Human-Robot Collaboration Design Space for Critical Large-Scale Infrastructures

Xinyu Chen[1,2,*], Ke Li[2], Raimund Kammering[1], Wim Leemans[1] and Frank Steinicke[2]

[1]German Electron Synchrotron (DESY), Notkestraße 85, 22607 Hamburg, Germany
[2]Department of Informatics, University of Hamburg, Bundesstraße 56B, 20146 Hamburg, Germany

Abstract

The vast scale and inaccessibility of critical large-scale infrastructures necessitate remote robotic applications to minimize human risk and improve efficiency. To support the development of a high-performance, resilient, and trustworthy human-robot remote collaboration platform, we evaluate three core design components derived from literature review in this position paper: level of robot autonomy, level of immersion, and interaction modalities. By synthesizing these themes, we identify the pivotal trade-offs between system performance, safety-critical reliability, and usability.

Keywords

Human-Robot Collaboration, Telepresence, Extended Reality, Interaction Design, Challenging Environment

1. Introduction

Critical large-scale infrastructures are vast, complex, and high-value physical infrastructures—such as particle accelerators, nuclear plants, and cable landing station—that are essential for national security, economic stability, and scientific advancement. Maintaining these facilities normally requires remote control due to hazard and inaccessibility. Still, human operators might have to enter the facility when faults occur, resulting in costly machine shut-down and possible harm on human safety [1].

Recently, the advancements in robotics allows deployment of robots to this critical and extreme environment [2], while the control panel enhanced by mixed reality facilitates user's presence of the remote environment and can improve operation precision [3]. Nonetheless, many challenges remain in developing a reliable and usable mixed reality-based human-robot collaboration platform. On one hand, these challenges involve technical aspects, such as latency [4], the dimensionality gap between user input and system output [5], and the need for robust robot performance in cluttered and extreme environments [6]; on the other hand, they also stem from human factors, such as unpredictable error commands [7], cognitive load [8], stress [9], cybersickness [10], and mistrust of the robot [11].

To build up a high-performance, resilient, and trustworthy human-robot collaboration platform for large-scale infrastructures, this position paper explore three major themes based on literature review: *i). Level of Robot Autonomy*, *ii). Level of Immersion*, and *iii). Interaction Modalities.*

2. Level of Robot Autonomy

Previous research has presented diverse discussion and insights regarding level of robot autonomy [12, 13, 14, 4]. Although phrased differently, three main paradigms can be summarized:

- Low-Level Autonomy: Robotic systems with low-level autonomy rely primarily on manual control. They may execute predefined motions or perform limited reactive behaviors, such as emergency stop for collision avoidance, but lack task-level planning capabilities and self-adaptability to changing environments [2].

CHI'26: XR4CE Workshop, April 14, 2026, Barcelona, Spain
*Corresponding author.
xinyu.chen@desy.de (X. Chen); ke.li@uni-hamburg.de (K. Li); raimund.kammering@desy.de (R. Kammering); wim.leemans@desy.de (W. Leemans); frank.steinicke@uni-hamburg.de (F. Steinicke)

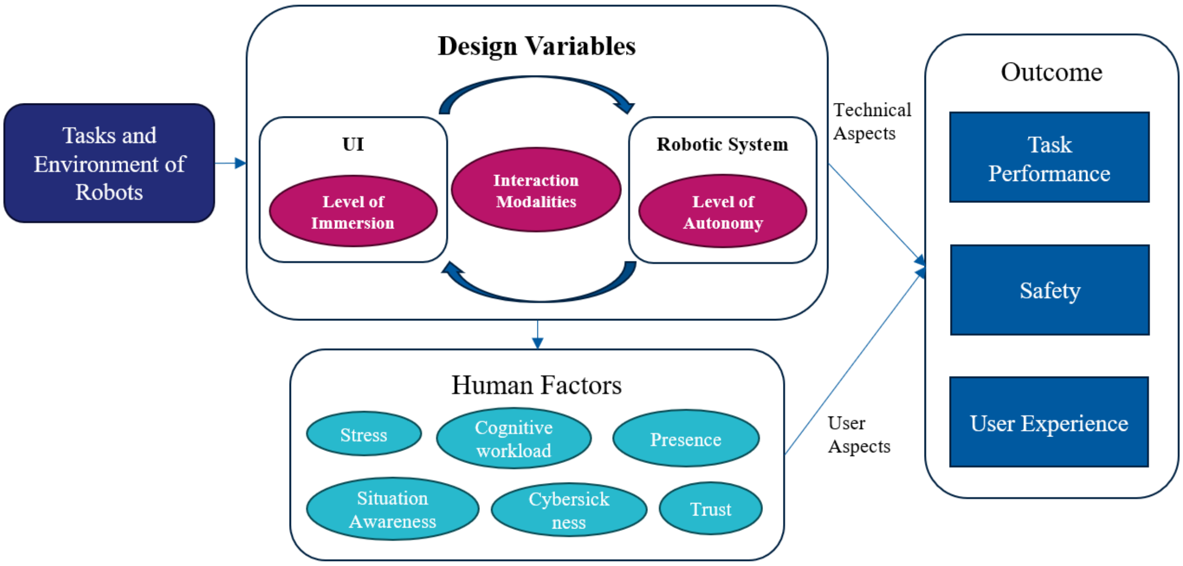

Figure 1: Design Variables of Human-Robot Collaboration for Critical Large-Scale Infrastructures and Their Impacts on Human Factors and the Final Outcome

- Shared Autonomy: Robotic systems with shared autonomy combine human guidance with autonomous capabilities. While operators specify high-level goals or provide supervisory input, the system autonomously handles subtasks such as motion planning, navigation, and obstacle avoidance [15, 16].
- Full Autonomy: Robotic systems with full autonomy operate without real-time human intervention. They are capable of perceiving the environment, planning and executing tasks independently, and adapting to unforeseen changes through integrated perception, reasoning, and control mechanisms [17].

Currently, to our best knowledge, most of robots deployed in critical large facilities are situated in low-level autonomy with concern of safety [7, 2], although robots with higher autonomy can significantly alleviate problems related to latency, cognitive load, and unpredictable human errors [13, 18, 4]. While reporting support for increased autonomy, Smith et al. [13] highlight the necessity of human involvement in complex tasks and safety-critical environment as experienced human operators can provide holistic understanding of the situations and therefore are more reliable to deal with unexpected events. Overly high robot autonomy can diminish the operator's situation awareness and, in the long term, may even lead to incapacitation [19, 13, 18]. Therefore, adopting shared autonomy offers significant potential benefits to improve usability and enhance resiliency in challenging operational environment.

As humans and robots increasingly operate as integrated teams, trust becomes a critical factor for effective collaborate. Overtrust may lead to complacency and insufficient monitoring, potentially resulting in catastrophic safety failures when the robot encounters situations beyond its capabilities; conversely, undertrust can cause disuse or unnecessary intervention, where users fail to leverage the robot's advantages, thereby reducing efficiency and increasing operator workload [20]. The level of autonomy is one factor influencing trust [11]. Nonetheless, research indicates that trust is not inherently proportional to the level of autonomy; even low-autonomy systems can still earn users' trust [21]. Instead, trust is contingent upon the alignment between human expectations and a robot's actual capabilities. This calibration ensures that operators can accurately discern the appropriate moments to intervene and assume manual control. Therefore, robotics engineers and domain experts must collaborate closely to define task allocation strategies and operational boundaries for robots within critical infrastructures, tailored to the inherent complexity of tasks and the specific attributes of the robotic systems.

3. Level of Immersion

As operators have limited access to the facilities during the machine runtime, the scene, audio, and other physical information inside are perceived by sensors and presented to operators via computer displays. Immersion is a metric to measure the capability of the displays to deliver an inclusive, extensive, surrounding, and vivid illusion of reality [22]. Displays are normally categorized in three immersion levels [10, 23]:

- *Low-Immersion Displays* normally only presents information of one sensory modality with limited field of view without depth sensing, such as traditional 2D monitors. The visualized scene does not match the user's proprioceptive feedback.
- *Moderate-Immersion Displays* normally presents information of one or two sensory modalities with extended field of view, such as concave displays [24]. The visualized scene can match the user's proprioceptive feedback.
- *High-Immersion Displays* normally presents information of more than two sensory modalities in 3D, such as Extend Reality Head-Mounted Display (HMD) or Cave Automatic Virtual Environment (CAVE) [25] with surround projection. The visualized scene closely match the user's proprioceptive feedback.

While low-immersion displays have traditionally served as the primary interface between the control room and the field, recent research has demonstrated the positive effects of high-immersion displays in collaboration with remote robots, particularly in improving task performance and reducing cognitive load [26, 27, 28]. Buss et al. [29] highlights multi-modal feedback is beneficial for telepresence and teleaction systems in large-scale environment. It is therefore reasonable to anticipate that immersive display technologies may become a mainstream interface for collaboration with telepresence robots.

However, existing studies also reveal that display with overly high immersion can lead to cybersickness, cognitive tunnel, and anxiety in emergency [10, 30, 31]. Especially, Souchet et al. [32] notes that 1/3 users experience discomfort during VR usage, and Kenan et al. [7] highlights that virtual reality is not suitable for extended use for teleoperating robots in nuclear facilities at the current state. Therefore, we suggest that the immersion level of displays should be adjustable according to the user comfort and the robotic tasks. For example, for people searching and rescuing which requires long-distance navigation and overview awareness, moderate immersion level would be more suitable than high immersion level; for fine manipulation requiring depth sensing for positioning, high immersion level would be beneficial.

4. Interaction Modalities

In critical large-scale infrastructures, the choice of interaction modalities between users and robots results in significant influence on operator's cognitive load, performance, trust, and operation safety [11]. Modalities include embodied interaction such as speech, gesture, gaze, and haptic control as well as disembodied interaction such as keyboard input and joystick input. The choice of the modality is related to the level of robot autonomy and the level of immersion. For instance, gaze and gesture control indicating human intention are more suitable if the robot has a certain level planning skills.

Advancements in robotics and extended reality communities facilitates applications with shared autonomy and high-immersion displays, presenting the opportunities to design more intuitive robot control and feedback methods [26]. For example, gestural control allows operators to more intuitively and naturally provide commands to robots [16]. However, in critical facilities with complex structure and expensive equipments, designing gesture-based controls requires carefully balancing the trade-off between intuitive interface design and potential losses in operational safety.

Speech interaction could complement manual control by enabling hands-free operation, which is essential when the operator needs to multi-task in a mission [33]. Rogalla et al. [34] demonstrated early on that combining speech with gestures can significantly reduce command ambiguity. However, reliability and the ambiguity of speech control shall remain a concern for critical infrastructure . As

Zinchenko et al. [35] investigated in the context of surgical robotics, speech recognition could be susceptible to errors caused by latency or changes in operator's voice pitch during high-stress situations. Therefore, speech command might best suited for discrete, high-level commands (e.g. "stop", "switch camera") rather than continuous, fine-grained control where a recognition error could lead to unintended robot actions.

Ultimately, a resilient teleoperation platform could not rely on a single modality. Therefore, we propose *multi-modal redundancy* for human-robot collaboration in critical large-scale infrastructures, where different inputs and outputs modality support each other in forming operator's final decision making [36]. For example, Gustavsson et al. [37] demonstrated the efficiency of combining speech recognition with haptic control, showing that multi-modal inputs could better support collaborative tasks. For critical large-scale infrastructures, this implies a design where operators could seamlessly switch modalities based on the context and scenarios of the tasks. For example, using gaze or gesture for coarse navigation, haptic control for fine manipulation, and speech control for system commands.

5. Conclusion

This position paper has examined three critical pillars—robot autonomy, immersion level, and interaction modalities—for developing a high-performance, resilient, and trustworthy human-robot collaboration framework in critical large-scale facilities. While current robotic interventions remain largely restricted to low-level autonomy due to safety concerns, we argue that transitioning toward shared autonomy is essential to mitigate the negative effects of latency and operator cognitive load while maintain high situation awareness. Regarding immersion, we propose an adaptive strategy according to specific tasks and user comfort. This balanced approach leverage the benefits of enhanced spatial perception for operation precision while address the limitations of cybersickness, cognitive tunnel, and anxiety. Furthermore, the integration of multi-modal interaction combining gestures, speech, gaze, and haptic controls could potentially provide the necessary redundancy to ensure reliable and intuitive command execution in high-stress, safety-critical scenarios. The synergy between these three dimensions is fundamental to solve challenges of developing human-robot remote collaboration platforms in both technical and human aspects. Ultimately, the successful deployment of such a platform requires continuous, cross-disciplinary collaboration between robotics engineers, specialists in mixed reality, and domain experts of the facilities. As a future step, we plan to extend this position paper to a larger literature review to provide a broader, more comprehensive analysis of the current and future research of human-robot collaboration in complex and inaccessible environments.

Acknowledgments

This work was supported by DASHH (HELMHOLTZ Graduate School for the Structure of Matter) with the Grant-No. HIDSS-0002.

Declaration on Generative AI

During the preparation of this work, the authors used OpenAI ChatGPT and Google Gemini in order to check grammar correctness and rephrase for clarity and writing flows. After using these services, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] Julien Branlard. "Low Level RF for SRF Accelerators". In: *LINAC14: Proceedings of the 27th International Linear Accelerator Conference*. 2014.

[2] S. Fargier et al. “Operational Controls for Robots Integrated in accelerator Complexes”. In: *Proceedings of the 19th International Conference on Accelerator and Large Experimental Physics Control Systems*. 19. 2023, pp. 423–428.

[3] Ke Li et al. “Reality Fusion: Robust Real-time Immersive Mobile Robot Teleoperation with Volumetric Visual Data Fusion”. In: *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (2024), pp. 8982–8989.

[4] Rafael Sousa Silva et al. “Evaluating Robotic Performative Autonomy in Collaborative Contexts Impacted by Latency”. In: *2025 IEEE International Conference on Robotics and Automation (ICRA)* (2025), pp. 3088–3094.

[5] Shalutha Rajapakshe. “Towards Accessible and Intuitive Shared Autonomy”. In: *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 2025, pp. 1879–1881.

[6] Changjoo Nam et al. “Fast and Resilient Manipulation Planning for Object Retrieval in Cluttered and Confined Environments”. In: *IEEE Transactions on Robotics* 37.5 (2021), pp. 1539–1552.

[7] Alperen Kenan et al. “Robot Teleoperation Design Requirements from End Users in Nuclear Facilities”. In: *2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2025, pp. 6749–6756.

[8] Juan Jose Garcia Cardenas et al. “Exploring Cognitive Load Dynamics in Human-Machine Interaction for Teleoperation: A User-Centric Perspective on Remote Operation System Design”. In: *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2024, pp. 12204–12211.

[9] Yi Ting Sam et al. “The Impact of Stress and Workload on Human Performance in Robot Teleoperation Tasks”. In: *IEEE Transactions on Robotics* 40 (2024), pp. 4725–4744.

[10] Sergo Martirosov et al. “Cyber sickness in low-immersive, semi-immersive, and fully immersive virtual reality”. In: *Virtual Reality* 26 (2021), pp. 15–32.

[11] Peter A. Hancock et al. “A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction”. In: *Human Factors: The Journal of Human Factors and Ergonomics Society* 53 (2011), pp. 517–527.

[12] Jenay M. Beer et al. “Toward a framework for levels of robot autonomy in human-robot interaction”. In: *Journal of Human-Robot Interaction* 3 (2014), pp. 74–99.

[13] Cailyn Smith et al. “Robot, Take the Joystick: Understanding Space Robotics Experts’ Views on Autonomy”. In: *2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN)* (2024), pp. 2182–2188.

[14] Raja Parasuraman et al. “A model for types and levels of human interaction with automation”. In: *IEEE transactions on systems, man, and cybernetics. Part A, Systems and humans : a publication of the IEEE Systems, Man, and Cybernetics Society* 30 3 (2000), pp. 286–97.

[15] Pinhao Song et al. “Robot Trajectron V2: A Probabilistic Shared Control Framework for Navigation”. In: *ArXiv* abs/2509.19954 (2025).

[16] Dennis Krupke et al. “Comparison of Multimodal Heading and Pointing Gestures for Co-Located Mixed Reality Human-Robot Interaction”. In: *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (2018), pp. 1–9.

[17] F. Paredes-Vallés et al. “Fully neuromorphic vision and control for autonomous drone flight”. In: *Science Robotics* 9.90 (2024), eadi0591. eprint: https://www.science.org/doi/pdf/10.1126/scirobotics.adi0591.

[18] Joost de Winter et al. “Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence”. In: *Transportation Research Part F-traffic Psychology and Behaviour* 27 (2014), pp. 196–217.

[19] Lisanne Bainbridge. “Ironies of automation”. In: *Automatica* 19.6 (1983), pp. 775–779.

[20] Giulio Campagna et al. “A Systematic Review of Trust Assessments in Human–Robot Interaction”. In: *ACM Transactions on Human-Robot Interaction* 14 (2024), pp. 1–35.

[21] Jiahe Pan et al. “Effects of Shared Control on Cognitive Load and Trust in Teleoperated Trajectory Tracking”. In: *IEEE Robotics and Automation Letters* 9.6 (2024), pp. 5863–5870.

[22] Mel Slater et al. “A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments”. In: *Presence: Teleoperators & Virtual Environments* 6 (1997), pp. 603–616.

[23] Haylie L. Miller et al. “Level of Immersion in Virtual Environments Impacts the Ability to Assess and Teach Social Skills in Autism Spectrum Disorder”. In: *Cyberpsychology, Behavior and Social Networking* 19 (2016), pp. 246–256.

[24] Mark H. Chignell et al. “Immersiveness and Perceptibility of Convex and Concave Displays”. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 63 (2019), pp. 396–400.

[25] Ross Tredinnick et al. “Uni-CAVE: A Unity3D plugin for non-head mounted VR display systems”. In: *2017 IEEE Virtual Reality (VR)* (2017), pp. 393–394.

[26] Rebecca Hetrick et al. “Comparing Virtual Reality Interfaces for the Teleoperation of Robots”. In: *2020 Systems and Information Engineering Design Symposium (SIEDS)* (2020), pp. 1–7.

[27] Dinh Tung Le et al. “Intuitive Virtual Reality based Control of a Real-world Mobile Manipulator”. In: *2020 16th International Conference on Control, Automation, Robotics and Vision (ICARCV)* (2020), pp. 767–772.

[28] Jaroslaw Jankowski et al. “Usability Evaluation of VR Interface for Mobile Robot Teleoperation”. In: *International Journal of Human-Computer Interaction* 31 (2015), pp. 882–889.

[29] Martin Buss et al. “Development of a Multi-modal Multi-user Telepresence and Teleaction System”. In: *The International Journal of Robotics Research* 29 (2010), pp. 1298–1316.

[30] Lisa C. Thomas et al. “Immersion and Battlefield Visualization: Frame of Reference Effects on Navigation Tasks and Cognitive Tunneling”. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 43 (1999), pp. 153–157.

[31] Pamela Barbadoro et al. “Stress responses in high-fidelity simulation and standard simulation training among medical students”. In: *BMC Medical Education* 23.1 (Feb. 17, 2023), p. 116.

[32] Alexis D. Souchet et al. “A narrative review of immersive virtual reality’s ergonomics and risks at the workplace: cybersickness, visual fatigue, muscular fatigue, acute stress, and mental overload”. In: *Virtual Reality* 27 (2022), pp. 19–50.

[33] Ángel-Gabriel Salinas-Martínez et al. “Multimodal Human–Robot Interaction Using Gestures and Speech: A Case Study for Printed Circuit Board Manufacturing”. In: *Journal of Manufacturing and Materials Processing* (2024).

[34] Oliver Rogalla et al. “Using gesture and speech control for commanding a robot assistant”. In: *Proceedings. 11th IEEE International Workshop on Robot and Human Interactive Communication* (2002), pp. 454–459.

[35] Kateryna Zinchenko et al. “A Study on Speech Recognition Control for a Surgical Robot”. In: *IEEE Transactions on Industrial Informatics* 13 (2017), pp. 607–615.

[36] Austin S. Wang et al. “Learning Robust Manipulation Strategies with Multimodal State Transition Models and Recovery Heuristics”. In: *2019 International Conference on Robotics and Automation (ICRA)* (2019), pp. 1309–1315.

[37] Patrik Gustavsson et al. “Human-Robot Collaboration Demonstrator Combining Speech Recognition and Haptic Control”. In: *Procedia CIRP* 63 (2017), pp. 396–401.