Human-Computer Interaction & Emerging Technologies, Vol. 233, 2026, 31–41
https://doi.org/10.54941/ahfe1007501

Understanding the Needs and Challenges of Developing Robot Teleoperation Applications Using Mixed Reality Headsets

Liuchuan Yu[1,2], Ke Jing[1], Zhigen Zhao[1,3], Ning Yang[1], and Zhicong Lu[2]

[1]ByteDance PICO, San Jose, CA 95110, USA
[2]Department of Computer Science, George Mason University, Fairfax, VA 22030, USA
[3]School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

ABSTRACT

Robot teleoperation enables humans to control robots to perform tasks and collect data to train robot intelligence. Compared to traditional interfaces, extended reality (XR)-based robot teleoperation offers more natural, efficient, and scalable interactions with reduced cognitive load. However, developing such applications involves interdisciplinary challenges across hardware, integration, interface design, and manipulation. To understand current practices and challenges, we conducted semi-structured interviews with 15 developers, ranging from novice prototypers to industry experts. While prior work focuses on end-user usability, this study explores the developer experience (DX) bottlenecks from the perspective of developers at varying levels of expertise, including undergraduate prototypers, graduate researchers, and industry practitioners. We identify a "Middleware Gap" where network instability and protocol mismatches hinder reproducibility, and a "Data Utility Crisis" where current XR tracking lacks the fidelity required for robust imitation learning. We contribute a refined taxonomy of XR teleoperation and a set of prioritized design implications, moving beyond generic wish lists to specific architectural requirements for interoperability, sensory substitution, and human-in-the-loop safety.

Keywords: Robot teleoperation, Extended reality, Mixed reality, Developer experience

INTRODUCTION

Robot teleoperation is evolving from a direct control mechanism into a data generation pipeline for embodied AI, supporting both real-time operation and the collection of high-quality interaction data for robot learning (Yeh et al., 2025). As a core problem in Human–Computer Interaction (HCI), teleoperation requires interfaces that effectively translate human intent into robotic action while addressing challenges such as latency, situational awareness, feedback, and user trust (Khasawneh et al., 2019; Louca et al., 2024; Bach et al., 2024). While traditional interfaces (e.g., joysticks and keyboards) suffer from poor spatial mapping and limited proprioceptive feedback (Whitney et al., 2019; Chen et al., 2007), extended reality (XR)

Received February 2, 2026; Revised April 5, 2026; Accepted April 23, 2026; Available online July 20, 2026

headsets offer embodied interaction that spatially aligns human perception with robot movement, improving intuitiveness and learning efficiency (Whitney et al., 2019; Chen et al., 2023; LeMasurier et al., 2024). Despite this promise, the creation of XR-based teleoperation systems remains fraught with technical and design friction. Fig. 1 demonstrates the diversity of XR-based teleoperation systems.

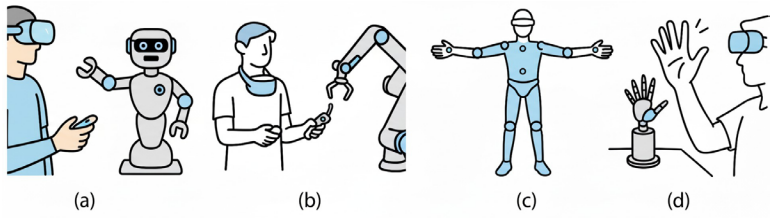

Figure 1: Robot teleoperation systems using mixed reality: (a) first-person control, (b) observation-driven control, (c) full-body teleoperation, and (d) dexterous hand manipulation.

This paper investigates the developmental challenges underlying XR teleoperation systems from a developer-centered perspective. Through interviews with a spectrum of developers, from undergraduates building ad hoc prototypes to academic researchers and industry practitioners deploying production-scale systems, we surface a persistent gap between "getting it to work" (novice priorities such as rapid setup and basic functionality) and "getting it to robustly scale" (expert priorities such as infrastructure stability, reproducibility, and data fidelity). This divide is exacerbated by infrastructural instability, operating system fragmentation, and hardware constraints, which are widely reported challenges in teleoperation practice (Rea and Seo, 2022; Wang et al., 2025).

In particular, we examine how these barriers hinder XR-based teleoperation development and limit its broader adoption, and how future toolkits must evolve to support the dual goals of intuitive real-time control and reliable, high-fidelity data collection for embodied AI. Accordingly, this work asks: (1) What infrastructural and hardware barriers impede the development of XR-based robot teleoperation systems? and (2) How should teleoperation toolkits evolve to reconcile the needs of novice prototypers and expert practitioners while supporting both control and data collection?

BACKGROUND AND RELATED WORK

The Fragmentation of XR Teleoperation

Recent advances in vision-language-action (VLA) models have intensified the need for large-scale, high-quality robot demonstration data (Zare et al., 2024; Black et al., 2024; Team et al., 2025). Robot teleoperation has therefore emerged as a practical mechanism for collecting multimodal training data that couples human intent with robot behavior. Prior work explores diverse

teleoperation approaches (Si et al., 2021; Darvish et al., 2023), including hardware-centric methods such as lead–follower arms (Fu et al., 2024) and universal grippers (Chi et al., 2024), vision-based hand tracking (Qin et al., 2023), and XR-based teleoperation leveraging immersive visual feedback and embodied interaction (Wang et al., 2024). While XR systems promise improved intuitiveness, they continue to face challenges related to setup complexity, usability, and scalability.

Table 1: Taxonomy of XR teleoperation.

Control Scale / Perspective	Fine Motor Control (Dexterous Manipulation)	Gross Motor Control (Whole-Body Retargeting)
Egocentric (First-Person/ Embodied)	**Embodied Manipulation.** Operator sees through robot vision and maps hand or finger movements to grippers. Inputs: Hand tracking, controllers. Use cases: precision assembly, surgery, complex grasping.	**Immersive Locomotion.** Operator embodies the robot's full frame and maps head or body pose to balance or gait. Inputs: XR headset, body tracking. Use cases: humanoid walking, drone FPV flight.
Exocentric (Third-Person/ External)	**Remote Manipulation.** Operator views the robot externally (camera or digital twin) and controls the end-effector. Inputs: 6DoF controllers, haptics. Use cases: pick-and-place, side-angle grasp inspection.	**Supervisory Navigation.** Operator views maps or external feeds and issues high-level path or posture commands. Inputs: joysticks, waypoints. Use cases: mobile base navigation, drone path planning.

Several XR teleoperation frameworks seek to unify development, including Open-TeleVision (Cheng et al., 2025), OPEN TEACH (Iyer et al., 2025), and XRoboToolkit (Zhao et al., 2025), yet adoption remains uneven and developer needs remain underexplored. Across this literature, control modality and operator perspective are often conflated, obscuring key design trade-offs. We therefore distinguish teleoperation systems along two orthogonal axes: (1) Perspective—egocentric (first-person) versus exocentric (third-person); and (2) Control Scale—fine motor control (dexterous manipulation) versus gross motor control (whole-body retargeting). This framing clarifies the XR teleoperation design space and motivates a developer-centered investigation of toolkit design (see Table 1).

Gaps in HCI and Developer Support

Beyond robotics pipelines and frameworks, HCI and Human–Robot Interaction (HRI) research emphasizes how people operate remote robots in practice, including input modalities, visual representations, and operational workflows (Si et al., 2021; Darvish et al., 2023; Moniruzzaman et al., 2022). While prior work extensively studies operator-facing concerns such as latency, situational awareness, and haptics (Luo et al., 2020), the developer burden of implementations remains underexplored. In particular, developers

must manually bridge what we term the *Embodiment Gap*: the mismatch between human perception and motor capabilities and robot morphology, often without standardized schemas or reusable guidance.

XR introduces embodied input options beyond traditional devices, including controller-based control, hand tracking for dexterous manipulation, and body or motion tracking for upper-body or humanoid control (Whitney et al., 2019; Chen et al., 2023; LeMasurier et al., 2024; Qin et al., 2023). Similarly, XR enables rich visualizations, such as mixed-reality overlays and spatial widgets, that can improve situational awareness (Ghimire et al., 2025; Whitney et al., 2019). However, existing systems rely on ad hoc, device-specific mappings and visualization strategies, and operational workflows (e.g., mode switching, sensitivity scaling, human takeover) remain inconsistent across toolkits (Iyer et al., 2025; Zhao et al., 2025). As a result, XR teleoperation systems are often developed in isolation for specific robot–input–display combinations (Si et al., 2021; Darvish et al., 2023). What is missing is a developer-oriented framework that systematically aligns robot targets, XR inputs, control mappings, workflows, and UI design, motivating our developer-centered investigation.

Table 2: Participant demographics including role, occupation, field, years of experience (YoE) in XR and robotics, and their projects.

ID	Group	Occupation	Field	XR YoE	Robot YoE	Project
P1	Prototyper	Undergraduate	Robotics Engineering	0	1–2	Pet litter scooping and petting
P2	Prototyper	Undergraduate	Embodied Intelligence	1–2	3–5	Robot arm for soldering assistance
P3	Prototyper	Undergraduate	Software Engineering	1–2	1–2	Robot arm teleoperation
P4	Prototyper	Undergraduate	Electrical Engineering	1–2	1–2	Elder care teleoperation robot
P5	Prototyper	Undergraduate	Computer Science	0	1–2	Farm surveillance robot
P6	Prototyper	Undergraduate	Internet of Things	0	0	Library book retrieval robot
P7	Prototyper	Undergraduate	Computer Science	3–5	1–2	Robotic glass-cleaning system
P8	Prototyper	Undergraduate	Mechanical Engineering	0	1–2	Teleoperated soldering robot
P9	Prototyper	Undergraduate	Automation Engineering	0	1–2	Home assistant teleoperation robot
P10	Prototyper	Undergraduate	Software Engineering	0	0	Drone teleoperation system
P11	Researcher	Graduate/Master	Robotics	3–5	1–2	Robot learning platform

(*Continued*)

Table 2: Continued.

ID	Group	Occupation	Field	XR YoE	Robot YoE	Project
P12	Researcher	Graduate/Master	Computer Science	1–2	1–2	Dexterous hand manipulation
P13	Researcher	Graduate/PhD	Artificial Intelligence	3–5	3–5	Humanoid robot teleoperation
P14	Practitioner	Manager	Humanoid Robotics	3–5	3–5	Robot fight teleoperation
P15	Practitioner	Developer	Mechanical Engineering	0	5+	Mechanical system teleoperation

METHOD

Participants

We recruited 15 participants (P1–P15) representing the full pipeline of development maturity.

- **Prototypers (N=10)**: Undergraduate students building ad-hoc systems for contests or course projects.
- **Researchers (N=3)**: PhD/Masters students focusing on VLA data quality and novel retargeting algorithms.
- **Practitioners (N=2)**: Industry experts managing deployed systems (e.g., robot fighting), focusing on reliability and safety.

The study was approved by the Institutional Review Board (IRB) at the first author's university. Table 2 summarizes the participant demographics.

Procedure

We conducted semi-structured interviews to investigate developers' workflows, hardware and software friction, and desired features for XR-based robot teleoperation systems. Participants first completed an IRB-approved consent process and a brief demographic questionnaire. Each interview lasted approximately 45–60 minutes and followed a consistent interview guide with 8 main questions and structured sub-questions, organized into three sections. All participants were asked the same core set of questions, with follow-up probes tailored to their specific experiences.

To ground the discussion, participants evaluated existing open-source teleoperation frameworks, including Open-TeleVision (Cheng et al., 2025), OPEN TEACH (Iyer et al., 2025), and XRoboToolkit (Zhao et al., 2025), reflecting on features that supported or hindered their development.

All interviews were audio-recorded, transcribed, and analyzed using an inductive, bottom-up thematic analysis approach (Corbin and Strauss, 1990), enabling us to identify recurring themes related to developer needs, pain points, and design implications.

FINDINGS

In this section, we present findings to answer the two research questions.

The Middleware Gap: Infrastructure as the Primary Barrier

A dominant theme was "Infrastructure Fragility." Unlike mature software domains, XR teleoperation lacks robust middleware.

Protocol Mismatches. Participants (P1, P4) described a "protocol hell" where XR headsets (often Android-based) failed to communicate with Robot Operating System (ROS) environments (Ubuntu) due to firewall rules and UDP/TCP mismatches.

> *On Windows, the connection works well. . . but in Ubuntu, the UDP connection cannot be established.* (P1)
>
> *The system can detect the specific devices, but it cannot transmit data from the headset (to the server).* (P4)

The "Good Enough" Trap. Prototypers frequently abandoned sophisticated frameworks like XRoboToolkit in favor of custom, fragile WebRTC scripts because the mature tools had steep learning curves or lacked "step-by-step" documentation.

> *Overall, the tool is convenient. . . However, its flexibility is limited. Modifying the code is difficult and debugging later is even harder. (P8)*
>
> *Many people have not used such systems before. . . A video demonstration would make a significant difference.* (P11)

Insight. The barrier to entry is not the robotics theory, but the network engineering required to bridge consumer headsets with research hardware.

> *Our teleoperation system has already been deployed. The only missing part is the connection between the headset and our server.* (P2)

The Data Utility Crisis

For researchers, the goal of teleoperation is data collection. Here, a critical tension emerged between immersive experience and data hygiene.

Tracking Noise vs. Learning Algorithms. P11 and P13 noted that while XR hand tracking feels "magical," it introduces noise and occlusion errors that are fatal for imitation learning algorithms. A dataset collected with jittery hand tracking is often useless for training precise policies.

Table 3: Synthesis of developer priorities.

Layer	Prototyper (Entry)	Researcher / Practitioner (Expert)
Infrastructure	Plug-and-play, Windows	Low-level access, Linux / ROS
Data	Visual confirmation	High quality, low latency
UI	Immersive 3D view	Telemetry, integrated control interfaces

> *If hand tracking is to be used for training dexterous hands, its accuracy must be high. . . even small errors in tracking may cause grasping to fail. (P11)*
>
> *If this cannot be solved from the data perspective, then the controller must be made more robust to tolerate imperfect inputs.* (P13)

Lack of Objective Latency Metrics. Surprisingly, developers lacked objective metrics for latency. P11 and P13 admitted to relying on "feeling" the lag rather than measuring it. This lack of instrumentation prevents systematic optimization, leaving developers to guess whether performance bottlenecks lie in the network, the render loop, or the robot controller.

> *The measurement of latency is still largely subjective.* (P13)
>
> *I think once teleoperation latency falls below a certain threshold, I can no longer perceive it.* (P11)

Cognitive and Sensory Augmentation

Participants envisioned future toolkits not just as controllers, but as augmented supervisors.

Sensory Substitution. Due to the lack of consumer-grade force feedback, participants (P3, P12) requested "Sensory Substitution"—using visual or auditory cues to represent weight, texture, or grip strength.

> *Force feedback from the robot hand or robotic arm can be conveyed in VR through vibration.* (P3)
>
> *Haptics should indeed be incorporated, adding tactile layers on top of the visual feedback.* (P12)

AI-in-the-Loop. P9 and P15 described a hybrid workflow where AI handles routine movement (using VLA models), and the human only intervenes for complex manipulation or error recovery. This shifts the developer's need from "continuous control" interfaces to "rapid intervention" interfaces.

> *Robots often encounter difficulties during grasping. . . Human operators can precisely control the gripper to complete the grasping task.* (P9)
>
> *When autonomous decision-making fails, teleoperation devices provide a safe fallback.* (P15)

DISCUSSION

Divergent Needs: Accessibility vs. Fidelity

Our analysis reveals a bifurcated set of developer requirements (Table 3) that challenges the notion of a single "universal" teleoperation toolkit. Prototypers primarily struggle with accessibility issues such as setup complexity, documentation, and network configuration, whereas researchers and practitioners contend with fidelity constraints, including latency, tracking occlusion, and data quality. While prior work examines teleoperation

interfaces and challenges in isolation (Arevalo Arboleda et al., 2021; Ghimire et al., 2025; Luo et al., 2020), it offers limited insight into how these trade-offs affect different developer groups in practice. Our findings suggest that future frameworks should explicitly support differentiated usage modes, such as simplified onboarding for prototyping and advanced configurations for rigorous experimentation and deployment.

Grounding Findings in HCI Literature

Our results align with and extend prior HCI and XR research. The lack of objective latency measurement mirrors findings by (Khasawneh et al., 2019), who show that human adaptation can obscure system inefficiencies. Similarly, participants' demand for haptics is consistent with (Louca et al., 2024), who report that users can adapt to visual-only feedback when visualization is effective. We extend this literature by highlighting a developer-centered implication: in the absence of reliable consumer haptics, toolkits must provide principled visual substitutes (e.g., force or compliance cues) to support usability and valid data collection.

The VLA Bottleneck

A key insight from this work is that the "Embodiment Gap" constrains progress in VLA models. Although large-scale datasets are critical for training generalist robots (Team, 2023), participants reported that difficulties in mapping human motion to diverse robot morphologies limit task diversity and data reuse. Unlike hand tracking, which benefits from emerging standards, whole-body retargeting lacks shared schemas, hindering scalable data collection. Without toolkits that better align embodiment, control, and data representation, XR teleoperation risks producing narrow datasets insufficient for advancing VLA models.

DESIGN IMPLICATIONS

Besides generic recommendations, we propose three architectural shifts:

Diagnostic Middleware

Toolkits must move beyond opaque, "black-box" connections and expose infrastructure as a first-class design concern. In addition to providing stable networking abstractions, frameworks should include built-in *diagnostic middleware*, such as latency probes that visualize round-trip time (RTT), packet loss, and jitter directly in the headset's head-up display (HUD).

This transforms subjective perceptions of "lag" into actionable engineering signals and aligns with calls for exposing networking and latency diagnostics within teleoperation workflows. Progressive onboarding mechanisms (e.g., setup wizards and video tutorials) can further reduce barriers for prototypers without obscuring low-level controls needed by experts.

Standardized Interaction Schemas

Just as OpenXR standardized headset inputs, the robotics community needs shared interaction schemas for XR teleoperation. Rather than ad hoc mappings, toolkits should define canonical control abstractions aligned with common robot morphologies (e.g., Humanoid UpperBody, Gripper Parallel) and support layered control mappings (direct, indirect). Such a "Teleop-XR" schema would mitigate today's interoperability crisis by enabling reusable mappings, canonical joint sets for full-body retargeting, and cross-project portability until broader standards mature.

Intervention-First Design

As teleoperation increasingly shifts toward AI-supervised workflows, interfaces should be designed around intervention rather than continuous control. Toolkits should codify human-in-the-loop patterns such as spectating previews, mode switching, and emergency freeze or retreat behaviors. The default state becomes monitoring, augmented by rich visual context and tunable feedback cues, while explicit UI affordances, such as a "Dead Man's Switch", enable immediate override of autonomous policies. This intervention-first approach prioritizes safety, reliability, and practical deployment while preserving human authority in failure cases.

LIMITATIONS AND FUTURE WORK

While our study presents empirical insights, our participant pool was small and imbalanced, with 10 of 15 participants being undergraduate students who served as prototypers, and 3 graduate researchers and 2 industry practitioners. Because the majority of participants are students rather than trained professional developers, the findings predominantly reflect entry-level development experiences and may not fully represent the challenges faced by experienced practitioners in production environments. As such, our findings should be interpreted as exploratory rather than exhaustive. However, this deliberate skew allows us to clearly identify barriers to entry, a critical but underexplored aspect of XR teleoperation adoption, and to contrast these with the high-fidelity needs of experts. Moreover, as shown in Table 2, even the undergraduate prototypers possess relevant hands-on experience in XR and/or robotics, suggesting that their feedback reflects informed development practice rather than uninformed opinion. Future work should expand recruitment to include a broader audience, explore toolkit adoption, and evaluate usability and effectiveness from end-user perspectives. Integrating these viewpoints will provide a more holistic understanding of XR-based teleoperation and inform the design of toolkits that serve both developers and end users.

CONCLUSION

This study shows that the primary challenges of XR-based robot teleoperation lie not only in interface design but in the reliability of the underlying development stack. By examining developer needs across prototyping, research, and deployment contexts, we identify a persistent middleware and data hygiene gap that limits XR teleoperation from serving as a robust control and data-collection pipeline. Our findings contribute design implications for future XR teleoperation toolkits and help bridge the gap between research prototypes and real-world systems, positioning XR as a viable foundation for next-generation robot learning.

REFERENCES

Arevalo Arboleda, S., Ru¨cker, F., Dierks, T. and Gerken, J. (2021), Assisting manipulation and grasping in robot teleoperation with augmented reality visual cues, *in* 'Proceedings of the 2021 CHI conference on human factors in computing systems', pp. 1–14.

Bach, T. A., Khan, A., Hallock, H., Beltr˜ao, G. and Sousa, S. (2024), 'A systematic literature review of user trust in ai-enabled systems: An hci perspective', *International Journal of Human–Computer Interaction* **40**(5), 1251–1266.

Black, K., Brown, N., Driess, D., Esmail, A., Equi, M., Finn, C., Fusai, N., Groom, L., Hausman, K., Ichter, B. et al. (2024), 'π_0: A vision-language-action flow model for general robot control', *arXiv preprint arXiv:2410.24164* .

Chen, J., Moemeni, A. and Caleb-Solly, P. (2023), Comparing a graphical user interface, hand gestures and controller in virtual reality for robot teleoperation, *in* 'Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction', pp. 644–648.

Chen, J. Y., Haas, E. C. and Barnes, M. J. (2007), 'Human performance issues and user interface design for teleoperated robots', *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* **37**(6), 1231–1245.

Cheng, X., Li, J., Yang, S., Yang, G. and Wang, X. (2025), Open-television: Teleoperation with immersive active visual feedback, *in* 'Conference on Robot Learning', PMLR, pp. 2729–2749.

Chi, C., Xu, Z., Pan, C., Cousineau, E., Burchfiel, B., Feng, S., Tedrake, R. and Song, S. (2024), 'Universal manipulation interface: In-the-wild robot teaching without in-the-wild robots', *arXiv preprint arXiv:2402.10329* .

Corbin, J. M. and Strauss, A. (1990), 'Grounded theory research: Procedures, canons, and evaluative criteria', *Qualitative sociology* **13**(1), 3–21.

Darvish, K., Penco, L., Ramos, J., Cisneros, R., Pratt, J., Yoshida, E., Ivaldi, S. and Pucci, D. (2023), 'Teleoperation of humanoid robots: A survey', *IEEE Transactions on Robotics* **39**(3), 1706–1727.

Fu, Z., Zhao, T. Z. and Finn, C. (2024), Mobile aloha: Learning bimanual mobile manipulation with low-cost whole-body teleoperation, *in* 'Conference on Robot Learning (CoRL)'.

Ghimire, A., Hou, A., Kim, I.-J. and Yoon, D. (2025), Avataroid: A motion-mapped ar overlay to bridge the embodiment gap between robots and teleoperators in robot-mediated telepresence, *in* 'Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems', pp. 1–26.

Iyer, A., Peng, Z., Dai, Y., Guzey, I., Haldar, S., Chintala, S. and Pinto, L. (2025), Open teach: A versatile teleoperation system for robotic manipulation, *in* 'Proc. Conf. Robot Learn.', PMLR, pp. 2372–2395.

Khasawneh, A., Rogers, H., Bertrand, J., Madathil, K. C. and Gramopadhye, A. (2019), 'Human adaptation to latency in teleoperated multi-robot human-agent search and rescue teams', *Automation in Construction* **99**, 265–277.

LeMasurier, G., Tukpah, J., Wonsick, M., Allspaw, J., Hertel, B., Epstein, J., Azadeh, R., Padir, T., Yanco, H. A. and Phillips, E. (2024), Comparing a 2d keyboard and mouse interface to virtual reality for human-in-the-loop robot planning for mobile manipulation, *in* '2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN)', IEEE, pp. 2197–2203.

Louca, J., Eder, K., Vrublevskis, J. and Tzemanaki, A. (2024), 'Impact of haptic feedback in high latency teleoperation for space applications', *ACM Transactions on Human-Robot Interaction* **13**(2), 1–21.

Luo, J., He, W. and Yang, C. (2020), 'Combined perception, control, and learning for teleoperation: key technologies, applications, and challenges', *Cognitive Computation and Systems* **2**(2), 33–43.

Moniruzzaman, M., Rassau, A., Chai, D. and Islam, S. M. S. (2022), 'Teleoperation methods and enhancement techniques for mobile robots: A comprehensive survey', *Robotics and Autonomous Systems* **150**, 103973.

Qin, Y., Yang, W., Huang, B., Van Wyk, K., Su, H., Wang, X., Chao, Y.-W. and Fox, D. (2023), Anyteleop: A general vision-based dexterous robot arm-hand teleoperation system, *in* 'Robotics: Science and Systems'.

Rea, D. J. and Seo, S. H. (2022), 'Still not solved: A call for renewed focus on user-centered teleoperation interfaces', *Frontiers in Robotics and AI* **9**, 704225.

Si, W., Wang, N. and Yang, C. (2021), 'A review on manipulation skill acquisition through teleoperation-based learning from demonstration', *Cognitive Computation and Systems* **3**(1), 1–16.

Team, G. R., Abeyruwan, S., Ainslie, J., Alayrac, J.-B., Arenas, M. G., Armstrong, T., Balakrishna, A., Baruch, R., Bauza, M., Blokzijl, M. et al. (2025), 'Gemini robotics: Bringing ai into the physical world', *arXiv preprint arXiv:2503.20020* .

Team, R.-X. (2023), 'Open-x-embodiment: Robotic learning datasets and rt-x models', *arXiv preprint arXiv:2310.08864*.

Wang, X., Shen, L. and Lee, L.-H. (2024), 'Towards massive interaction with generalist robotics: a systematic review of xr-enabled remote human-robot interaction systems', *arXiv preprint arXiv:2403.11384*.

Wang, X., Shen, L. and Lee, L.-H. (2025), 'A systematic review of xr-enabled remote human-robot interaction systems', *ACM Computing Surveys* **57**(11), 1–37.

Whitney, D., Rosen, E., Phillips, E., Konidaris, G. and Tellex, S. (2019), Comparing robot grasping teleoperation across desktop and virtual reality with ros reality, *in* 'Robotics research: the 18th international symposium ISRR', Springer, pp. 335–350.

Yeh, H.-H., Chang, Y.-W. and Liu, Y.-C. (2025), 'Intuitive hand motion-based teleoperation system for human-mobile manipulator interaction using mixed reality', *Control Engineering Practice* **164**, 106467.

Zare, M., Kebria, P. M., Khosravi, A. and Nahavandi, S. (2024), 'A survey of imitation learning: Algorithms, recent developments, and challenges', *IEEE Transactions on Cybernetics*.

Zhao, Z., Yu, L., Jing, K. and Yang, N. (2025), 'Xrobotoolkit: A cross-platform framework for robot teleoperation', *arXiv preprint arXiv:2508.00097*.