Wheel2VR: Gaze-Controlled Dual Robotic Arms for Assistive Wheelchair Interaction in Virtual Reality

Ahmed Seyit Kücük
Universität Bremen
Bremen, Germany

Dmitry Tschernobai
Universität Bremen
Bremen, Germany

Hannah Köper
Universität Bremen
Bremen, Germany

Xinyun Yu
Universität Bremen
Bremen, Germany

Jayon K Vinod
Universität Bremen
Bremen, Germany

Ahmed Ibrahim
Universität Bremen
Bremen, Germany

Arne Winter
Universität Bremen
Bremen, Germany

Abstract

People with tetraplegia have a strong requirement for independence in daily activities using wheelchair-mounted assistive robotic arms. However, most arms utilize inaccessible inputs, such as joysticks, and offer limited functionality due to their single-arm configuration. To address these issues, we present a gaze-controlled dual robotic arm system in VR using a shared control strategy, which integrates an HTC Vive Pro Eye headset with two 6-DOF Kinova JACO arms based on inverse kinematics and behavior trees. Moreover, eye-tracking and head-tracking were employed as gaze input modalities, featuring both Manual and AI modes. To evaluate the system, 18 participants performed a pill bottle opening task under both manual and shared conditions with their selected input. Evaluation metrics included completion time, NASA-TLX, SUS, input mode preference, prior VR experience, and qualitative interviews across four interaction combinations: eye-manual, eye-shared, head-manual, and head-shared. Results show that 62% of VR novices preferred eye-tracking, compared to only 20% of experienced VR users, though no significant differences occurred in task performance, such as completion time or the number of buttons pressed. Furthermore, the Shared condition demonstrated significantly lower workload (NASA-TLX median: 36.67 vs. 44.17), descriptively higher usability (SUS median: 68.75 vs. 65.00), and shorter completion times (3.29 vs. 4.43 min), despite participant feedback regarding unpredictable or inaccurate intelligent arm behaviors.

1 Introduction

Spinal cord injuries (SCI) have many forms. One of them being tetraplegia, which results from damage to the cervical spine, causing partial or complete loss of function in all four extremities. SCI is affecting about 15.4 million individuals around the globe with steadily increasing numbers of new cases according to the World Health Organization [Fosbrooke et al. 2026]. Furthermore Germany alone has about 7.8 million people with severe disabilities (as of 2021) [Goldau 2024]. Affected people with tetraplegia face many challenges in their daily lives, including limitations in mobility and social participation as well as personal care [Fosbrooke et al. 2026]. The use of robots in an assistive capacity could provide these people with increased independence and overall autonomy especially when receiving individual care. Currently there are technologies which rely on the use of robotic arms on a wheelchair (WMRA) that can be controlled via the wheelchair joystick [Goldau 2024]. People with tetraplegia form an exception because they can't use these kind of controls. In our Masters project Wheel2VR we want to implement a prototype of the WMRA system in Unity VR which includes two robot arms mounted on a wheelchair that can be controlled via eye tracking or head tracking. While one robotic arm is manually guided to execute a primary task, the second arm acts as a reactive assistant, semi-autonomously coordinating its movements to support the primary action. For instance, if one arm secures a pill bottle, the other can independently unscrew the cap and dispense the medication. Having implemented this system in Unity VR, we plan to conduct a user study to investigate whether gaze-based shared assistance for dual-arm manipulation lowers the user's workload and interaction effort compared to fully manual control. We hypothesize that participants will indeed experience reduced workload and effort in the shared-assistance condition.

2 Related Work

Prior work has shown that gaze can serve as a viable input modality for assistive robotic manipulation, particularly for object selection and high-level action triggering. Several systems use eye tracking to identify user-selected targets, often combining gaze with confirmation mechanisms or additional modalities to avoid unintended activation.

Hansen et al. present a single-arm assistive system that combines eye tracking, muscle activity (sEMG), and head movements for everyday tasks, showing that gaze can be integrated into accessible control workflows for users with tetraplegia while also underlining the practical role of multimodal input [Hansen et al. 2024]. Wang et al. similarly demonstrate direct gaze-guided target selection for robotic grasping [Wang et al. 2023]. At a broader level, the scoping review by Fischer-Janzen et al. synthesizes 39 approaches and identifies recurring patterns such as object selection by gaze, command triggering through fixation, and the use of additional confirmation steps or other modalities [Fischer-Janzen et al. 2024]. Across this literature, the Midas Touch problem remains a central challenge, which is why many systems rely on dwell times, confirmation mechanisms, or multimodal control.

Taken together, these studies establish the feasibility of gaze-based assistive interaction, but they also highlight recurring challenges related to unintended activation, precision, and interaction effort.

A second line of work focuses on shared control and transparency in human-robot interaction. These approaches do not rely on purely manual control but instead combine user input with robotic assistance. A central concern in this literature is how the robot communicates intent so that assistance remains understandable and trustworthy.

Kronhardt et al. describe shared control as a way to preserve user autonomy while reducing low-level control demands in assistive manipulation [Kronhardt et al. 2023]. Zolotas and Demiris similarly argue that assistance becomes easier to trust when the robot makes its planned motions and grasp goals visible to the user through transparent intent communication [Zolotas and Demiris 2020]. Fischer-Janzen et al. further explore how gaze can be used to select targets while the robot contributes assisted execution [Fischer-Janzen et al. 2026]. Fuchs and Belardinelli extend this direction by using eye movements to infer likely user goals during teleoperated manipulation [Fuchs and Belardinelli 2021]. Together, these studies show that the quality of assistance and the clarity of intention communication are closely linked.

While prior work suggests that transparent intent communication can improve acceptance of robotic assistance, less is known about how users perceive proposal-based assistance in a gaze-driven interaction workflow.

Further relevant work examines either dual-arm manipulation or virtual reality as a testbed for assistive interaction. Research on dual-arm manipulation shows that two-arm coordination can be beneficial for certain tasks, but also introduces additional complexity in planning and coordination. VR-based systems, in turn, provide a controlled way to prototype and evaluate assistive interaction concepts before deployment in real robotic settings.

Wan and Harada compare single-arm and dual-arm regrasp planning and show that dual-arm manipulation can be useful for some tasks, but also introduces geometric constraints and collision risks [Wan and Harada 2016]. Zhou and Aburumman discuss realistic grasping in virtual reality and outline how VR can be used to model physically plausible human and robotic grasp interactions [Zhou and Aburumman 2024]. Wan et al. present a VR-based teleoperation system for robotic manipulation and highlight how immersive simulation can support intuitive interaction design and safe early-stage evaluation before deployment on physical hardware [Wan et al. 2024].

However, these lines of work rarely address the combination of dual-arm assistive interaction, gaze-based input, and user-supervised shared assistance in a single experimental setup.

Taken together, prior work motivates gaze-based assistive control, transparent shared autonomy, and coordinated manipulation, but leaves open how users experience preview-based shared assistance in a dual-arm assistive task. Wheel2VR addresses this gap through a VR prototype that compares fully manual control with gaze-based shared assistance.

3 System Design

3.1 System Overview

Wheel2VR simulates a wheelchair-mounted assistive manipulation system in virtual reality. The environment contains two robotic arms, a pill bottle with a removable cap, and a small set of pills on a kitchen counter. The core task is a coordinated two-handed manipulation sequence in which the bottle is stabilized, the cap is removed, and the pills are dispensed.

The system compares two autonomy conditions. In the manual condition, users control both robotic arms sequentially by moving a ghost preview through gaze-activated directional controls. In the AI-assisted condition, users can request shared-control suggestions for the currently selected arm through an AI button instead of positioning that arm manually step by step. In specific task situations, most notably when the pill bottle is selected and the task state is sufficiently clear, the system can also propose a coordinated two-arm action. In both conditions, robot motion is mediated through gaze-based interaction and a preview-before-confirmation workflow.

Two gaze input modalities are supported: eye tracking and head-direction tracking. Combined with the two autonomy conditions, this yields four possible interaction conditions that are later evaluated in the study.

3.2 Hardware Setup

The prototype runs on an HTC Vive Pro Eye headset with integrated Tobii eye tracking. Participants remain seated throughout the experiment. No handheld controllers are used; all interaction is performed through either eye tracking or head-direction tracking, depending on the active input mode.

3.3 Scene Setup

The virtual scene is implemented as a minimal kitchen environment with reduced visual clutter so that attention remains on the manipulation task. Two six-degree-of-freedom robotic arms are positioned in front of a countertop. A pill bottle, its cap, and several pill capsules serve as the task objects.

The robotic arm models are based on the Kinova JACO design, which is representative of assistive arms mounted on powered wheelchairs. For the user interface, we use MRTK3 components to provide a consistent world-space UI and robust gaze-based interaction elements.

All arm motions follow a ghost-preview-then-confirm pattern. When the user initiates a manual repositioning step or requests AI support, a semi-transparent preview arm moves to the proposed pose first, while the physical arm model remains stationary. The user must explicitly confirm the preview before the real arm executes the movement. Separate ghost colors are used for manual and AI-generated proposals so that the origin of each action remains visually clear.

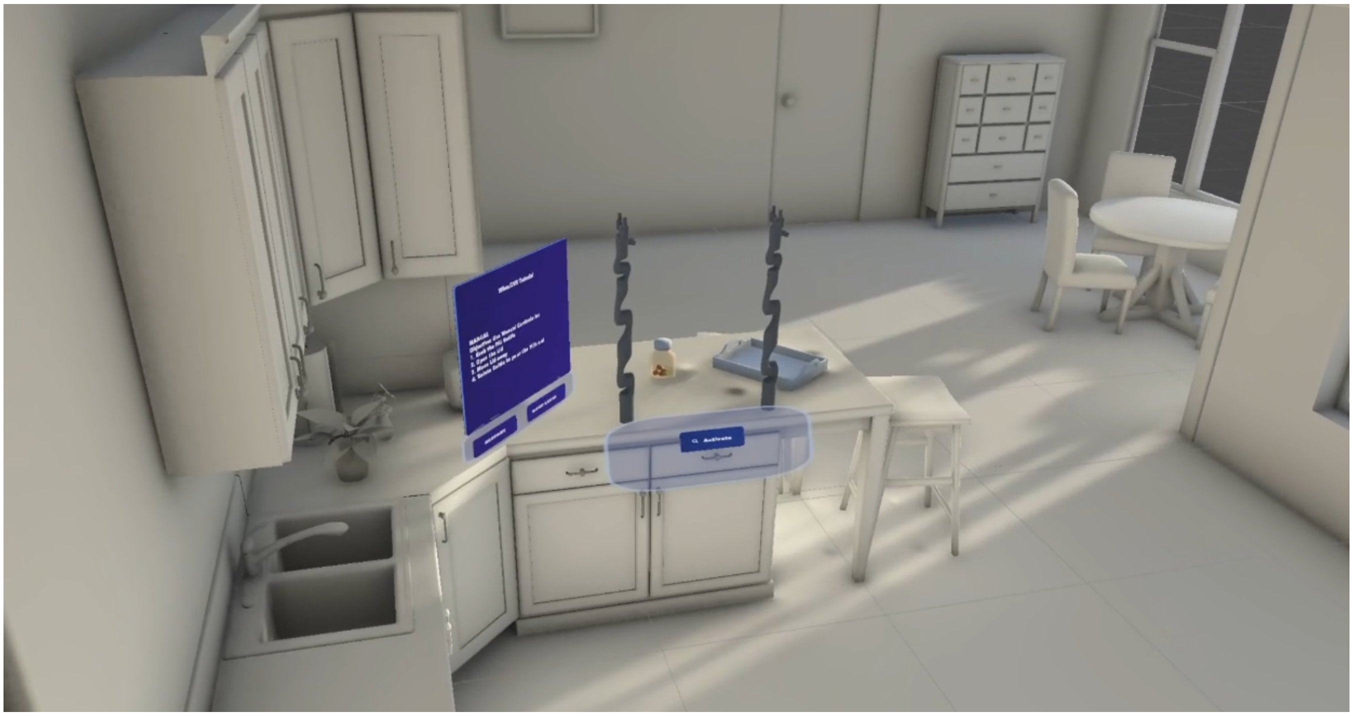

Figure 1: Overview of the Wheel2VR prototype environment with the dual-arm workspace, pill-bottle task setup, and world-space control interface.

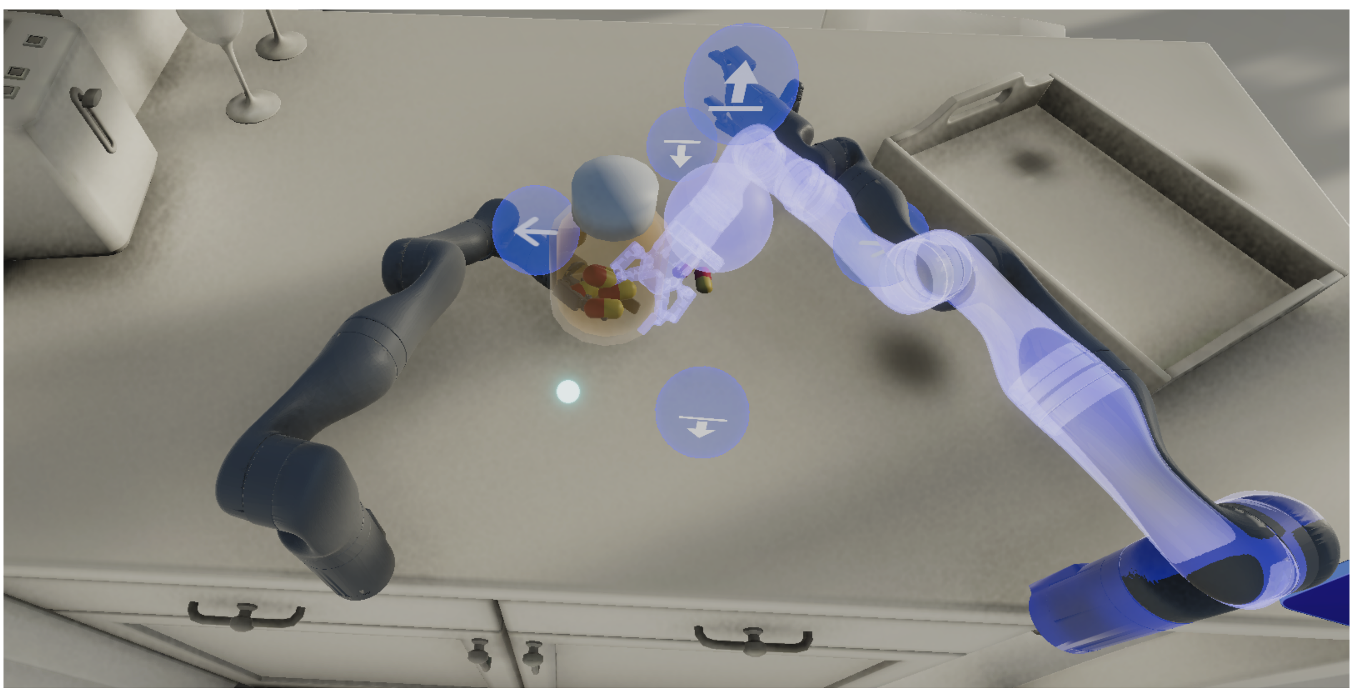

Figure 2: Manual positioning with directional gaze controls and a semi-transparent ghost preview before movement confirmation.

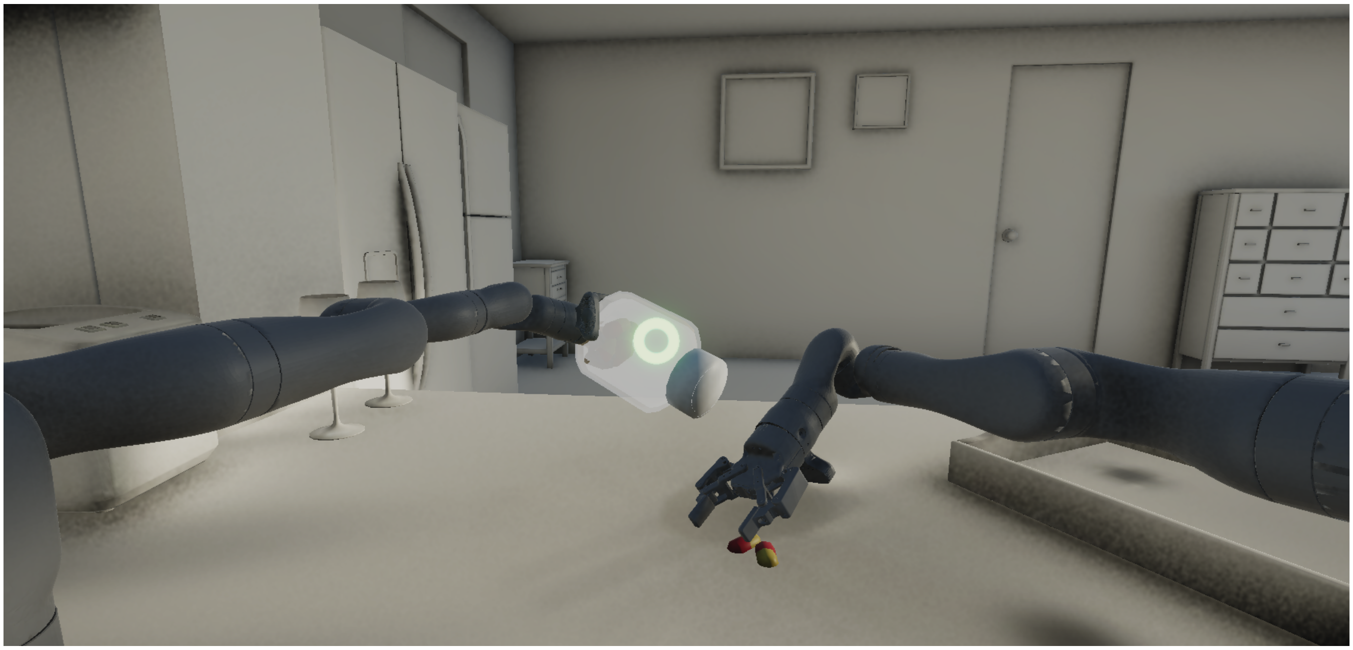

Figure 3: User's point of view while rotating the pill bottle to dispense the pills.

3.4 Control Strategy

Two control strategies are compared.

In the manual condition, users operate both robotic arms sequentially. After selecting an arm, they reposition its ghost target through six gaze-activated directional controls around the wrist area of the robot. As long as the user fixates on one of these controls, the ghost preview moves continuously with a small constant velocity in the corresponding direction. Once the desired pose is reached, the user confirms the previewed movement and then switches to the other arm as needed to complete the task.

In the AI-assisted condition, users still select which arm they want to work with, but they can invoke AI support for that arm instead of positioning it manually. The system then proposes a suitable target pose based on the selected object, the current task context, and the inferred interaction goal. These proposals are not executed automatically. Instead, they are shown through the same ghost-preview mechanism and require explicit confirmation. In a small number of clearly defined cases, such as coordinated bottle-and-cap manipulation, the system may generate a combined proposal that involves both arms.

3.5 Software Architecture

The prototype is implemented in Unity and organized into four main modules: a gaze interaction layer, robot control based on inverse kinematics, a shared-control assistance module, and a world-space user interface.

The gaze interaction layer identifies the current target of the user's gaze and converts sustained fixation into interaction events. The robot control layer translates target positions into joint configurations for the two arms. The assistance module generates pose suggestions for the selected arm and, in specific task states, coordinated proposals for both arms. The user interface exposes all relevant actions, confirmations, and system states inside the VR scene.

3.6 Interaction Techniques

All interaction relies on gaze-directed dwell selection. An `XRRay Interactor` attached to the camera rig casts a ray in the current gaze direction. The `GazeDwellTimer` measures how long the user fixates on a tagged target. After an onset delay of 150 ms, the system starts accumulating dwell time. Once the dwell threshold of 1.5 s is reached, an `OnDwellComplete` event is broadcast to the corresponding UI element or scene object. The cursor also provides progressive visual feedback by expanding while the dwell timer is running.

Robot motion is controlled through inverse kinematics. Instead of manipulating joints directly, the user moves a spatial target that represents the desired gripper position. In manual mode, this target is shifted incrementally through the six directional gaze controls, while in AI-assisted mode it can also be set through a system-generated proposal. The IK solver then computes the corresponding joint configuration automatically. When the gripper is sufficiently close to an object, a grab action attaches that object to the gripper, allowing the bottle and cap to be manipulated as part of the task.

3.7 Shared-Control Assistance

In the AI-assisted condition, support is implemented as a shared-control mechanism rather than as a permanently autonomous helper arm. When the user selects an arm and activates the AI function, the system generates a movement proposal for that arm instead of requiring purely manual positioning. The implementation uses a behaviour tree with four node types: `BehaviourTree` as the root, `Selector` to test alternative branches, `Sequence` to execute ordered action chains, and `Leaf` nodes that perform concrete checks or actions and return `Success`, `Failure`, or `Running`.

The behaviour tree evaluates the selected object, the current task state, and the likely interaction goal in order to generate a suitable target pose. In the common case, this means finding a good position for the currently selected arm from which the object can be approached or manipulated through inverse kinematics. In clearly structured scenarios, such as pill-bottle handling, the same logic can also generate a coordinated two-arm proposal, for example when one arm should stabilize the bottle while the other opens the cap. This provides a transparent rule-based structure that is suitable for a prototype because the resulting behavior remains relatively predictable and easy to debug.

All AI-generated actions are mediated through user approval. Before execution, the system presents the proposed motion as a ghost preview. Users can then confirm or reject the action via gaze interaction, ensuring that the assistance remains user-initiated and supervisory rather than fully automatic.

3.8 3D User Interfaces

The main interface is presented as a world-space panel within the user's field of view so that interactions can be performed without looking away from the workspace. The menu contains gaze-activated controls for selecting the active arm, enabling manual positioning, grabbing or releasing objects, confirming a ghost pose, and activating the AI-supported shared-control workflow.

Object states are communicated directly in the scene through visual highlighting. A blinking white outline marks an object as available for selection, a yellow outline marks the currently selected target, a red outline indicates that the object is outside the reachable workspace, a green highlight indicates that the object can be grasped, and a black highlight indicates that the object is currently attached to a gripper.

Only one robotic arm can be actively controlled at a time. In manual positioning mode, six directional controls appear around the ghost target near the robot wrist and represent translations along the three spatial axes. Fixating on one of these controls moves the ghost target continuously in that direction at a low constant speed, and looking away stops the motion immediately. This allows the user to approximate controller-like directional input through gaze alone. Once the desired pose is reached, the user confirms it through dwell selection, after which the real arm moves to the previewed position.

3.9 Design Decisions

Gaze interaction was chosen because the target application concerns users with limited or no hand-based input capabilities. The interface therefore prioritizes hands-free command selection and explicit confirmation.

The ghost-preview mechanism was introduced to reduce unintended robot motion and to increase transparency. Users can inspect each proposed pose before execution, which is especially important in the AI-assisted condition where suggested movements are generated by the assistance logic.

Behaviour trees were selected for the assistance logic because they provide a readable and deterministic decision structure. For a prototype that is intended to study user perception and usability rather than maximize autonomy, this design offers a practical balance between transparency and implementation effort.

3.10 Limitations

The current system is a simplified simulation of an assistive robotic setup rather than a physically accurate robotic platform. Object interaction, collision handling, and grasping are abstracted compared to real-world manipulation.

The assistance logic relies on a predefined behaviour tree and does not use perception-driven planning or learning-based adaptation. As a result, the system can demonstrate the interaction concept, but it does not yet represent the full complexity of real assistive autonomy.

Finally, the interface and task are deliberately constrained to a single tabletop scenario. This improves experimental control, but limits how far the findings can be generalized to broader daily-living activities.

4 Methodology

This section describes the study design, participant recruitment, experimental procedure and metrics used to evaluate the Wheel2VR system. The study is exploratory in nature and aims to assess the general usability of a gaze-controlled dual-arm wheelchair-mounted robotic assistant in virtual reality. As a secondary focus, we examine user preferences between two gaze-based input modalities: eye tracking and head tracking.

4.1 Study Design

We conducted a controlled lab study using a within-subjects design with respect to the autonomy condition. Each participant performed the main task twice: once with both arms controlled manually and once with AI-assisted control of the secondary arm. The order of these two conditions was randomized per participant via coin toss to counterbalance potential learning or fatigue effects.

The choice of input modality (eye tracking vs. head tracking) followed a different structure. During the preceding familiarization phase, each participant experienced both input modes in randomized order. After completing both familiarization sessions, participants stated which input mode they preferred. The main task was then carried out entirely in the preferred mode. As a consequence, the comparison between eye tracking and head tracking is not a controlled within-subjects comparison on the main task but rather a self-reported preference measure collected after the familiarization phase.

This design results in four possible condition combinations, shown in Table 1. However, each participant only occupies one column (their preferred input mode) and both rows (manual and semi-autonomous).

Table 1: Condition grid. Each participant completes both autonomy conditions (rows) but only in their preferred input mode (one column).

	Head Tracking	Eye Tracking
Manual	HM	EM
Semi-Autonomous	HA	EA

4.2 Participants

We recruited 18 participants from the university campus, mostly fellow students and acquaintances. The age range was approximately 20 to 35 years. Prior VR experience varied across participants and was recorded as part of the pre-study questionnaire.

The target user group of the Wheel2VR system are people with tetraplegia or comparable motor impairments. We did not recruit from this population due to limited access within the scope of a university project. Instead, we evaluated the system with able-bodied participants as a first step to assess general usability, interaction design and task flow before conducting studies with the intended user group. This is a clear limitation: results regarding physical comfort, fatigue and real-world applicability cannot be transferred directly to users with spinal cord injuries. We discuss this further in Section 6.

4.3 Apparatus

Participants were seated at a desk throughout the experiment and did not use hand controllers at any point. All interaction was performed through gaze input exclusively.

4.4 Procedure

Each session followed a fixed sequence and lasted approximately 40 to 50 minutes per participant.

4.4.1 Introduction. Participants received a verbal explanation of the system concept, the task they would perform, and the two input modes available to them. They were informed that participation was voluntary and that they could stop at any time.

4.4.2 Familiarization. The familiarization phase introduced the core interactions through three tutorial levels of increasing complexity:

- **Calibration:** Calibration of the system depending on input mode.
- **Level 1:** Select an object by gazing at it (dwell-based selection).
- **Level 2:** Manually control the right arm to reach a glass and grab it.
- **Level 3:** Use the AI mode to open a kettle lid, grab it and lift it.

Each participant completed the familiarization phase twice, once with eye tracking and once with head tracking. The order was randomized by coin toss. After experiencing both input modes, participants were asked which mode they preferred.

4.4.3 Main Task. The main task simulated a pill bottle interaction requiring coordination of both robotic arms. The task consisted of the following steps:

(1) Grab the pill bottle with one arm.
(2) Grab the bottle lid with the other arm.
(3) Lift the lid off the bottle.
(4) Lift the pill bottle.
(5) Rotate the pill bottle to dispense the pills.

Each participant performed this task twice in their preferred input mode: once with fully manual control of both arms, and once with AI-assisted control mode. The order of these two runs was determined by coin toss.

4.4.4 Per-Condition Questionnaires. Immediately after completing each condition, participants filled out the NASA Task Load Index (NASA-TLX) and the System Usability Scale (SUS). This was done twice per participant: once after the manual condition and once after the AI-assisted condition, before proceeding to the next run. Administering the questionnaires directly after each condition rather than once at the end allows for a paired comparison of perceived workload and usability between the two autonomy modes.

To reduce potential confusion between the two administrations, questionnaire sheets were visually distinguished by color coding per condition.

4.4.5 Interview. After both task runs and both sets of questionnaires were completed, a short informal interview was conducted to gather qualitative feedback on the interaction, perceived strengths and weaknesses of the system, and general impressions of the two autonomy modes.

4.5 Metrics

We collected both quantitative and qualitative data.

4.5.1 Quantitative Metrics.

- **Task completion time:** Recorded separately for the manual and the AI-assisted condition. This is the primary metric for comparing autonomy modes within subjects.
- **NASA-TLX:** Six subscales (mental demand, physical demand, temporal demand, performance, effort, frustration) rated on a 21-point scale. Collected separately after each autonomy condition, enabling a paired comparison.
- **SUS:** Ten-item questionnaire yielding a single usability score between 0 and 100. Collected separately after each autonomy condition, enabling a paired comparison.
- **Input mode preference:** Binary choice (eye tracking or head tracking) stated after the familiarization phase.

4.5.2 Qualitative Metrics.

- **Prior VR experience:** Self-reported in the pre-study questionnaire.
- **Interview responses:** Open-ended feedback on system usability, preferences, difficulties, and suggestions for improvement. Responses were collected as notes and later grouped thematically.

4.6 Analysis Approach

Task completion times, NASA-TLX scores and SUS scores between the manual and AI-assisted conditions are compared using paired statistical tests, as each participant completed both conditions and filled out questionnaires after each one. For NASA-TLX raw scores are used to identify which aspects of workload are most affected by the autonomy mode.

Input mode preference is reported descriptively as a frequency count. Since participants only performed the main task in their preferred mode, no direct performance comparison between eye tracking and head tracking on the main task is possible.

Qualitative interview data is summarized by identifying recurring themes and notable observations across participants.

5 Results

5.1 Demographics

Out of 18 participants only one was female and all were between 18 and 34 years old. Eight of the participants had never used a VR setup before, nine have used it up to five times total and only one person regularly using VR. No participants were left handed, with one person being ambidextrous and the other participants being right handed. Since we didn't have a substantial diversity of ages, genders and handedness these factors won't be included in our analysis.

5.2 Choice of input method

After the familiarisation period seven participants chose eye-tracking as their preferred input method, most often citing it being more easy to use or more intuitive as well as it being faster as reasons for choosing it, as well as some participants considering it having more accuracy for the tasks.

The eleven participants that chose head-tracking also cited their preferred input method as being easier to use or feeling more intuitive, with the two other most mentioned reasons being it feeling more stable than eye-tracking and preferring not having to focus on the buttons with their eyes, instead being able to look around while completing the tasks.

The sequence of introduction to the input methods doesn't seem to have a big effect on what input method was preferred, see table 2. It is an almost even split between people choosing the first or second method they tried, in both cases only one person more choosing the second input method they tried.

But there is a clear distinction between the choices of participants that were completely unfamiliar with VR and participants who had tried VR before. 62% of participants that had never tried VR chose eye-tracking as their preferred input method, while only 20% of participants who had used VR previously chose eye-tracking.

Sequence	Input method	Absolute	Relative
Eye, Head	Eye-tracking	3	33%
Eye, Head	Head-tracking	6	67%
Head, Eye	Eye-tracking	4	44%
Head, Eye	Head-tracking	5	56%
Never used	Eye-tracking	5	62%
Never used	Head-tracking	3	38%
Used before	Eye-tracking	2	20%
Used before	Head-tracking	8	80%

Table 2: Choice of preferred input method by sequence of familiarisation and familiarity with VR

	TLX			SUS		
	Avg.	**Med.**	**STD**	**Avg.**	**Med.**	**STD**
Total	43	42	13	68	68	16
Eye, Manual	45	43	10	71	65	15
Eye, Shared	33	37	12	74	68	18
Head, Manual	48	47	15	65	60	18
Head, Shared	42	42	13	66	70	14
Eye	39	38	12	73	66	16
Head	45	45	14	65	68	16
Manual	47	44	13	67	65	17
Shared	38	37	13	69	69	16
Never used	40	38	12	70	69	16
Used before	45	45	15	66	65	16

Table 3: Total and stratified averages, medians and standard deviations of NASA TLX and SUS

5.3 Task Load & Usability

The average NASA TLX score for our system is 43 out of 100 points, with the mental load being the highest sub score with 5.6 out of ten points, and frustration being the lowest sub score with 3.1 out of ten points.

On the System Usability Score our system received an average of 68 out of 100 points, see table 3.

Across the 18 paired questionnaire responses, participants reported lower workload in the shared condition than in the manual condition, as reflected in NASA-TLX scores (44 vs. 37 median; Wilcoxon signed-rank test, W = 31.5, p = .032). SUS scores were descriptively higher in the shared condition (69 vs. 65 median), but this difference was not statistically significant (W = 47.5, p = .493). Participants who selected eye tracking showed lower workload and higher usability than participants who selected head tracking. However, this comparison is exploratory only, because participants completed the main task in their preferred input mode rather than under a controlled within-subject comparison. The same observations and limitations apply to participants who had no prior experience with VR, compared to participants who had previous VR experience.

5.4 Task data

Task-performance analyses were based on a reduced core sample of 16 participants after excluding two datasets due to logger failure and incomplete recording, whereas questionnaire analyses were based on 18 paired post-task responses. The data gathered during the experiment supports the shared mode being less work for participants. Participants using shared mode needed a median of 60 button clicks until successfully pouring out pills, when using manual mode they needed a median of 241 button clicks until finishing the task. This persists regardless of the sequence in which the participant experienced the two modes, see figure 4.

This pattern also persists in completion time, although not to a significant level. It was lower in the shared condition (median 3.29 vs. 4.43 minutes), see figure 5, but this difference was not statistically significant (Wilcoxon signed-rank test, W = 54.0, p = .495; n = 16).

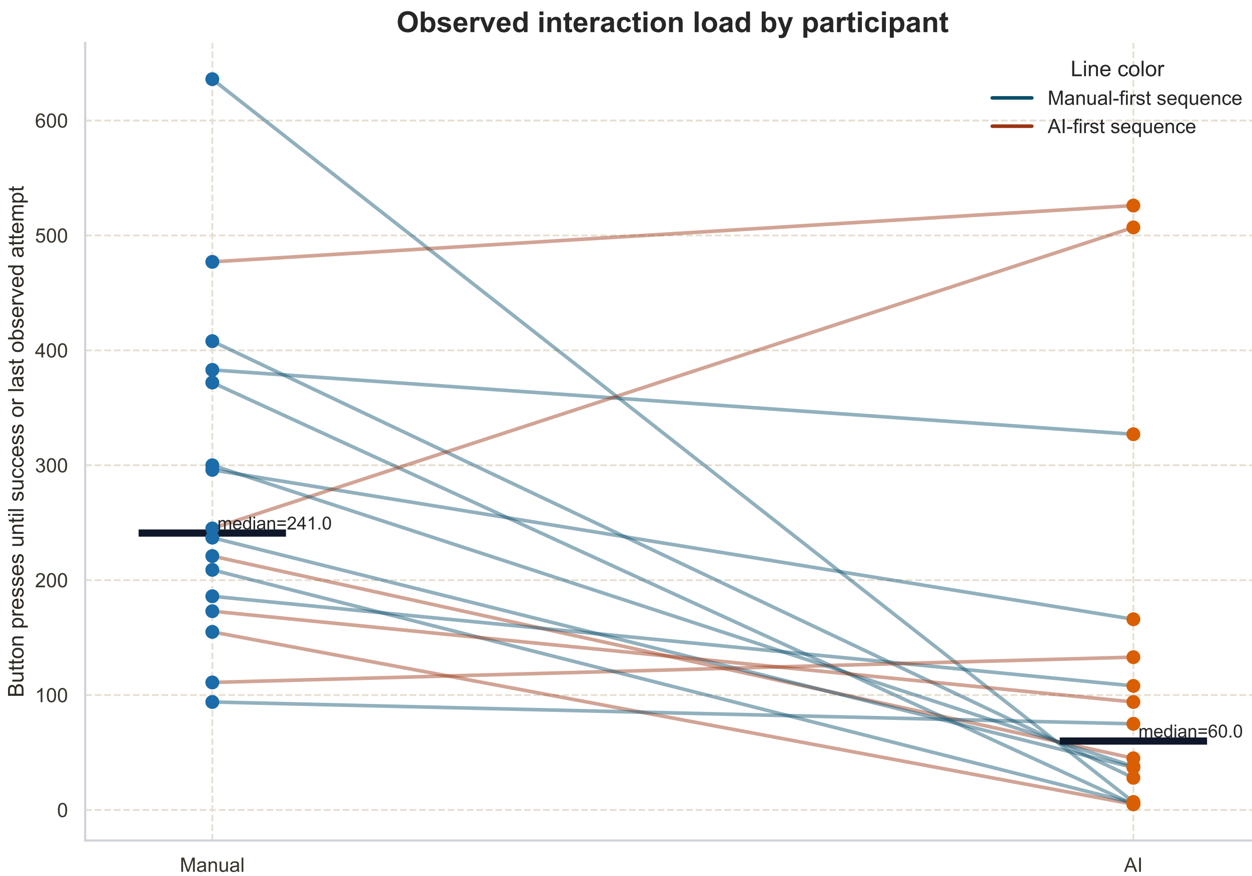

Figure 4: Observed interaction load by participant

When comparing the completion times of the first and second experiment of the participants a learning effect can be observed, with participants generally completing the second task more quickly or efficiently($W = 24.0$, $p = .041$) and required fewer button presses ($W = 23.0$, $p = .035$), regardless of whether the second experiment was in shared or manual mode, see figure 6.

In the main paired task analysis, participants required substantially fewer button presses in the shared condition than in the manual condition (median 60 vs. 241; Wilcoxon signed-rank test, $W = 17.5$, $p = .009$; $n = 16$). Completion time was descriptively lower in the shared condition (median 3.29 vs. 4.43 minutes), but this difference was not statistically significant ($W = 54.0$, $p = .495$; $n = 16$). The number of attempts until first success did not differ between conditions (both medians = 1; $W = 17.5$, $p = .943$; $n = 15$).

5.5 Post experiment interviews

During the post experiment interview half of all participants expressed that the intelligent arm behaved in unexpected ways. Participants still saw the utility of the intelligent features reducing the workload of the user. Complaints about utility were about the specific implementation, not about intelligent features generally.

Participants felt generally in control of the arm, even if the intelligent arm behaved unexpectedly, since the ghost arm showed a preview of the behaviour that participants could still adjust or cancel, before moving on to the robot arm executing the previewed movement.

When discussing the differences between manual and shared mode participants mentioned most often how convenient the shared mode is, reducing time spent trying to position the arms correctly in space than when using manual mode.

Participants also reported the shared mode to not always be accurate, sometimes needing manual adjustment to be able to do the task, this happened the most often when trying to grab the bottle cap.

The automatic opening of the bottle cap by the intelligent arm did not always work, leaving participants confused about the arm behaviour and making the behaviour feel unpredictable.

Participants who preferred doing the task manually highlighted that the manual mode felt more accurate and straightforward to them.

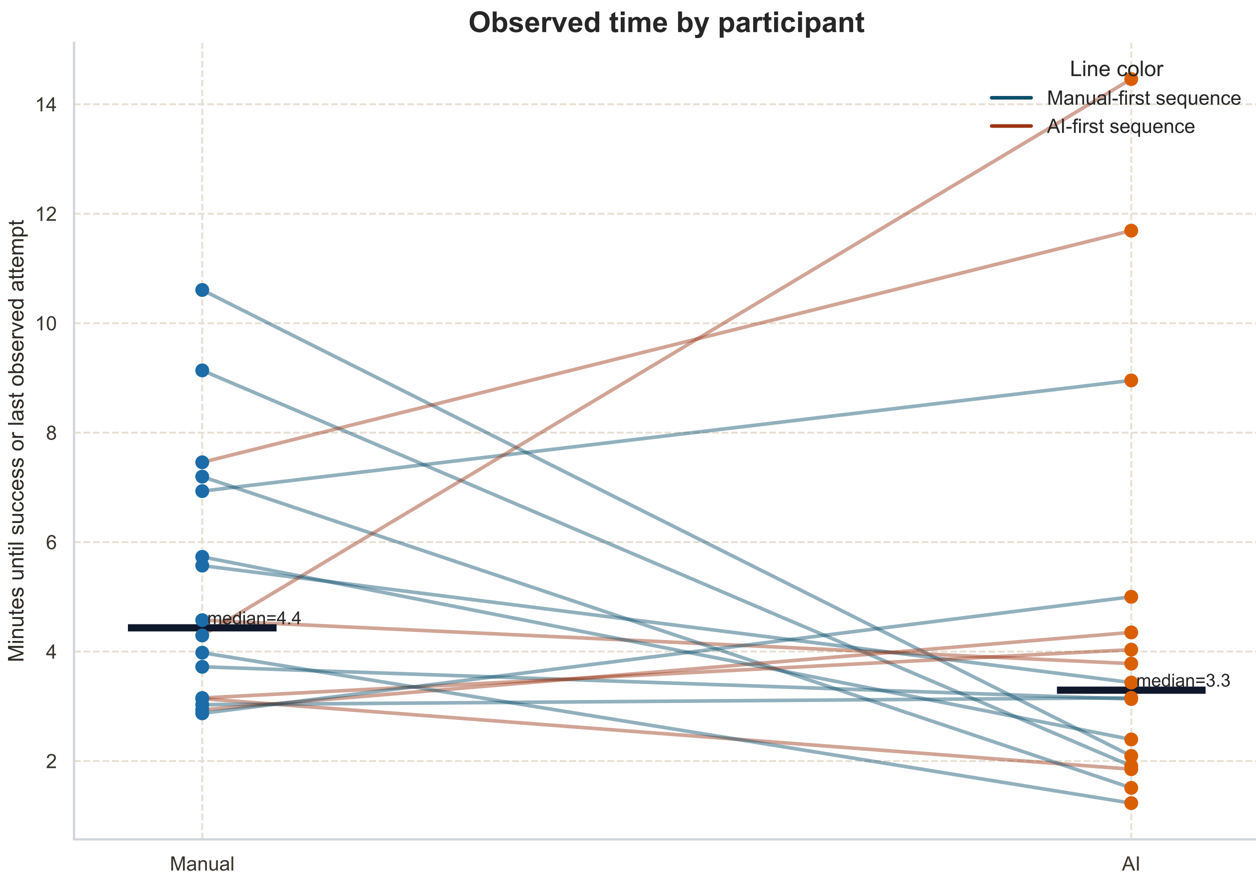

Figure 5: Observed time by participant

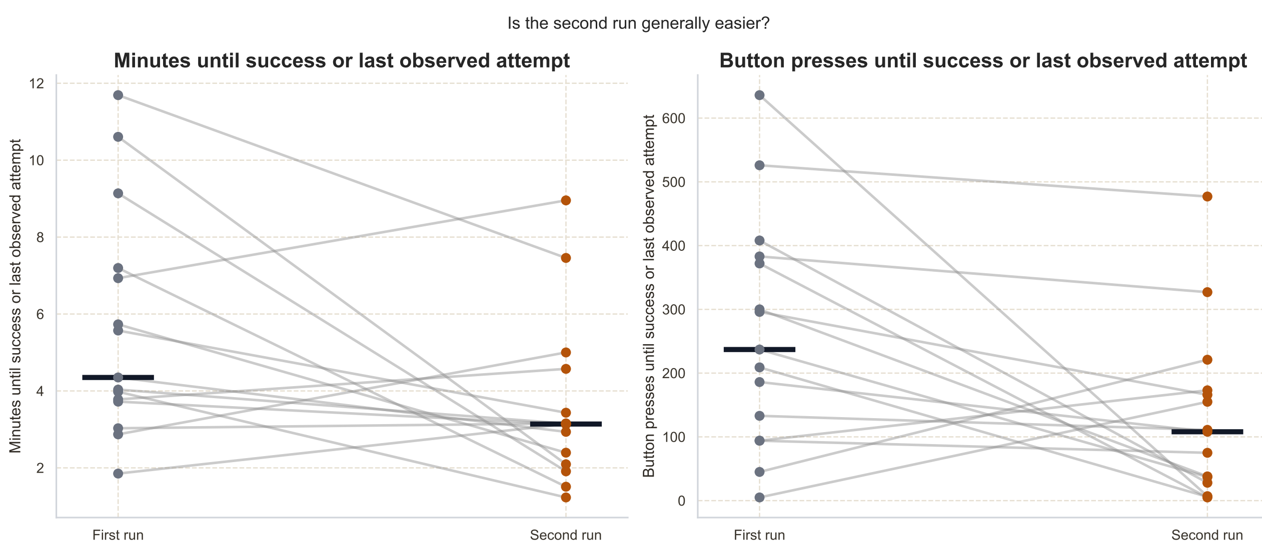

Figure 6: Observed interaction load and time by sequence and participant

6 Discussion

6.1 Head-tracking and Eye-tracking

Before the experiment, we hypothesized that participants would prefer head-tracking over eye-tracking. Our hypothesis was confirmed, as 7 participants chose eye-tracking while 11 participants chose head-tracking following the familiarization phase. According to the feedback, most participants preferred head-tracking because it is more intuitive, stable, and provides a flexible field of fixation, allowing them to look around while completing the task.

On the other hand, the cursor frequently jittered in the eye-tracking condition during the experiment, which may be a hardware limitation. As Qian and Teather concluded that head-only selection is likely to continue dominating VR interaction due to the absence of more precise eye trackers with superior calibration methods [Qian and Teather 2017].

Furthermore, we surprisingly found that prior VR experience is a factor in the selection of input methods. The Table 2 shows that 62% of VR novices chose eye-tracking, compared to only 20% of experienced VR users. This difference can be explained by comparing participants' personal expectations with their previous experiences, which will be investigated in future work.

In addition, a contradiction was observed that eye-tracking reported a lower workload and better usability than head-tracking by six and eight points respectively, even though there were no differences in performance metrics, such as speed and the number of buttons pressed. This indicates that when participants choose their preferred input, they can achieve similar task performance. However, it remains to be seen if the subjective advantages of eye-tracking outweigh the reliability of head-tracking under equal performance and user preference since our current data is unbalanced (eye-tracking vs. head-tracking is 7 : 11).

6.2 Manual Control and Shared Control

In this section, we discuss how the integration of autonomous behaviors in a dual-arm system influences task efficiency, workload, and usability compared to a fully manual control strategy in VR.

First, shared control helps shorten the procedure of spatial positioning and decreases the interaction load, as shown in Figures 4 and 5, resulting in improved task performance. However, certain disadvantages remain, such as unexpected behaviors and inaccurate positioning, which introduced noise and necessitated extra adjustment operations. Observations suggest that unexpected behaviors were primarily caused by the large step length of the AI suggestions, which merged steps 2 and 3 in 4.4.3 . Furthermore, participants occasionally confused the "ghost arm" with the real arm, leading to a cognitive mismatch regarding the task status. Although the system provided status feedback via a highlighting shader, the bottle lip is a relatively small component. Consequently, the current visual effect was not sufficiently prominent. A more precise algorithm combined with multimodal feedback would likely further improve task efficiency.

Second, shared control significantly reduced task load, scoring nine points lower than manual control. It suggests that the cooperation between AI-driven suggestions and participant decision-making conserves more cognitive resources than fully manual operation. Furthermore, an interesting contradiction emerged that most participants felt the shared control was under control despite the autonomous behaviors being unpredictable. In the post-experiment interviews, 16 out of 18 participants reported that autonomous behaviors were under control. Even though 9 of 16 found these behaviors unpredictable, only 1 felt that the AI was not a beneficial partner or tool. Since all participants successfully completed the task, this sense of agency likely stems from control over the outcome rather than the process.

Third, neither manual control nor shared control significantly influenced the usability of the system, as shown in Table 3. This indicates that unpredictable autonomous behaviors potentially canceled out the benefits of reduced objective or cognitive effort. Since all participants were tasked with the same procedure and usability scores remained close, a correlation appears to exist between task difficulty and usability. Moreover, this relationship may function similarly to a learning curve between two test rounds following identical execution steps. As seen in Figure 6, the experience gained in the first turn significantly decreased completion time and the number of buttons pressed in the second turn, even with different control strategies in a counter-balanced order.

In summary, shared control results in higher task efficiency and lower workload but maintains similar usability compared to fully manual control. This suggests that reduced workload does not always equate to a higher subjective perception of the system. Ultimately, the difficulty of the task and the interface design appear to influence usability more significantly than the underlying control strategy.

6.3 Future Works

The ultimate goal of the Wheel2VR project is to migrate the gaze-controlled dual robotic arm framework from VR to AR. The findings of this study provide several pathways for refining semi-autonomous dual-arm coordination and transitioning these technologies from virtual simulations to real-world applications. While VR provided a controlled and cost-effective environment for initial testing, an AR implementation will allow users with tetraplegia to interact with physical objects in their actual living spaces.

Current results show an average NASA-TLX score of 43/100 and an average System Usability Score of 68/100, suggesting moderate workload and usability. While these figures are acceptable for an early-stage prototype, there remains significant room for improvement. Future research will focus on clarifying the factors influencing user preference, task difficulty, and the impact of outcome vs. process on the Sense of Agency.

First, the hypothesis that participants achieve similar task performance when using their preferred input modality needs further validation. If confirmed, user preference should be integrated into a brief probe at the start of each session to identify the most viable modality [Kourtesis et al. 2025]. Once a technically and ergonomically appropriate input is identified, it will be locked in to reduce cognitive overhead and maintain interaction times within clinical constraints.

Second, different levels of task difficulty must be tested within the same control strategy group to further investigate the correlation between difficulty and usability. This evaluation of difficulty will include variables such as the step length of AI suggestions and the complexity of the required solution for a given task.

Third, research into the distinct influences of outcome and process on the Sense of Agency is critical. These results will be synthesized with the task difficulty research to guide the flexibility and interaction design of autonomous arms.

Based on these future research findings, subsequent iterations will focus on improving pose generation and clarifying the system state through multimodal feedback.

7 Limitations

Despite the promising results regarding task efficiency and workload, several limitations of the current study must be addressed.

First, a significant constraint of this study is the spatial arrangement of targets. In the current experimental setup, all interactive targets were programmed to appear exclusively in front of the viewer. In daily life, users may interact with objects located outside their immediate field of view or at varying heights. Therefore, the current results do not fully account for the neck strain or complex gaze-shifting required for 360-degree environmental interaction.

Second, the generalizability of the results is limited by the participant demographic. The evaluation was conducted with able-bodied participants in a controlled laboratory setting. Consequently, the findings cannot be directly transferred to the individuals with tetraplegia. Factors such as localized muscle fatigue, varying degrees of head mobility, and the long-term tolerance of wearing a VR headset may differ significantly for users with specific physiological constraints.

Third, the real-world applicability of the system remains to be validated. While the VR simulation provides a high-fidelity proxy for robotic interaction, current system lacks the physical risks, such as varying lighting conditions or moving obstacles.

8 Conclusion

This paper presented Wheel2VR, a VR-based prototype of a gaze-controlled wheelchair-mounted dual-arm assistive robotic system that combines manual operation with shared autonomy. In our exploratory user study with 18 participants we gained valuable insights. As discussed earlier, the shared control results in higher efficiency and lower workload with similar usability compared to fully manual control. The UI as well as the difficulty of the task appear to influence usability more significantly. Improving pose generation, clarifying system state through stronger feedback, and refining the interaction flow are essential next steps. Since the current evaluation was conducted with able-bodied participants in a controlled lab setting, the findings cannot yet be generalized to people with tetraplegia. Nevertheless, the study provides initial evidence that gaze-based shared control in VR is a viable direction for accessible dual-arm assistive systems and offers a strong foundation for future evaluations with the target user group and, ultimately, real robotic hardware.

References

Anke Fischer-Janzen, Thomas M. Wendt, Daniel G"orlich, and Kristof Van Laerhoven. 2026. Eye-Tracking-Driven Shared Control for Robotic Arms: Wizard of Oz Studies to Assess Design Choices. *ACM Transactions on Human-Robot Interaction* 15, 3 (2026), 1–31. doi:10.1145/3796524

Anke Fischer-Janzen, Thomas M. Wendt, and Kristof Van Laerhoven. 2024. A scoping review of gaze and eye tracking-based control methods for assistive robotic arms. *Frontiers in Robotics and AI* 11 (2024), 1326670. doi:10.3389/frobt.2024.1326670

Vera Fosbrooke, Aline Christen, Barbara Catherine Wortmann, Iris Theodora Maria de Boer, Raphael Rätz, Julian Frederik Rösch, Gabriel Gruener, and Anja M Raab. 2026. Assistive Robotic Arm to Support Activities of Daily Living in Individuals With Tetraplegia: Protocol for a Real-World Convergent Parallel Mixed Methods Feasibility Study. *JMIR Research Protocols* 15, 1 (2026), e78339.

Stefan Fuchs and Anna Belardinelli. 2021. Gaze-Based Intention Estimation for Shared Autonomy in Pick-and-Place Tasks. *Frontiers in Neurorobotics* 15 (2021), 647930. doi:10.3389/fnbot.2021.647930

Felix Ferdinand Goldau. 2024. *Adaptive Degrees of Freedom for Shared Control of Assistive Robots.* Doctoral Thesis. University of Bremen, Bremen, Germany. Submitted: August 21, 2024; Defended: October 2, 2024.

Taylor C. Hansen, Troy N. Tully, V. John Mathews, and David J. Warren. 2024. A Multimodal Assistive-Robotic-Arm Control System to Increase Independence After Tetraplegia. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 32 (2024), 2124–2133. doi:10.1109/TNSRE.2024.3408833

Panagiotis Kourtesis, Evgenia Giatzoglou, Panagiotis Vorias, Katerina Alkisti Gounari, Eleni Orfanidou, and Chrysanthi Nega. 2025. Examination of Eye-Tracking, Head-Gaze, and Controller-Based Ray-Casting in TMT-VR: Performance and Usability Across Adulthood. *Multimodal Technologies and Interaction* 9, 8 (2025). doi:10.3390/mti9080076

Kirill Kronhardt, Max Pascher, and Jens Gerken. 2023. Understanding Shared Control for Assistive Robotic Arms. *arXiv preprint arXiv:2303.01993* (2023). doi:10.48550/arXiv.2303.01993 Presented at VAT-HRI 2023 Workshop on Variable Autonomy for Human-Robot Teaming.

Yuan Yuan Qian and Robert J. Teather. 2017. The eyes don't have it: an empirical comparison of head-based and eye-based selection in virtual reality. In *Proceedings of the 5th Symposium on Spatial User Interaction* (Brighton, United Kingdom) *(SUI '17)*. Association for Computing Machinery, New York, NY, USA, 91–98. doi:10.1145/3131277.3132182

Ke Wan, Chengxi Li, Fo-Sing Lo, and Pai Zheng. 2024. A virtual reality-based immersive teleoperation system for remote human-robot collaborative manufacturing. *Manufacturing Letters* 41 (2024), 43–50. doi:10.1016/j.mfglet.2024.09.008

Weiwei Wan and Kensuke Harada. 2016. Developing and Comparing Single-Arm and Dual-Arm Regrasp. *IEEE Robotics and Automation Letters* 1, 1 (2016), 243–250. doi:10.1109/LRA.2016.2517147

Shaochen Wang, Wei Zhang, Zhangli Zhou, Jiaxi Cao, Ziyang Chen, Kang Chen, Bin Li, and Zhen Kan. 2023. What You See is What You Grasp: User-Friendly Grasping Guided by Near-eye-tracking. *arXiv preprint arXiv:2209.06122* (2023). doi:10.48550/arXiv.2209.06122

Mingzhao Zhou and Nadine Aburumman. 2024. Grasping Object: Challenges and Innovations in Robotics and Virtual Reality. *arXiv preprint arXiv:2411.06244* (2024). doi:10.48550/arXiv.2411.06244

Mark Zolotas and Yiannis Demiris. 2020. Transparent Intent for Explainable Shared Control in Assistive Robotics. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence Doctoral Consortium.* 5184–5185. doi:10.24963/ijcai.2020/732