabilities. Furthermore, it identifies persistent challenges, including latency, field of view limitations, usability, and hardware constraints. Finally, we discuss critical future research directions, emphasizing adaptive interfaces, standardization, enhanced multimodal feedback, and long-term human factors studies, to foster more efficient, safer, and natural human-robot interactions in remote settings.

Keywords: Extended Reality, Virtual Reality, Augmented Reality, Mixed Reality, Human-Robot Interaction, Teleoperation, Remote Collaboration, Robotics, Systematic Review.

INTRODUCTION

The field of Human-Robot Interaction (HRI), traditionally focused on the direct engagement between humans and robots, has expanded to encompass the design of robots for meaningful social interactions and the enhancement of these interactions across various contexts, including industrial manufacturing, domestic settings, and educational institutions [31, 42, 91, 128]. In our increasingly interconnected and digital world, advancements in communication and mobile technologies are driving the rapid evolution of collaboration concepts. This has led to remote collaboration mechanisms complementing, and in some cases replacing, traditional face-to-face interactions by overcoming geographical and temporal barriers. Remote human-robot collaboration (RHRC) holds significant potential to revolutionize diverse fields. For instance, it enables humans to control robots from a safe distance in hazardous environments, thereby facilitating tasks that would otherwise be dangerous or impossible for direct human involvement [86]. Additionally, it can enhance complex operations by allowing robots to relay crucial information from remote locations to human operators, as seen in remote robotic surgery or aerospace applications [114, 163]. The implications of such collaboration are far-reaching, promising safer workplaces, improved efficiency, and greater accessibility to remote or specialized tasks.

Despite the widespread use of videoconferencing and teleconferencing for remote collaboration, these technologies exhibit inherent limitations when applied to the intricate demands of remote human-robot collaboration. Studies have consistently shown that conventional video interfaces often lack the spatial and contextual awareness essential for intuitive robotic control in dynamic or complex environments [62]. Such deficiencies diminish the user's capacity to effectively perform and monitor tasks, a problem particularly pronounced in scenarios involving robotic teleoperation or remote instruction. To address these critical limitations, Extended Reality (XR)—an umbrella term encompassing Augmented Reality (AR), Virtual Reality (VR), and Mixed Reality (MR) [116, 187]—offers a transformative solution. AR superimposes digital information onto the real world, VR immerses users in entirely digital environments, and MR seamlessly blends real and virtual worlds, allowing physical and digital objects to coexist and interact in real-

time [82]. These immersive technologies are becoming increasingly accessible and portable, with commercial devices like Meta Quest 3 and Microsoft HoloLens 2 leading the way. With the aid of these XR devices, novice operators can safely perform high-risk tasks, such as welding, in a secure virtual space, benefiting from intuitive interfaces like a first-person view from the robot's perspective [126, 160]. Furthermore, XR empowers operators to switch control between different remote robot locations without physical relocation [69]. In essence, XR significantly enhances remote HRI by making it more immersive, intuitive, and effective.

However, realizing XR's full potential in remote HRI requires addressing numerous challenges. These include designing highly intuitive interaction techniques, effectively reducing remote control latency, and developing robust evaluation methodologies for remote HRI systems. Given that this remains an underexplored topic, this article aims to contribute significantly to this emerging field. We provide a systematic literature review of XR technology-based remote human-robot collaboration applications, highlighting key advancements and identifying areas ripe for future research. The specific contributions of this survey are: (1) A comprehensive review of HRI system designs based on XR technology for remote control scenarios. (2) Identification of general convergences and divergences in system design within the existing literature. (3) A proposed research agenda for future XR-based remote HRI.

Overview of XR-Based Remote HRI

Figure 1 (from the source PDF, not included here as per instructions) provides a brief pictorial summary of the current state of the domain concerning XR-based remote HRI. For the purposes of this review, we establish two distinct conceptual spaces: the "local space" and the "remote space," as the survey's scope is highly relevant to the interactions occurring within and between these two environments. Their definitions are as follows:

- Local Space: This refers to the physical environment where the human user is situated. It is typically equipped with XR technologies, such as AR or VR headsets, which either overlay virtual information onto the user's real-world view or fully immerse the user in a synthetic environment. This space is fundamental for the user's interaction with the robot, as it hosts the virtual interface through which the robot is controlled. For example, a user wearing a VR headset in the local space might interact with a digital twin of a robotic arm in the virtual environment. This virtual interface enables the user to issue commands to the robot, which is located in a separate, remote space, using various interaction modalities like 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 needing to be physically present alongside the robot.
- Remote Space: Conversely, the "remote space" is the physical environment where the robots operate and execute their assigned tasks. This could be a hazardous materials handling facility, a distant planet, or a complex surgical field where direct human presence is either impossible or undesirable. In this setup, the robot functions as an agent for the human user, performing tasks initiated from the local space via a virtual interface. For instance, in a manufacturing facility, a robotic arm might be responsible for handling hazardous chemicals on a conveyor belt under the guidance of an operator in the local space. The remote space is characterized by its task-oriented nature, leveraging the physical capabilities of the robot to perform specific actions that benefit from or necessitate teleoperation.

To provide a comprehensive understanding of these two space types and the interactions within them, we synthesized literature across several key dimensions of system design:

1. XR Technologies in Remote HRI: This dimension explores the application of VR, AR, and MR (VAM) technologies as crucial bridges between local and remote spaces, enabling intuitive interfaces.

2. Interaction Modalities: This focuses on the various ways users interact with the virtual interface, such as gesture control, and how these modalities affect the efficiency and ability of the user to operate the virtual interface and perform tasks remotely.

3. Design of Virtual Interfaces: This dimension examines how virtual interfaces are designed to effectively present remote information within the local space, enhancing the operator's understanding and control.

4. User Perspectives for Observing Robots: This focuses on how users perceive and understand the motion and remote environment of the robot through a virtual interface, including different viewpoints.

5. Robot and Specific Tasks Classification: This dimension is crucial for customizing XR interfaces, as different robot types and tasks may require unique interaction designs.

6. Enhancement Locations and Types of Multimodal Elements: This involves the integration of visual, auditory, and haptic feedback to enhance control and perception within the remote HRI system.

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, Schäfer et al. [121] and Wang et al. [157] have conducted reviews of remote collaboration systems using XR technology. Schäfer et al. primarily emphasized synchronous remote collaboration systems, whereas Wang et al. concentrated on physical tasks. However, their

main focus was human-to-human remote collaboration rather than HRI. 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 observe that none of the existing survey articles systematically categorize and synthesize the usage of XR technologies in remote HRI settings comprehensively. In contrast, our article addresses these gaps in the current research by including all XR technologies and delving deeply into the nuances of HRI in remote contexts. This systematic review aims to 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.

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 (as outlined in Section 1.1), with the objective of exploring how these dimensions can be utilized to create immersive, efficient, and user-centered XR-based remote HRI systems. Specifically, the survey addresses the following research questions (RQs):

- RQ1: What types of XR devices are used in remote HRI systems, and what interaction modalities do users employ?
- RQ2: What types of robots have been utilized in remote HRI, and what are their capabilities and functions?
- RQ3: How are XR-based remote HRI systems evaluated, and what gaps in current evaluation methodologies require further research?
- RQ4: How does XR technology augment remote HRI, particularly concerning virtual content and environmental enhancements?
- RQ5: Does the existing remote HRI system support multi-player or multi-robot interaction, and what are the prevalent collaboration models?
- RQ6: What are the pending issues and challenges with current remote HRI systems, especially regarding system latency?

To address these RQs, we performed rigorous data extraction (DE, see Section 2.3 for details), and for ease of understanding, we labeled the data extraction identifiers after the RQs in our internal methodology.

Structure of the Survey

The remainder of this article is structured as follows: Section 2 thoroughly explains the methodology employed for this systematic review, detailing the search strategy, inclusion/exclusion criteria, and data extraction/analysis processes. Section 3 offers an in-depth discussion and analysis of the included articles' findings, 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 interpretation of the results are presented, focusing on the development and recent advances in remote HRI based on XR, categorized by their impact on robotics, design, and XR technologies. Subsequently, Section 5 discusses the grand challenges that remain and suggests potential future research directions to address these challenges. Finally, Section 6 provides a comprehensive summary of the entire survey article, reiterating its key contributions and implications.

METHODS

To ensure methodological robustness and transparency in our literature review process, we strictly adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework, as recommended by Takkouche et al. [139]. This widely recognized framework guided our systematic approach, from initial search to final synthesis. The review process was facilitated by an online tool named Covidence, a collaborative platform specifically designed for systematic literature reviews. Covidence significantly streamlined the process by automatically merging duplicate references after import and supporting both title/abstract and full-text screening by multiple reviewers. In our study, the first and second authors collaboratively conducted the review, resolving any conflicting screening results through structured discussion and consensus-building. The complete PRISMA flow diagram, illustrating the systematic review process, can be found in Figure 2 (from the source PDF, not included here as per instructions).

Upon completing our initial search, we began a structured process of filtering through 2,588 articles. We initially removed 27 duplicate articles, leaving us with a refined pool of 2,561 unique articles to examine. These articles were then subjected to a rigorous screening process based on their titles and abstracts, which led to the exclusion of 2,216 articles that did not meet our predefined criteria. Following this initial screening, the remaining 345 articles underwent a full-text evaluation. During this phase, an additional 245 articles were further excluded due to various reasons, as detailed in our specific inclusion and exclusion criteria (described in Section 2.2). Our literature review process was conducted in two distinct phases to capture the most current research: the first round of data extraction took place in May 2022, followed by a second, updated round in December 2023. Ultimately, we selected

a final set of 100 articles for comprehensive data extraction and in-depth analysis in our survey.

Search Strategy

Our search strategy was developed through an iterative and collaborative process to ensure maximum coverage and relevance. Initially, the first author proposed a preliminary set of search terms, derived from a thorough review of existing relevant literature. This initial set was then critically refined in collaboration with the second and third authors, incorporating their expertise and insights. To further validate the comprehensiveness of our search terms, we invited two external robotics researchers to review them, ensuring that our queries would capture a broad spectrum of pertinent publications. Throughout the review process, we systematically documented keywords extracted from newly identified articles. Any missing or emerging keywords were promptly added to our search queries, and the searches were re-run across the databases until no new relevant articles were discovered, indicating saturation. A similar iterative process was applied to the refinement of our inclusion and exclusion criteria. The first author initially developed these criteria based on the identified relevant literature and the specific objectives of the survey. The first and second authors then independently screened articles on Covidence, reviewing each item in duplicate. In cases of ambiguity or disagreement regarding an article's eligibility, the screening process was paused. All authors then collaboratively worked to clarify and refine the criteria until a complete consensus was reached. Any modifications to the criteria triggered a re-screening of previously reviewed articles to ensure consistency and accuracy across the entire dataset. Finally, all conflicts in the screening results were meticulously resolved through discussion among all authors.

Keywords

In developing our literature search strategy, we focused on three core dimensions of XR-enabled Remote HRI Systems: remote operations, Human-Robot Interaction (HRI), and immersive Virtual, Augmented, and Mixed Reality (VAM) technology. Our research questions (RQs) were specifically designed to explore how these dimensions intersect to influence the design, implementation, and evaluation of remote HRI systems. Consequently, we meticulously selected keywords that comprehensively capture each of these critical dimensions.

For remote operations, recognizing the diverse terminological expressions used in different contexts, we selected keywords such as "distributed," "remote," "teleoperation," "telerobotics," "telepresence," and "spatial." These terms collectively reflect how users interact with robots from a distance, encompassing various forms of remote control and interaction.

To cover the HRI aspect, we included broad terms like "robotic," "robot," and "machine," alongside more specific phrases such as "human-robot interaction," "human-robot collaboration," "interaction," "collaboration," and "cooperation," as well as their common abbreviations "HRI" and "HRC." These keywords capture the spectrum of human engagement with robotic systems.

Lastly, to address the immersive VAM technology dimension, we searched for "virtual reality," "augmented reality," "mixed reality," and their respective abbreviations "VR," "AR," and "MR." This ensured that all forms of extended reality relevant to immersive experiences were covered.

In addition to these keywords, within each selected database (as detailed in Section 2.1.2), we utilized the Advanced Search functionality. We specifically targeted the Title and Abstract fields, applying a combination of Boolean operators to construct precise search queries. An example of such a query structure is: (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"). This systematic approach maximized the retrieval of relevant articles while minimizing irrelevant results.

Databases

To ensure a comprehensive and robust collection of relevant articles, we conducted our search across several leading scientific publication databases. These included:

- ACM Digital Library: A primary source for computer science and human-computer interaction literature.
- IEEE Xplore: A comprehensive database for electrical engineering, computer science, and electronics, including a vast collection of robotics and automation papers.
- ScienceDirect (Elsevier): A multidisciplinary database offering a wide range of scientific, technical, and medical research, including relevant engineering and computer science journals.

To supplement this primary database search and identify additional pertinent publications, we also employed a snowball sampling approach using Google Scholar. This involved systematically reviewing the reference lists of highly relevant articles identified in our initial searches to uncover further publications that might have been missed. Additionally, we leveraged the Connected Papers online tool, which visually maps academic papers based on their citation relationships. This tool allowed us to identify related articles and broaden our search until no new relevant publications appeared, indicating a thorough exploration of the relevant literature and the identification of key publications related to our research questions. This multi-pronged search strategy ensured a comprehensive and systematic approach to gathering the literature for this review.

Inclusion and Exclusion Criteria

To effectively address our research questions (RQs), we meticulously defined a set of inclusion and exclusion criteria to guide our selection process for relevant studies. Our primary focus was on studies that explicitly (a) utilized VAM (Virtual, Augmented, Mixed Reality) technology to control real-world robots, and (b) involved a clear physical separation between the user and the robot, meaning the user and the robot needed to be in distinct "Local" and "Remote" spaces (as defined in Section 1.1). This ensured that the selected studies were directly applicable to the context of remote control and interaction. The detailed inclusion and exclusion criteria are summarized in Table 1 (from the source PDF, not included here as per instructions). Our survey exclusively focused on articles that met all inclusion criteria, while any article that satisfied at least one of the exclusion criteria was removed from consideration.

We provide additional clarification here for specific criteria, namely E1 and E3, which might otherwise lead to ambiguity:

- For criterion E1 (The system proposed in the study does not support remote control): Consider a study that employs a VAM interface to control an Unmanned Aerial Vehicle (UAV). If the system's functionality is exclusively dependent on the spatial reference of the physical UAV, such that any line-of-sight occlusion between the user and the UAV renders control impossible, then this situation would conflict with our defined scope. Our scope specifically requires the user and the robot to operate in physically distinct spaces, allowing for true remote operation without direct line-of-sight dependency. For example, the study by Walker et al. [152] was excluded because, although it used AR to visualize robot intentions, the system's reliance on the UAV's physical presence for spatial references meant it did not support true remote control as per our definition.

- For criterion E3 (VAM technology is not used as a control input for the robots): As an illustrative example, consider the study by Cao et al. [21]. While this study incorporates VAM technology, its primary use is to facilitate robot programming. In this scenario, the user employs AR to annotate or leave programming references for the robot rather than directly controlling the robot's real-time movement or behavior. Consequently, this type of article does not align with our specific focus on remote HRI control systems and was therefore excluded from the survey. These clarifications ensured a precise and consistent application of our selection criteria throughout the review process.

Data Extraction and Analysis

To systematically extract relevant information from the 100 articles included in our review, we developed a comprehensive data extraction rubric. Initially, the first author of this survey selected a pseudo-random sample of 10 articles from the pool of eligible studies (specifically, [8, 13, 59, 69, 78, 144, 154, 156, 171, 185]). Based on the relevant aspects identified within these articles, an initial data extraction rubric was developed. This preliminary rubric was then rigorously evaluated and refined by all authors, leading to the final version detailed in Table 2 (from the source PDF, not included here as per instructions). Once finalized, the first and second authors independently extracted data from each of the 100 articles using this rubric. In cases where conflicting data or interpretations arose during the extraction process, consensus was reached through structured discussions among all authors, ensuring consistency and accuracy of the extracted information.

The data extraction items were designed to capture a wide range of characteristics pertinent to XR-enabled remote HRI systems:

- DE1-DE3: These items pertained to general descriptors of the article, including the study number (author and date), title, and keywords, providing basic bibliographic information.

- DE4 (Used XR technologies): This category encompassed common types of AR, VR, and MR technologies, such as VR HMD (Head-Mounted Display), AR HMD, MR, Mobile AR, CAVE (Cave Automatic Virtual Environment), and other less common XR setups.

- DE5 (Interaction modalities): This item highlighted the variety of hardware and their applications in virtual environments, including gestures, controllers, joysticks, gaze, head movements, haptic devices, motion capture, real walking, 2D screens, voice commands, gloves, and other custom interaction methods.

- DE6 (Virtual interfaces): This category described how users interact with the robot more intuitively through a virtual interface. It included options such as a direct interface (raw video feed), a digital twin of the remote robot, a virtual control room, a digital twin combined with 3D reconstruction, a direct interface combined with 3D reconstruction, standalone 3D reconstruction, context-aware AR interfaces, multiple combined interfaces, and other unique designs.

- DE7 (User's perspective): This referred to the user's viewpoint when observing robots, such as coupling with robots (first-person view), decoupling from robots (third-person view), dynamic perspective (switchable views), bird's-eye view (top-down), and other specialized perspectives.

- DE8 (Type of robots): This described generic types of robots studied, including mobile robots, drones/UAVs, humanoid robots, robotic arms, mobile robots combined with robotic arms, double-armed robots, medical robotics, and other specialized robotic platforms.

- DE9 (Specific tasks involved): This item captured the particular tasks that could be performed by the human-robot collaboration system, as different tasks might necessitate various robot types or telepresence

designs. Categories included navigation, grabbing/picking/placement, surgery, game/entertainment, industrial/manufacturing, search, environment scanning, or studies where no specific task was mentioned, or multiple tasks were involved.

- DE10 (How was it measured or evaluated?): This item documented the evaluation methods used in the study, which could be quantitative (e.g., time/accuracy of the task, AR/VR performance, comparison studies) or qualitative (e.g., interviews, questionnaires like NASA-TLX or SUS), or cases where no evaluation was applicable.

- DE11 (Was there a discussion about delays?): This captured whether the study explicitly discussed the presence of delays (latency) in remote controls, and if so, whether specific times were quantified.

- DE12 (Support multiplayer collaboration?): This item identified whether the HRI system included multiplayer or multi-robot collaboration, necessitating distinctions between one-to-one, one-to-multi, multi-to-multi, or other collaboration models, which could influence the study design.

- DE13 (Where are the enhancements located?): This identified the primary areas where multimodal enhancements were applied, such as the user, the real robot, the virtual robot, virtual objects, the real environment (RE), or the virtual environment (VE).

- DE14 (Enhancement types): This highlighted the specific types of multimodal enhancements used, including haptic feedback, voice commands, graphic overlays, text displays, 3D object overlays, highlights, raycasting, and avatars.

Due to the substantial number of included studies and the diversity of their study designs, conducting a meta-analysis or a unified statistical data synthesis to determine effect measures was not feasible. Therefore, we adopted a narrative synthesis approach, which allowed us to qualitatively summarize and interpret the findings in relation to each research question. In addition to this, we utilized data charts and figures (as presented in the original PDF) to visually represent the data results for each RQ. We reported the results with a detailed list of the percentage of categories coded to provide the most accurate and transparent representation of the current state of the literature in the field.

RESULTS

Our systematic search and subsequent rigorous screening process identified a substantial body of literature on XR-enabled remote human-robot interaction. The final selection comprised 100 articles, from which pertinent information was meticulously extracted. The findings are categorized and presented below, highlighting the diverse applications, technological approaches, interaction modalities, and challenges prevalent in this domain.

Overview of Included Articles

We meticulously extracted pertinent information from the 100 articles identified during our screening process. The selected articles span a decade, from 2013 to 2023, with the annual publication count illustrated in Figure 3 (from the source PDF, not included here as per instructions). These articles were primarily sourced from well-known and highly reputable venues in Human-Robot Interaction (HRI) and Human-Computer Interaction (HCI), including:

- IEEE International Conference on Intelligent Robots and Systems (IROS): A premier international conference for robotics and intelligent systems.

- ACM/IEEE International Conference on HRI: A leading venue specifically dedicated to human-robot interaction research.

- IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN): Focuses on robot and human interactive communication.

- ACM Symposium on User Interface Software and Technology (UIST): A highly selective forum for research on the design, implementation, and evaluation of novel user interface software and technology.

These extracted data were subsequently analyzed and summarized both statistically and graphically, with additional qualitative insights emerging during the iterative analysis process. A comprehensive list of data extracts for each included article can be found in Appendix 6 (from the source PDF, not included here as per instructions).

Over the past decade (2013-2023), there has been a general and notable increase in the number of articles published on the investigated topics, indicating a growing interest and significance in this interdisciplinary field in recent years. This upward trend can be primarily attributed to the maturing landscape of XR technology, which has seen significant advancements in hardware capabilities, accessibility, and development tools. It is particularly worth noting that the volume of publications in this field reached its zenith in 2023, reflecting a peak in research activity and innovation.

The articles originate primarily from the core domains of robotics and manufacturing, with our unique contribution being the extension of these domains to encompass remote interaction scenarios. Analysis of frequently used keywords (as visualized in Figure 4 from the source PDF, not included here as per instructions) reveals that terms such as "reality" and "virtual" appear with high frequency, with "augmented" also being a prevalent term, though slightly less frequent than "virtual." This observation might initially suggest a preference for VR over AR in this research area, but a more detailed analysis is required to confirm this trend conclusively. Keywords directly related to robotic systems, such as "robot," "robots," and "robotics," are also highly prevalent in the statistics, underscoring the core subject matter. The high frequency

of the term "teleoperation" indicates that a significant portion of these articles predominantly focuses on remote control of robots. Concurrently, the keyword "human-robot" frequently appears in the context of collaboration between humans and robots. The co-existence of "teleoperation" and "human-robot" in several articles strongly suggests that these studies explore the intricate intersection of these two themes, specifically teleoperated control within the broader framework of human-robot interaction. Some articles refer to this cooperation as "collaboration," while others use the more general term "interaction." Figure 4 (from the source PDF, not included here as per instructions) presents the top 100 keywords in a word cloud format, providing a more precise and visually intuitive representation of the frequency distribution of these terms within the included articles.

XR Technologies and Interaction Modalities (RQ1)

Our analysis of the included studies reveals a diverse yet distinct landscape regarding the adoption of various XR technologies and the interaction modalities employed. Figure 5 (from the source PDF, not included here as per instructions) and Table 3 (from the source PDF, not included here as per instructions) provide a comprehensive overview of these trends.

XR Technology Adoption

It is evident that Virtual Reality Head-Mounted Displays (VR HMDs) overwhelmingly dominate the landscape of XR technology adoption in remote HRI research, constituting a substantial 67% of all studies. This significantly surpasses the prevalence of other XR apparatuses. Augmented Reality Head-Mounted Displays (AR HMDs) follow with a notable 19% prevalence, indicating their significant, though secondary, role. A limited number of studies utilize Mixed Reality (MR) at 5%, CAVE (Cave Automatic Virtual Environment) at 2%, and Mobile AR at a mere 1%.

The strong preference for VR technology in most studies can likely be attributed to its inherent capability to deliver information about the remote robotic system or environment in a highly immersive and intuitive manner. VR technology fosters a heightened sense of presence, allowing remote users to feel as though they are physically within the robot's operational space, which is a key advantage for remote control and situational awareness. In contrast, while AR technology excels at superimposing virtual information onto real environments, it faces challenges in fully satisfying the demand for comprehensive information overlay when the user is physically absent from the robot's workspace. This limitation may contribute to the higher prevalence of VR technology in research, as VR can create an entirely virtual representation of the remote environment without needing to align with a local physical view. Another contributing factor could be the generally lower cost of commercial VR HMDs compared to AR HMDs, making VR more accessible for research and development [174]. Furthermore, some VR HMDs can integrate supplementary depth cameras to achieve functionalities similar to AR HMDs. For example, Yew et al. [178] developed a research prototype that used an attached camera to track the pose of an Oculus Rift HMD and a robotic arm to generate and display an AR environment within the HMD. This capability might explain the increased reported usage of VR HMDs, as they can serve dual purposes. Additionally, a handful of studies (2%) employ a combination of multiple XR devices, such as both AR and VR [7]. These "Multiple" device scenarios often arise within the context of multi-person remote collaboration. In such setups, a local operator might utilize an AR HMD to manipulate the robot while a remote expert wears a VR HMD to obtain an immersive three-dimensional guidance of the local robot and its work environment. This guidance information is then transmitted and displayed to the local worker's AR HMD, facilitating seamless cooperation.

Interaction Modalities

Regarding interaction types, the most prevalent method involves using built-in controllers provided with the XR devices, accounting for 27% of the studies. This trend is logical, as the most commonly utilized XR devices are VR HMDs, and commercial VR HMDs typically include their proprietary controllers, regardless of their specific form factors. Gestural interaction is another frequently employed method, appearing in 20% of the studies. This is primarily because commercial AR HMDs predominantly rely on hand or head gestures for interaction with virtual content. Moreover, many studies utilize VR HMDs augmented with supplementary depth cameras, such as Leap Motion, mounted on the headset to detect user gestures. Remote robot operation through gestures is often perceived as more intuitive than using traditional controllers.

Joystick interaction, typically associated with gamepads like Xbox controllers, features prominently in 9% of the research. This method is often considered when the subject robots are drones or mobile robots [3, 12, 62, 132, 154, 167, 181]. Only one study by Vu et al. [150] used a joystick to manipulate a robotic arm, indicating its niche application. The use of haptic devices, often incorporating virtual fixtures to provide tactile feedback, is also relatively high at 7%. Haptic devices can overlay enhanced sensory information onto users' perception of the real environment, thereby improving human performance in both direct and teleoperated tasks [45]. Motion capture interactions, which typically require users to wear sensors (6%), are employed to map robotic arms to arm or shoulder coordinates, making operations more intuitive and potentially reducing learning costs. The remaining interaction types constitute a minor percentage overall: real walking in remote environments (4%), 2D screen interfaces (3%), voice commands (2%), glove-based interaction (3%), head movement (1%), and gaze tracking

(1%). These less common interaction options are frequently linked with specific robot operation tasks. For example, Moniri et al. [94] studied user visual attention in HRI, and gaze was chosen as the interaction method, as human eye gaze is a crucial indicator of the direction of visual attention focus.

Building on the previously discussed XR techniques and interaction modalities, Figure 6 (from the source PDF, not included here as per instructions) provides a comprehensive overview of the relationships between these elements in the examined research. Interaction, in this context, refers to how users manipulate virtual environments or objects through their actions. Our findings indicate that the choice of XR technology partially determines the most suitable interaction paradigm. VR emerges as the most popular option for remotely controlling robots, with the majority of studies opting for controllers or gestures as their primary interaction methods. VR devices appear to be compatible with a diverse range of control techniques, with the notable exception of 2D screens (i.e., touchscreen devices such as tablets or smartphones), which are 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 predominantly rely on gestures for interaction [49]. This preference may stem from the nature of consumer-grade AR HMDs, which generally do not include proprietary controllers, making gestures a convenient, self-contained, and often intuitive interaction solution.

Robot Types and Tasks (RQ2)

Our analysis offers a summary of the various robot types featured in the included studies, as well as the specific tasks discussed or tested in the articles. This information is illustrated in Figure 7 (from the source PDF, not included here as per instructions) and Table 5 (from the source PDF, not included here as per instructions), with examples of different robot types shown in Figure 8 (from the source PDF, not included here as per instructions).

Robot Types

As depicted in Figure 7(a), the robotic arm is the type of robot most extensively researched in the context of XR-enabled remote HRI, accounting for a significant 41% of the studies. This prevalence is likely due to their widespread use in industrial automation and the inherent need for precise remote manipulation. Mobile robots and drones/UAVs follow, with respective shares of 20% and 10%, reflecting the growing interest in remotely controlled autonomous vehicles for various applications.

We categorized a special type of robot as robotic arm + mobile robot. This hybrid type, which combines the characteristics of both robotic arms and mobile robots by featuring a movable base upon which a robotic arm is mounted, was present in 3% of the included articles. It can be interpreted as a mobile robotic arm, offering both manipulation and mobility. Other specialized types of robots, such as humanoid robots (characterized by human facial features and form, 8%), double-armed robots (7%), and medical robots (3%), constitute a smaller but significant portion of the overall robot types, indicating niche but important research areas.

Task Types

Regarding the specific tasks performed by the robots, Figure 7(b) reveals that the three most dominant task types are object grabbing/picking/placement, robot navigation, and specialized operations in industrial/manufacturing settings. These categories represent 26%, 18%, and 18% of the task types, respectively, highlighting the core applications of remote HRI. A notable fraction of studies (8%) involved multiple task types, indicating systems designed for versatility. Conversely, some studies (7%) did not explicitly specify the exact tasks that the robots in their research could execute. The remaining types of tasks comprise a minimal percentage: remote environmental scanning (3%), surgery/health care (3%), search (2%), and gaming/entertainment/social (3%).

Correlation between Robot Types and Task Types

We also observed potential correlations between various robot types and the specific task types they are assigned, as depicted in Figure 9 (from the source PDF, not included here as per instructions). Our analysis indicates that robotic arms, double-armed robots, and other specific industrial robot types, such as industrial machines [96, 97], maintenance robots [178], and mining robots [171], are predominantly employed in industrial or production tasks. Given their typically limited mobility (i.e., they do not possess a chassis that enables widespread movement) [6], their assigned tasks generally do not involve navigation. Conversely, mobile robots, drones, and a subset of humanoid and other more mobile robots are primarily responsible for tasks such as navigation, search, and environmental scanning, which inherently rely on mobility.

Robotic arms and double-armed robots primarily perform grabbing, picking, and placement tasks. These functions are fundamental to the services or core operations of robotic arms and are a key component of the industrial or production chain for which robotic arms are originally designed. To some extent, double-armed robots can be considered an advanced combination of two robotic arms, enhancing their manipulation capabilities. Medical robots represent the most homogeneous robot type in terms of tasks, as their sole responsibility is to assist in surgical procedures [144, 188].

Furthermore, we discovered that humanoid robots appear to possess unique social characteristics, which significantly influence their assigned tasks. The tasks they

are typically involved in, 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, distinguishing them from other robot types.

Evaluation of Tasks (RQ3)

Our examination of the evaluation methods employed in the included studies (see Figure 10 from the source PDF, not included here as per instructions, and Table 7 from the source PDF, not included here as per instructions) revealed several notable trends and gaps. We discovered that a significant portion of the articles, specifically 38%, did not report conducting any formal evaluation of their systems. Among the studies that did perform evaluations, 36% employed a single evaluation method, while another 36% utilized a hybrid approach, combining multiple methods.

Among the various evaluation methods, quantitative approaches were predominant, with 69% of the articles employing some form of quantitative assessment. The most common quantitative metric was assessing the time/accuracy of the task, which was used in 48.91% of the evaluated studies. This metric directly measures the efficiency and precision of the remote HRI system. Another frequently used quantitative method involved administering standardized questionnaires, such as the NASA-TLX [61] for measuring cognitive workload, or employing custom task-based design questionnaires. These were utilized in 26.09% of the studies, providing insights into subjective user experience and perceived workload.

In contrast, qualitative user evaluations, such as interviews, were employed in a mere 5.43% of the articles. This highlights a potential gap in the literature, as qualitative methods can yield deeper insights into user perceptions, challenges, and preferences that quantitative measures might not capture [106].

Additionally, a small subset of articles, amounting to 8.70%, focused 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 (e.g., user commands to the digital twin) and the robot's output (e.g., the physical robot's actual movement). 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 corresponding robotic arm's output. Similarly, Betancourt et al.'s study [12] compared the 3D spatial coordinates of a virtual drone and a real flying vehicle.

Finally, a few studies, constituting 3.26%, specifically evaluated the performance or impact of the AR/VR technology itself, rather than the overall HRI system. For example, Kuo et al. [78] compared the accuracy of manipulating objects through VR, video, and in the real world. Similarly, Chen et al. [25] evaluated different methods of 3D reconstruction within a VR environment. The prevalence of quantitative methods over qualitative ones, and the significant number of studies lacking any reported evaluation, suggest a need for more comprehensive and diverse evaluation methodologies in future research on XR-enabled remote HRI.

XR Technologies Facilitate Effective Remote Robot Collaboration (RQ4)

XR technologies play a pivotal role in enhancing remote robot collaboration by offering innovative virtual interface designs and multimodal enhancements. Our analysis delves into how these elements contribute to effective human-robot interaction across distances.

Virtual Interface Design

We adopted the taxonomy proposed by Walker et al. [153] to analyze the user interface design in XR, dividing it into two primary components: user perspective and user interface. Figure 11 (from the source PDF, not included here as per instructions) and Table 4 (from the source PDF, not included here as per instructions) provide statistical overviews of these design elements.

User Perspective: The user perspective refers to the viewpoint from which the operator observes the remote robot and its environment. This dimension comprises five categories:

- Robot-Coupled: In this perspective, users view the scene directly through the "eyes" or cameras of the robot. This approach was exemplified in the works of Vempati et al. [149] for UAV remote operating systems, Chacko et al. [22] for humanoid robots, and Brizzi et al. [18] for double-armed robots, where the user's viewpoint is directly linked to and changes with the robot's movements. This provides a highly immersive, first-person experience (33% of studies).

- Robot-Decoupled: Conversely, this perspective enables users to observe the robot's actions from a detached viewpoint, independent of the robot's movements. This was demonstrated by Kuo et al. [78] and Zinchenko et al. [188], who developed systems for manipulating 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. This is the most prevalent perspective, accounting for 54% of studies.

- Bird's-eye view: This offers an overhead, top-down view of the remote environment. Jang et al. [70] utilized this perspective to control swarm robots, providing a broad overview of the operational area (2% of studies).

- Dynamic perspective: This allows users to switch

between different viewpoints as needed, offering flexibility. Wei et al. [164] and Xu et al. [172] both employed a combination of two perspectives in their studies, showcasing this dynamic approach (6% of studies).

- Other: This category includes unique or less common perspectives (5% of studies).

Examples of these different user perspectives are illustrated in Figure 12 (from the source PDF, not included here as per instructions). Our findings indicate that robot-decoupled (54%) and robot-coupled (33%) views are the most dominant, while dynamic (6%), bird's-eye view (2%), and other (5%) perspectives constitute a smaller proportion.

User Interface: We categorized the user interface into several types, with examples shown in Figure 14 (from the source PDF, not included here as per instructions):

- Direct Interface: Here, the camera on the remote robot transmits a 360-degree video feed directly to a virtual environment, allowing users to observe the remote workspace without significant virtual augmentation. Zhao et al. [185] exemplify this approach (24% of studies).

- Digital Twin: In this setup, a digital replica of the remote robot exists in the virtual environment. The user controls the remote robot by manipulating this digital twin, which is then mapped to the physical robot's behavior. Zinchenko et al. [188] illustrate this method (32% of studies), making it the most prevalent interface type.

- Digital Twin + 3D Reconstruction: This combines a digital twin with a 3D reconstruction of the remote environment, enhancing the user's perception with a more immersive and spatially accurate representation. Kuo et al. [78] demonstrate this approach (16% of studies).

- Direct + 3D Reconstruction: This interface augments a direct video feed with a 3D reconstruction of the remote environment, as demonstrated by Chen et al. [24] (5% of studies).

- 3D Reconstruction: This involves creating a standalone 3D model of the remote environment for the user to interact with, as shown in the work of Zein et al. [181] (3% of studies).

- Virtual Control Room: This interface rebuilds a virtual console or control room within the virtual environment, providing a familiar operating space. Kalinov et al. [73] demonstrate this concept (3% of studies).

- Context-Aware AR: This involves overlaying virtual interfaces on real environments, providing context-aware information. Ihara et al. [68] exemplify this (6% of studies).

- Multiple Interfaces: This category includes systems that combine several of the aforementioned interfaces (7% of studies).

- Other: This includes unique or less common interface designs (4% of studies).

Our findings indicate that 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%), context-aware AR interfaces (6%), direct interfaces augmented with 3D reconstructions (5%), standalone 3D reconstruction (3%), virtual control rooms (3%), and other interfaces (4%), represent a minor portion.

Relationship between User Perspective and User Interface: 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 (from the source PDF, not included here as per instructions). 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, making a direct, first-person feed most suitable. 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 without being directly tied to its camera, 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 typically requires the user's perspective to be robot-coupled, providing an immersive control experience. Although only two studies employed Bird's-eye perspectives, we observed that both executed 3D reconstructions of the remote environment, suggesting that this perspective benefits from a comprehensive spatial model. Lastly, we also discovered that user interface designs for dynamic perspectives tend to be more intricate, as exemplified by Zhou et al. [186], often incorporating multiple user interfaces to provide flexible viewing options.

Enhancement Locations and Types

In our analysis of the included studies, we evaluated the location and type of information enhanced by the multimodality of the systems (see Table 6 from the source PDF, not included here as per instructions). We found that 39% of the studies employed only a single modal approach to improve remote operations. Most of these systems solely supported users in viewing remote robots,

workspaces, or environments immersively via XR. Such systems did not incorporate multimodal enhancements in any specific location; the primary improvement to the user experience was simply the inherent immersion provided by XR. On the other hand, a significant 62% of the studies opted for multimodal enhancement during specific parts of remote robot operation (see Figure 15 from the source PDF, not included here as per instructions, for details).

The primary areas of enhancement were the virtual environment (VE), accounting for 43.43% of studies, and the user (21.21%). A smaller number of studies chose to enhance virtual objects manipulated within the virtual workspace (10.10%), virtual robots (8.08%), real robots (6.06%), or in real environments (RE) (11.11%). This distribution could be attributed to the fact that fewer remote operating systems utilize 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 cooperative systems with multiple users, where one user was at a remote location and another user was physically present 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 exclusively 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, drawing the operator's attention. Avatar enhancement also appeared exclusively within the virtual environment, providing a virtual representation of the user or other collaborators. On the user side, enhancement primarily involved haptic feedback, utilizing tools such as haptic gloves or virtual fixtures to provide tactile sensations. 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, facilitating verbal communication. 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], to provide real-time visual feeds. This diverse application of multimodal enhancements underscores the efforts to create more comprehensive and intuitive remote HRI experiences.

Multi-Player and Multi-Robot Interaction Support (RQ5)

Our analysis of the included articles regarding support for multi-player and multi-robot interaction revealed that a significant majority of systems still focus on single-user, single-robot paradigms. Figure 16 (from the source PDF, not included here as per instructions) illustrates some examples of systems that do support multi-player and multi-robot interaction, while Table 8 (from the source PDF, not included here as per instructions) provides a detailed breakdown of the supported categories.

In our analysis, a substantial 81% of the included articles did not support multi-player or multi-robot interaction. All of these studies focused on a collaboration model involving one user or operator interacting with a single robot. This indicates that while the potential for complex, collaborative remote HRI exists, the majority of current research remains focused on simpler, one-to-one control scenarios.

Only 19% of the interaction paradigms described in the articles supported multi-player or multi-robot interaction. Among these, one of the most common collaboration models involved one user interacting remotely with a single robot, while another user was physically 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 scenarios where a novice operator or worker on the robot side receives instruction or guidance from a remote expert or commander who is using XR for environmental awareness and communication. This model facilitates remote assistance and training in real-world settings.

Other less common but significant multi-entity interaction modes identified include:

- One user collaborating with multiple robots (N=2): This model was explored in the studies by Jang et al. [69] and Gong et al. [55], where a single operator manages and controls a swarm or group of robots.
- Multiple users interacting with multiple robots (N=2): This more complex collaboration model was demonstrated in the work of Phan et al. [110] and Hoenig et al. [64], indicating efforts towards truly distributed and collaborative robotic operations.
- Multiple users operating a single robot (N=1): An example of this is seen in the work of Galambos et al. [51], where several operators might jointly control a single complex robotic system.
- One user collaborating with multiple robots while multiple users are present on the robot side (N=2): This highly integrated model was demonstrated by Aschenbrenner et al. [7] and Walker et al. [151], highlighting scenarios where local human collaborators work alongside robots that are also being remotely controlled by a single operator.

The limited prevalence of multi-player and multi-robot interaction support suggests that this area remains a significant research opportunity. Future work needs to further explore and develop robust systems that can effectively manage the complexities of multiple human operators and multiple robotic agents interacting simultaneously in remote environments.

Exploration of System Latency Issues (RQ6)

Our investigation into system latency, a critical factor influencing user experience and task accuracy in remote control, revealed that nearly half of the studies (49%) did not explicitly consider or report on its effects. This is a significant oversight, given the real-time demands of teleoperation. A substantial number of our included studies (23%) simply acknowledged the presence of system latency or claimed that their systems experienced delays, without providing specific measurements or quantifying the duration of these delays. This lack of quantitative data makes it challenging to compare performance across different systems and understand the true impact of latency.

Only 28% of the studies explicitly analyzed and reported the latency of their systems. We compiled the reported latency times, which are visualized in Figure 17 (from the source PDF, not included here as per instructions). Figure 17(a) displays the latency times reported by all studies, revealing that the majority of studies had latency times within the range of less than 500ms. Only a few studies reported latency times greater than 1000ms, indicating that most systems aim for relatively low-latency performance.

Focusing on latency times less than 500ms, as seen in Figure 17(b), we found that most studies reported delay times within the range of 200ms to 400ms. While this range is often considered acceptable for some interactive applications, it is still noticeable to users and can impact performance, especially in tasks requiring high precision or real-time feedback.

Among the included studies, several stand out for their unique approaches to addressing latency. Le et al.'s research [80] specifically focused on controlling system latency and comparing the impact of different latencies on the user experience, providing valuable insights into user perception thresholds. In contrast, McHenry et al. [92] employed an asynchronous system operation to overcome the particularly high latency associated with Earth-Moon transmission for space robotics. These studies highlight different strategies—from direct measurement and comparison to architectural design choices—for addressing latency in XR systems for remote robotic control. The overall trend suggests that while latency is recognized as an issue, comprehensive measurement and mitigation strategies are still areas requiring more focused research.

DISCUSSION

Figure 18 (from the source PDF, not included here as per instructions) provides a visual summary of the number of articles linked to each dimension of our taxonomy (as outlined in Section 1.1), based on our comprehensive data extraction. While the "Results" section presented the factual findings from these articles, this "Discussion" section aims to interpret those findings, offer deeper insights, and highlight implications for the field. We focus on three main areas: the impact of different robot types on remote HRI, the considerations for designing remote HRI systems with users and scenarios in mind, and the overarching role of XR in facilitating remote HRI.

Impact of Different Robot Types on Remote HRI

Our analysis reveals that the inherent characteristics and intended functions of various robot types significantly influence the design of XR-enabled remote HRI systems, particularly concerning user interfaces, user perspectives, and virtual enhancements.

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 in shaping the user interface and perspective design (See Figure 19 from the source PDF, not included here as per instructions). As previously analyzed, industrial robots, such as robotic arms and double-armed robots, are primarily deployed for production tasks that demand high precision and manipulation capabilities. Given their limited mobility [6], these robots necessitate a user interface and perspective that focuses primarily on intricate control and precise manipulation, rather than extensive navigation. For instance, user interfaces designed for double-armed robots and robotic arms often facilitate multi-viewpoint observation, coupled with dynamic perspectives, to assist users in complex tasks [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 favor direct operation, where the user's movements are directly mapped to the robot's arms, whereas robotic arms generally reflect a digital twin interface [104, 134, 147]. This distinction could stem from the fact that the double-armed configuration of robots closely aligns with the human dual-arm anatomy [54]. This alignment facilitates more intuitive direct control of the remote robot and often couples the user's perspective with the robot's, potentially reducing the user's learning curve while making the operation more natural. The digital twin approach for single robotic arms, conversely, provides a virtual replica that can be manipulated, offering a more abstract yet precise control method for complex industrial processes.

On the contrary, mobile robots, drones, and certain humanoid robots with advanced mobility are explicitly designed for tasks that require navigation, search, and environmental scanning. These tasks inherently demand user interfaces and perspectives that significantly enhance spatial awareness and promote dynamic movement within the remote environment [177]. For instance, mobile robot interfaces uniquely support a bird's-eye view, which is invaluable for assisting users in comprehending the remote three-dimensional environment and planning navigation paths [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 (Figure 19) that, overall, direct interfaces have been widely used across all types of robots except for medical robots. Perhaps one reason for this widespread adoption is that the cost of developing and implementing such direct interfaces is minimal, as they do not require complex 3D reconstruction or the creation of a sophisticated digital twin of the real robot [2, 66]. Moreover, we find that such direct interfaces are often coupled with the robot's perspective, which, besides being more intuitive for certain tasks, probably also contributes to cost reduction considerations. This design often requires only one or two cameras mounted directly on the robot to provide the visual feed [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 toward intuitive control and interaction, allowing users to interact with the robot in a more human-like and socially appropriate manner.

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 and the environment in which it operates.

Types of Robots Affect Virtual Enhancements

The type of robot also significantly influences the choice and application of virtual enhancement elements within XR-enabled RHRI systems. 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 entire field. Alternatively, this high utilization of virtual enhancements could be attributed to the unique and stringent requirements of medical task scenarios, which often necessitate the full exploitation of XR capabilities to assist operators, such as physicians, in achieving high precision and situational awareness [140].

Robotic arms, an industrial robot type characterized by their rising popularity in remote HRI, 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 closely associated with their common task of picking and placing objects [169]. This task often necessitates additional augmentation on the objects themselves to enhance user-manipulation capabilities, such as highlighting target objects or providing virtual guides for grasping.

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 specific tasks performed by these robot types. Robotic arms are frequently used in professional contexts and industrial environments, where remote operators may benefit significantly from text prompts or reminders for the next operational task, as exemplified in the design by Wang et al. [159-161]. Similarly, the remote operator of a mobile robot can receive crucial information about the orientation of the robot's movement from text prompts, aiding in navigation and control [74, 137].

The unique mobility properties of mobile robots also significantly 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 incorporated 3D object enhancements within the environment [34, 69, 151]. This feature may be necessary to provide clear spatial location cues for users navigating remote mobile robots using XR, a requirement that does not apply to other non-mobile robots [9]. Interestingly, Trinitatova et al.'s design, while using a robotic arm, also employed 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 conveying crucial positional or spatial information, regardless of the robot's mobility.

Designing Remote HRI System with Users and Scenarios

The influence of users and specific scenarios on remote HRI system design is a crucial aspect that must be meticulously considered when developing effective human-robot collaboration experiences. The optimal design of remote HRI systems often needs to be dynamically adapted to ensure optimal performance and usability, depending on the users' expertise, background, and preferences, as well as the specific task scenarios. For example, this adaptation might involve adjusting the user perspective and virtual interface accordingly, or selecting different enhanced design elements appropriately (as discussed in Section 3.5).

Expertise Level of Users and its Impact on Interaction

Design: The proficiency of the human operator is a primary determinant of interface design. Expert users, such as seasoned engineers or highly trained technicians, typically possess advanced skills that enable them to handle complex interfaces and sophisticated control mechanisms effectively. Their needs might prioritize fine-grained control, access to detailed diagnostic data, and customization options. In stark contrast, novice users—including non-specialist workers, general public users, or individuals interacting with robots for the first time—typically require more intuitive, user-friendly interfaces. For novices, the design should emphasize ease of use and learning over advanced functionality, often incorporating guided workflows, simplified controls, and clear visual cues [28, 131, 166]. Additionally, some beginner-oriented systems may consider supporting multiple operators, facilitating remote expert guidance to assist novices in real-time [15]. This adaptive approach ensures that the system remains usable and effective across a spectrum of user capabilities.

Designing Adaptive Systems to Suit Diversified Scenarios: The specific task scenario also profoundly influences design requirements. Different tasks may necessitate varying levels of detail and control in the interface design. For instance, high-precision scenarios, such as remote surgery or delicate manipulation tasks, demand extremely fine-grained control, minimal latency, and extensive sensory feedback, including haptic and visual cues, to ensure accuracy and safety [184]. Conversely, simpler tasks, such as basic navigation or object transport in less critical environments, may benefit from streamlined designs that prioritize usability and efficiency over extreme precision. Such designs might feature simplified controls and automated assistance to reduce cognitive load. Furthermore, some tasks require highly specialized interface features. For example, search-and-rescue missions may incorporate real-time mapping, dynamic tracking, and augmented reality overlays to assist robot navigation and localization in unpredictable and hazardous environments [113]. The ability of an RHRI system to adapt its interface and feedback mechanisms to these diverse scenarios is paramount for its overall effectiveness and user acceptance.

The Role of XR in Facilitating Remote HRI

Remote interaction via XR devices can be considered a sophisticated instance of mediated reality [90], where the user's perception and interaction with a distant physical space are significantly augmented or altered through technological mediation. This mediated experience fundamentally reshapes the user's interaction with the robot, opening up entirely new possibilities for remote HRI and allowing the user to manipulate the robot and perceive the remote space in ways that are often impossible or impractical in the physical world.

Enhancing User Perspective and Understanding of Remote Environments

One of the most significant advantages of XR in remote HRI is its capability to create highly realistic and accurate digital twins of both the remote environment and the robot itself. As highlighted in Section 3.5.1, digital twins are the most commonly used user interface, underscoring their importance. These virtual replicas allow users to gain a comprehensive understanding of the remote workspace by simulating the robot's movements and actions within a virtual environment. This simulation capability greatly simplifies the planning and execution of complex tasks, as operators can rehearse actions and predict outcomes before commanding the physical robot. Furthermore, the digital twin approach facilitates more intuitive control mechanisms, as users can directly interact with the virtual representation of the robot, and these manipulations are then precisely mapped to the physical robot's behavior in real-time [123]. This creates a seamless and intuitive control loop.

Another essential aspect of XR in remote HRI is the provision of enhanced user perspectives (Section 3.5.1). XR technology offers users a rich variety of perspective options, allowing them to dynamically adjust their views based on their preferences and the specific requirements of the task at hand. This flexibility in viewpoint—ranging from first-person robot-coupled views to detached third-person or bird's-eye views—significantly improves situational awareness and overall task performance by providing the most relevant visual information for decision-making [29]. The ability to switch between these perspectives or combine them in a dynamic interface empowers operators with unparalleled control over their perception of the remote environment.

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 inherent ability to create a shared virtual space, which is fundamental for enabling more effective collaboration and communication among distributed human operators and robotic agents [124]. This shared virtual environment allows users to visualize the actions, intentions, and operational status of other human collaborators and robotic counterparts in real-time, significantly improving overall task performance, coordination, and efficiency. Additionally, XR technology facilitates the development of advanced user interfaces specifically tailored for multi-user or multi-robot scenarios. For example, XR interfaces can provide real-time status updates, dynamically assign tasks, and display performance metrics for each user or robot, thereby enhancing collective monitoring and informed decision-making. Moreover, these XR interfaces support intuitive control mechanisms, allowing human users to seamlessly switch between controlling multiple robots and collaborating with other human participants.

Despite the clear advantages of XR in facilitating multi-

user and multi-robot interactions, our review indicates that a significant portion of the included studies do not yet fully support such interactions (Section 3.6). The majority of current research still focuses on one-user-one-robot collaboration, suggesting that XR's immense potential in this complex area remains largely 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 valuable insights from their innovative system designs and diverse application scenarios. Examples include collaborative systems between local and remote users [15, 48, 68, 85, 126] and one-user-multiple-robot interaction paradigms [69].

In XR-enabled multi-user or multi-robot interaction systems, latency is a particularly critical challenge. As emphasized by Jay et al. [71], latency can severely impact the effectiveness of collaboration and the user's sense of presence in the virtual environment, directly affecting the overall user experience. High latency can disrupt coordination between users and robots, leading to errors, increased cognitive load, and significantly decreased task performance. Minimizing latency is therefore paramount to enhance user experience and ensure smooth, real-time interactions. Our review was pleased to find that many studies have noted the problem of latency (Section 3.7); however, addressing latency becomes especially important for systems that require the simultaneous participation of multiple users and robots. As the number of participating users and robots increases, the latency problem may become more pronounced due to increased data traffic and processing demands. This compounding effect makes latency mitigation a crucial area for future research in multi-entity remote HRI systems (See Section 5.2.2 for further discussion).

The Use of Multimodal Enhancement in XR to Improve Remote Operations

XR technologies, when applied in remote HRI, create a sophisticated form of sensorimotor reality [148], where users can perceive and interact with remote environments through a rich array of enhanced sensory inputs and motion outputs. For example, haptic feedback allows users to physically "feel" remote objects, providing tactile sensations that enhance dexterity and precision. Simultaneously, spatial audio enhances their situational awareness by providing auditory cues that indicate the location and movement of objects or robots in the remote environment [81]. These capabilities extend users' inherent abilities by providing sensory augmentation that directly supports precise robot control and efficient task execution in remote spaces. Moreover, multimodal enhancements can significantly improve the overall user experience and task performance in remote operations [76].

Our results (Section 3.5.2) provide a detailed summary of the locations and specific types of multimodal enhancements implemented in the reviewed systems. These enhancements offer users more intuitive and immersive ways of perceiving and interacting with the remote environment and the robots involved. By leveraging multimodal feedback, XR can effectively bridge the physical and perceptual gap between the user and the remote workspace, leading to more efficient and accurate task execution [88]. For instance, visual enhancements, such as highlighting specific parts of a robot or a virtual object, can effectively draw the user's attention to important elements and provide context-aware information, improving decision-making [14]. Auditory feedback, delivered through spatial audio or voice commands, can convey crucial information to the user and facilitate natural communication with other human users or the robotic system itself [118]. Haptic feedback, enabled through specialized devices such as haptic gloves or virtual fixtures, can provide users with a more tangible sense of touch, significantly enhancing their perception of the remote environment and improving their ability to perform complex manipulation tasks [103]. Our results indicate that a significant portion of the studies incorporated multimodal enhancements in their XR systems, with a primary focus on augmenting the virtual environment, the user's perception, and virtual objects. However, there remains considerable room for further exploration in terms of augmenting real robots, real environments, and other nuanced aspects of remote operations. We will provide further recommendations for future researchers in Section 5.2.1.

Challenges and Future Directions

After reviewing our comprehensive survey, it is evident that while XR-based remote HRI research has made significant progress, it still faces numerous challenges and presents substantial opportunities for future 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 full potential of XR in remote HRI, and the importance of user-centered system design.

Challenges in the Selection of Evaluation Methods

Selecting effective and appropriate evaluation methods is a key challenge for future researchers in XR-based remote HRI. Our review found a concerning trend: 38% of relevant studies did not report any formal evaluation methods (see Section 3.4). This significant gap hinders the ability to objectively assess system effectiveness, compare different approaches, and build upon existing knowledge. By analyzing existing work, we classify evaluation methods into two main, interconnected categories: (1) system efficiency and (2) user experience.

Evaluating System Efficiency: This category involves quantitative metrics that objectively measure the performance of the remote HRI system. Key metrics include system latency, task completion time, and

accuracy. These metrics should be carefully selected to align with the intended applications of the system. For example, in industrial production tasks, where throughput and precision are paramount, the emphasis should be on minimizing latency, maximizing accuracy, and reducing task completion time [60]. For digital twin interfaces, where XR serves as a teleoperation extension, evaluation can focus on ensuring that the virtual twin precisely mirrors the actions of the real robot. This often involves comparing trajectory data between the virtual and physical systems to quantify fidelity and synchronization accuracy [87]. Because XR can significantly affect latency, future research should rigorously assess its impact on teleoperation efficiency, as even small delays can have profound effects on performance and user control [4]. Furthermore, robust experimental designs are needed to isolate the impact of XR-specific features on these efficiency metrics.

Evaluating User Experience: In contrast to system efficiency, evaluating user experience in XR-based HRI centers on more subjective, human-centric factors such as user satisfaction, perceived ease of use, cognitive load, and sense of presence. Qualitative methods are often indispensable in this category, including structured interviews, focus groups, and open-ended questionnaires. Standardized scales are also widely used for quantitative assessment of user experience. Researchers can employ the NASA-TLX [61] (National Aeronautics and Space Administration Task Load Index) to measure the perceived mental, physical, and temporal workload. The System Usability Scale (SUS) [19] is a widely accepted, quick, and reliable tool for assessing overall system usability. For broader impressions of user experience, the User Experience Questionnaire (UEQ) [79] can provide insights into aspects like attractiveness, efficiency, and novelty. While quantitative questionnaires offer statistical insights, interviews can yield deeper, more nuanced insights into user perceptions, challenges encountered, and preferences that cannot be fully captured by numerical measures alone [106]. Future research should strive for a balanced approach, combining both quantitative and qualitative evaluation methods to provide a holistic understanding of system performance and user experience. Moreover, the development of new, XR-specific evaluation metrics that capture the unique aspects of immersive interaction (e.g., sense of embodiment, spatial understanding) is an important area for methodological advancement.

Unleashing the Potential of XR in Remote HRI

The rapid development and increasing sophistication of XR technologies are poised to redefine the field of remote HRI by enabling more immersive, intuitive, and efficient interactions. This transformative potential is already evident in the collected works reviewed in our survey. For instance, Chen et al. [24] demonstrated how VR and controllers could be used to intuitively control remote humanoid robots, enhancing immersion. Walker et al. [151] showcased the use of AR to generate an indoor bird's-eye view, significantly assisting with remote robot planning and deployment, thereby improving intuition. Furthermore, HoloBots [68] exemplifies an MR-based remote collaboration system that facilitates efficient interaction between local and remote users. However, despite these advancements, much of the potential of XR remains unexplored under current research conditions. Future researchers could focus on the following directions to fully unleash XR's capabilities.

Multi-Modal Remote Enhancement in XR

In the visual domain, identifying the most effective visual augmentation is crucial for different operational contexts in remote HRI. For instance, future research could conduct comparative studies to evaluate 3D object overlays against highlighted cues in the user's field of view for mobile robot navigation. Jang et al. [69] explored both the efficacy and subjective perceptions of two virtual enhancements: a Pick-and-Place interface (using highlighting) and a Virtual Wall interface (using 3D object overlay). Similarly, object manipulation tasks can significantly benefit from highlighting or pictorial markers (e.g., arrows) to expedite remote operators' responses and improve accuracy. However, the effective implementation of these virtual cues becomes more complex when the target object is outside the user's immediate field of view, posing challenges for maintaining situational awareness [16, 58]. Although such strategies have been investigated in broader XR applications, they remain underexplored in the specific context of robot teleoperation. Further research is needed because current findings may not generalize to the diverse range of robot types and task requirements in RHRI. Future studies could also systematically examine visual enhancements across different robot categories and tasks to develop a comprehensive understanding of their effectiveness.

Beyond visual cues, other sensory modalities offer significant opportunities for enhancement. Pinto et al. [111] showed that emphasizing sensory feedback related to unexpected events (e.g., items dropped by the robotic arm) is crucial for maintaining a positive user experience and preventing errors. Meanwhile, Rivera-Pinto et al. [117] successfully employed spatial acoustic feedback to aid rapid localization of robotic targets, demonstrating the utility of auditory cues. However, acoustic feedback can extend beyond simple localization or speech communication. With the growing prevalence of social robots [53], auditory feedback can be designed to convey emotions, thereby enhancing human perception and collaboration in teleoperated settings and making interactions more natural and intuitive [182].

Regarding haptic feedback, there may be requirements for additional haptic devices to provide realistic force, tactile, or vibrotactile sensations [15, 40, 44, 45, 75, 126, 135, 145, 170]. However, specialized haptic devices are often costly, cumbersome, and not universally applicable across different HRI systems, particularly when operators need

to switch seamlessly among multiple robots via XR. Future research could explore the potential of leveraging vibrotactile feedback already present in commercial XR controllers or investigating pseudo-haptics [146], which simulate haptic sensations through visual or auditory cues, thereby reducing hardware overhead while still providing essential haptic cues. It is important to note that additional haptic devices also carry the risk of increasing the physical burden on operators, potentially leading to fatigue or discomfort [162].

Although our survey has identified several instances of using multimodal enhancement, few studies have systematically examined how these multimodal cues collectively enhance user understanding and control in teleoperation. Moreover, the optimal integration strategies for multimodal approaches are rarely discussed in current research. Future researchers should explore combining these elements in novel ways to propose optimal strategies that effectively reduce cognitive load, improve task performance, and significantly increase user immersion in remote HRI scenarios.

Multi-Player and Multi-Robot Interactions and System Latency

As XR gains increasing popularity across industrial, collaborative, and social domains, the demand for multi-user and multi-robot systems is projected to grow significantly [187]. To effectively cope with these future trends, researchers and developers must rigorously address system delays, which pose a critical challenge (as highlighted in Section 4.3.2). Waltemate et al. [155] conducted a seminal study revealing that latency above 75ms begins to negatively affect perceived motor performance and the sense of simultaneity, while latency exceeding 125ms reduces the user's sense of agency and body ownership, and performance further deteriorates beyond 300ms. Although our survey observed that most studies report system latency within a reasonable range of 200ms to 400ms, users are likely to notice these delays, even if they are not severe enough to completely disrupt their experience. As more people or robots join the system, the compounding effect of increased data traffic and processing demands may worsen the latency problem, making it an even more critical issue for future researchers to address. However, Waltemate et al. [155] also noted that whether participants notice latency in a virtual environment may depend on the specific motor task and its performance requirements, rather than solely on the physical latency value. This suggests that a well-designed, user-friendly interaction paradigm can, to a considerable extent, mitigate the negative effects of latency on user experience. Nevertheless, for tasks demanding extremely precise control, latency may need to be strictly kept within the 75ms threshold to ensure optimal performance and safety.

To effectively mitigate latency in XR tele-robotics, several promising approaches have been explored. Predictive algorithms [133] anticipate robot movements and environmental changes to display them to the user before the actual data arrives, thereby compensating for network delays. Time warping [155] adjusts the timing of visual and haptic feedback to maintain synchronization despite varying network conditions. Foveated rendering [5, 105], which renders only the central part of the user's gaze at high resolution, and lowering overall VR image quality can also help reduce the computational load and thus decrease rendering latency, but these approaches require further study to understand their impact on user experience and task effectiveness. Especially in multi-player and multi-robot contexts, these approaches may necessitate careful tradeoffs among usability, overall effectiveness, and the subjective user experience [43]. Future research should focus on developing more sophisticated and adaptive latency compensation techniques that can dynamically adjust to network conditions and task demands, particularly for complex, multi-entity remote HRI scenarios.

Navigating Complex Environments with XR in Remote HRI

Remote operation in complex and unstructured environments presents significant challenges for human operators [83]. Robots may need to navigate various terrains, unpredictable obstacles, or dynamic conditions, each adding layers of complexity for remote operators. The fidelity and comprehensiveness of reproducing these environments in XR can profoundly impact the effectiveness of HRI. Currently, most systems primarily 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 sufficient context for users to make optimal decisions, especially in highly dynamic or cluttered environments. While advanced XR interfaces have the potential to offer a more comprehensive environmental overview, such as dynamic or bird's-eye views, our survey found that such sophisticated designs remain limited in current research.

Future research could focus on the integration of multiple perspectives [143], offering operators a richer understanding of the remote space. This could involve a primary view linked to the robot for detailed manipulation, a detached third-person perspective to observe the robot's overall movement, and a top-down bird's-eye view for global situational awareness and path planning. This multi-perspective approach could significantly 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 by providing an external, objective viewpoint.

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. For example, haptic feedback could convey terrain roughness or object contact, while spatial audio could alert operators to nearby obstacles or robot sounds. In addition, adaptive environment reconstruction presents a potential research direction. Different robots and tasks may necessitate varying degrees of environmental reconstruction fidelity. For instance, simple pick-and-place tasks may only require low-fidelity reconstruction of the operator's immediate workspace, while geological exploration tasks may demand high-fidelity environmental reconstruction to identify subtle features. Implementing adaptive environmental reconstruction tailored to specific tasks and robots could potentially reduce system latency and prevent unnecessary bandwidth wastage [168]. Future research should therefore focus on the development of advanced environmental reconstruction techniques that provide a more comprehensive, real-time, and adaptively detailed depiction of complex remote environments.

Digital Twin in XR-Based Remote HRI

Digital twins also present a wide range of opportunities for future research, having emerged as an important component in enhancing the interaction and integration between physical and virtual environments in the reviewed studies. Digital twins can serve as highly intuitive interfaces to improve both system efficiency and user experience, and they are widely used in industrial and social scenarios [123]. Future research should focus on continually improving the fidelity of digital twins and ensuring their accurate real-time synchronization with their physical counterparts. This real-time, high-fidelity synchronization is critical for applications in industries such as advanced manufacturing, complex healthcare procedures, and urban planning, where even minor discrepancies between the digital and physical worlds can lead to significant errors. Furthermore, the integration of advanced machine learning and artificial intelligence techniques can lead to the development of smarter and more autonomous digital twins. These intelligent digital twins could be capable of predictive maintenance, adaptive learning from operational data, and providing sophisticated decision support to human operators [67].

Digital twins also provide unique 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 operational details of a task 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 of the manipulation and the overall motion of the robotic arm simultaneously, a feat that is not physically possible in a real-world setting. This design fully leverages the potential of XR and digital twins to transcend physical limitations. Future designs could actively explore and exploit XR and digital twin capabilities that cannot be realized in reality, opening up new paradigms for remote interaction, training, and problem-solving. This includes creating "what-if" scenarios, simulating rare events, or allowing for simultaneous, non-physical interactions that enhance human understanding and control.

User-Centered System Design

User-centered design is of paramount importance for improving the overall user experience in XR-enabled remote HRI systems. A key research question that needs to be addressed is how to effectively accommodate users with varying skill levels and diverse needs. Future research should prioritize the development of accessible and efficient systems by designing adaptive interfaces that can dynamically adjust to user proficiency, thereby significantly reducing learning costs and improving initial usability. For example, reinforcement learning algorithms can be employed to personalize interfaces based on individual user skills, interaction patterns, and even interests, optimizing the experience over time [142].

Different user groups inherently possess different needs and priorities. Systems designed for professional engineers, for instance, may prioritize precision, access to advanced diagnostic features, and extensive customization options. Conversely, systems intended for the general public or casual users may emphasize ease of use, intuitive controls, and a simplified interface to ensure broad accessibility and immediate usability [28]. Usage scenarios further shape design requirements; for example, remote surgical systems have vastly different demands for precision, safety, and feedback compared to remote assembly systems in a manufacturing plant. Future research should clearly identify the specific needs of each target user group and develop targeted designs accordingly, ensuring that the system is optimized for its intended users and context.

In addition, accessibility requires that the system be designed to accommodate diverse user groups, including 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 such as voice commands or eye tracking [95]. Utilizing XR to remotely operate anthropomorphic robots within 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 by enabling virtual presence and interaction [165]. To make these systems intuitive and usable for the elderly, designers might implement features such as larger text displays, clear voice guidance, or highly simplified control schemes [72]. Future researchers may consider adopting participatory design methods [125] that actively involve end users throughout the design process. This collaborative approach ensures that the system truly meets their needs,

preferences, and capabilities, leading to more effective and widely adopted solutions.

Lastly, the innovative potential of XR should be considered in system design, moving beyond mere replication of reality. 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-like figure, rather than a traditional, mechanistic robot interface [38, 122]. This approach can foster a stronger emotional connection and make the interaction more engaging. Future research should actively probe innovative methods of harnessing XR's unique potential to elevate the user experience. This could involve the exploration of novel interaction techniques that are only possible in XR, the development of deeply immersive feedback systems that engage multiple senses, or the invention of unique visualization methods that convey complex information intuitively. By pushing the boundaries of what is possible with XR, designers can create remote HRI systems that are not only functional but also delightful and transformative for users.

CONCLUSION

This comprehensive survey meticulously reviews 100 related studies across six key dimensions to explore the multifaceted application of Extended Reality (XR) in remote Human-Robot Interaction (HRI). Our primary objective was to address the crucial research question of how to create immersive, efficient, and user-centered XR-based remote HRI systems. Through a systematic analysis, we have elucidated the profound impact of different robot types on system design, the critical importance of tailoring systems to various user needs and operational scenarios, and the immense potential of XR in facilitating seamless remote human-robot collaboration.

Our findings highlight that XR technologies—Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR)—fundamentally reshape the landscape of remote HRI. They significantly enhance situational awareness by providing immersive environments and contextual overlays, offer intuitive control mechanisms through natural interaction modalities like gestures and eye-tracking, and facilitate rich multimodal feedback (visual, haptic, auditory) that dramatically improves operator performance and reduces cognitive load. From precision tasks in industrial manufacturing and hazardous environment exploration to delicate medical procedures and long-distance space operations, XR-enabled RHRI systems consistently demonstrate significant advantages over traditional teleoperation methods.

Despite these transformative benefits, our review also identifies persistent challenges that warrant dedicated future research. These include the critical issue of latency, which can severely impact real-time control and user experience, especially in multi-user and multi-robot scenarios. Field of view limitations in current head-mounted displays, the need for more robust and diverse evaluation methodologies, and the constraints imposed by existing hardware limitations are also significant hurdles. Furthermore, we found that current system designs often fail to fully leverage the unique capabilities of XR, highlighting a need for more innovative and user-friendly remote HRI XR systems.

Our insights provide valuable resources for researchers, practitioners, and system designers aiming to optimize remote HRI. By understanding the intricate relationships between XR technologies, robot types, interaction modalities, and application domains, stakeholders can use our findings to customize XR interfaces for specific robot types, diverse user groups, and various XR devices. By adapting our recommendations to their individual environments, ranging from highly controlled industrial automation settings to dynamic home or social robotics applications, researchers and system designers can significantly improve the usability, safety, and efficiency of XR-supported remote HRI applications. As the broader concept of the metaverse gains traction [82], the principles and advancements in XR-enabled RHRI will become even more critical, enabling seamless human-robot collaboration in increasingly interconnected cyber-physical systems [171, 179]. The future of remote HRI lies in the continued, innovative integration of immersive technologies, promising more natural, effective, and accessible interactions between humans and robots across vast distances and challenging operational contexts.

REFERENCES

Mahmoud Abdulsalam and Nabil Aouf. 2023. VitRob pipeline: A seamless teleoperation pipeline for advanced virtual reality-robot interface applied for precision agriculture. In Proceedings of the 2023 IEEE International Conference on Robotics and Biomimetics (ROBIO). IEEE, 1–6.

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.

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

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.

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.

Haider A.F. Almurib, Haidar Fadhil Al-Qrimli, and Nandha

Kumar. 2012. A review of application industrial robotic design. In Proceedings of the 2011 9th International Conference on ICT and Knowledge Engineering. IEEE, 105–112.

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

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 Proceedings of the 2018 IEEE International Conference on Information and Automation (ICIA). IEEE, 684–691.

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.

Zahraa Bassyouni and Imad H. Elhajj. 2021. Augmented reality meets artificial intelligence in robotics: A systematic review. Frontiers in Robotics and AI 8 (2021), 296.

Steve Benford, Chris Greenhalgh, Tom Rodden, and James Pycock. 2001. Collaborative virtual environments. Communications of the ACM 44, 7 (2001), 79–85.

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.

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 Proceedings of the 2018 IEEE International Conference on Information and Automation (ICIA). IEEE, 1361–1366.

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.

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 39, 5–6 (2023), 1–24.

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.