# Designing Remote Assistance Interfaces for Automated Vehicles

A Planning-Oriented Prototype with Real-World Driving Stimuli

Alexandra Nick
Karlsruhe Institute of Technology (KIT)
Institute of Human and Industrial Engineering (ifab)
Karlsruhe, Germany
alexandra.nick@kit.edu

Leon Lamesic*
Karlsruhe Institute of Technology (KIT)
Institute of Human and Industrial Engineering (ifab)
Karlsruhe, Germany
uclzt@student.kit.edu

Barbara Deml*
Karlsruhe Institute of Technology (KIT)
Institute of Human and Industrial Engineering (ifab)
Karlsruhe, Germany
barbara.deml@kit.edu

## Abstract

Automated vehicles may encounter situations at system limits that require external human support without continuous remote driving. We present a lab-based remote assistance setup for investigating planning-oriented interaction methods. The setup combines real-world driving video stimuli, contextual dashboard information, a pen-based planning display, and instrumentation for logging interaction and attention data in a multi-screen operator workplace. It supports two assistance concepts: collaborative planning, in which operators select system-generated maneuver suggestions, and trajectory guidance, in which operators draw a desired path for the vehicle. We conducted a formative within-subjects evaluation with 35 participants to assess the setup and explore perceived usability and user experience of two implemented interaction conditions. No statistically significant differences were found between the two implemented conditions in perceived usability or user experience. The findings suggest that both concepts warrant further investigation and motivate future setup iterations that combine interaction methods according to scenario demands and operator preferences.

## CCS Concepts

• **Social and professional topics**; • **Human-centered computing** → **Graphical user interfaces**; **User studies**; *User interface programming*;

## Keywords

teleoperation, remote assistance, automated vehicles, usability, human–automation interaction

**ACM Reference Format:**
Alexandra Nick, Leon Lamesic, and Barbara Deml. 2026. Designing Remote Assistance Interfaces for Automated Vehicles: A Planning-Oriented Prototype with Real-World Driving Stimuli. In *Proceedings of Mensch und Computer 2026 (MuC '26), August 30–September 02, 2026, Duisburg, Germany.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3820253.3831374

## 1 Introduction

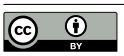
Highly automated vehicles (AVs) are designed to operate without continuous human supervision within their operational design domain. However, in open road environments, they may still encounter situations that are difficult to resolve autonomously, including ambiguous traffic situations, perception errors, planning conflicts, roadworks, occlusions, and interactions with vulnerable road users [2, 8]. Such situations underline the continuing need for event-based human support in future automated mobility systems [10].

Teleoperation has been proposed as a fallback mechanism when AVs reach their operational limits [2, 9]. In contrast to remote driving, remote assistance (RA) provides high-level, event-based support without continuous manual vehicle control: the remote operator supplies information, advice, or planning input, while the AV remains responsible for validating and executing the maneuver [1, 12]. Previous work has provided teleoperation taxonomies [1, 12], scenario classifications [8, 17], and prototype RA interfaces [11, 14]. However, empirical comparisons of RA interaction methods remain scarce [2], and many prototype evaluations rely on simulated or abstracted traffic environments [22]. Therefore, study setups are needed that combine technically plausible RA interaction concepts with realistic traffic stimuli and human-centered evaluation.

## 2 Background and Scope

### 2.1 Remote Assistance and Interaction Concepts

Instead of replicating the in-vehicle driving task remotely, RA shifts the interaction toward supervisory and planning-oriented human–automation collaboration: the operator assesses the assistance request and provides input that the AV can interpret and execute. RA is therefore not a single interaction paradigm. Prior work has proposed waypoint guidance, trajectory guidance, perception modification, and collaborative planning [2, 9]. These concepts differ in the allocation of responsibility, information requirements, and the human planning effort required [10, 21]. Waypoint and trajectory-based approaches specify spatial guidance, whereas perception modification supports environmental interpretation. Collaborative planning reduces planning effort by presenting system-generated suggestions that can be assessed, selected, or modified [2, 9]. Consequently, RA usability depends on the interaction concept, scenario characteristics, and available interface information [10]. Because these concepts differ in operator involvement and the allocation of responsibility, their evaluation may also require task-specific constructs such as perceived safety, decision confidence, perceived responsibility, and situation awareness [2]. Prior work has investigated RA usability and workload across implementation approaches



This work is licensed under a Creative Commons Attribution 4.0 International License.
*MuC '26, Duisburg, Germany*
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2611-8/26/08
https://doi.org/10.1145/3820253.3831374

[11, 14, 19], but the suitability of specific RA methods remains closely tied to task and scenario demands.

### 2.2 Disengagement Scenarios

Disengagements may result from perception failures, path-planning problems, sensor malfunctions, or unexpected events [2, 10]. Scenario-oriented work shows that not all disengagements are suitable for RA [8] or teleoperation in general [2]. RA is particularly relevant when the AV cannot continue autonomously, but the situation can still be resolved through high-level human support, for example, in unclear road layouts, roadworks, obstacles, occlusions, poor visibility, or ambiguous interactions with other road users [7, 8, 18]. Human-centered approaches further emphasize task demands such as predictability, available decision time, degrees of freedom, situational accessibility, and environmental complexity [10]. The present study focuses on event-based assistance for a single AV across two scenario clusters: unclear perception and blocked road situations. Real-world driving videos were used to preserve visual complexity, while the assistance task remained controlled through an abstract planning interface rather than direct interaction with the recorded vehicle [22].

### 2.3 Research Gap and Contributions

Prior work has provided teleoperation taxonomies [1, 12], scenario classifications [2, 7, 13, 18], and RA workplace prototypes [11, 14, 22]. However, less is known about how to implement planning-oriented RA methods in a controlled laboratory setting while preserving the visual complexity of real-world traffic stimuli. This short paper addresses this gap by presenting a RA workstation for studying planning-oriented interaction methods in highly automated driving. The setup integrates real-world driving videos, contextual dashboard information, and standardized pen-based planning interaction in a multi-screen operator workplace. It is used to explore, in a formative way, two implemented interaction conditions: trajectory guidance and collaborative planning. Both conditions were tested in scenarios drawn from the same two problem clusters, unclear perception and blocked road situations, but with scenario-specific realizations for each interaction method. The primary contribution of this paper is the design and formative evaluation of a prototypical RA setup for planning-oriented remote assistance. The accompanying within-subject evaluation reports initial usability and user experience results for two implemented interaction conditions and provides guidance for future scenario-based comparisons of RA interaction methods.

## 3 Setup

### 3.1 Remote Assistance Setup

The RA workplace consisted of three road-view displays (24-inch LG displays arranged in a curved layout) and one interaction display (a 15.6-inch Asus ProArt pen-based touch display). A stationary SmartEye eye-tracking system with four cameras and three infrared flashes was mounted around the displays to enable later analyses of visual attention. All screens were fixed in their position to allow the world coordinate system to be calibrated in SmartEye. The road-view displays presented synchronized real-world driving videos from different viewing angles, while the interaction display combined contextual dashboard information with a pen-based planning interface. This separation was intended to approximate the information structure of an RA workplace: the video displays supported situation assessment, whereas the interaction display supported the formulation or confirmation of a planning decision.

The setup supported two planning-oriented RA methods: collaborative planning and trajectory guidance. In the collaborative planning condition, participants selected one of several predefined maneuver suggestions provided by the system. The operator's task was therefore to assess the traffic situation and select the most appropriate system-generated decision. In the trajectory guidance condition, participants manually drew the trajectory the vehicle should follow to resolve the assistance request. Here, the operator specified the intended path more directly, while the vehicle was assumed to validate and execute the submitted trajectory.

In both conditions, participants could also refuse to pursue the maneuver if they considered the situation unsafe to solve with RA. In this case, they could abort the maneuver. The intended system response was for the vehicle to stop at a safe position on the roadside, terminate the current trip, and wait for external support. These design choices were intended to reduce avoidable interaction variance and to separate situation assessment from planning input. Figure 1 (a) shows the collaborative planning condition, while Figure 1 (b) shows the trajectory guidance condition. The driving examples shown in the figures were anonymized for publication. Detailed screenshots of the interaction display are provided in Figure 2.

### 3.2 Experimental Control Software

The study was controlled by a custom Windows-based application that coordinated the multi-screen RA setup. At the start of each RA session and each trial block, the experimenter entered the participant's ID, the randomized order of the trials, and the participant's dominant hand (right/left). The dashboard adjusted accordingly and positioned the interactive unit to the right or left of the dashboard to enable more intuitive pen interactions (See Figure 5 (a)).

For each trial, the software presented three synchronized video streams (front view with integrated rear mirror and side mirrors) on separate road-view displays, and loaded the corresponding dashboard and planning interface on the interaction display. After the video sequence reached the assistance point, the planning interface was unlocked, and participants could provide input via the interaction space. The software logged trial-level interaction data, including the time until the first input, the processing time until submission, the number of undo actions, the submitted decision type, and the perceived scenario difficulty. In addition, the final interaction input was saved as an image, allowing later inspection of the submitted trajectory or planning decision. Between the trials, there was always a break of a random length between 10 and 15 seconds. The next incoming trial was signaled by a beep and followed by the ready screen for the next scenario (Mattias "MATRIXXX" Lahoud, 2020). The procedure of a trial is illustrated in Figure 5 from b) to d). The software, therefore, served as an orchestration and logging layer for the experimental setup. The code is provided as a zip file in the supplementary material in the OSF project.



**(a)** Collaborative planning (Scenario BD)



**(b)** Trajectory-based interaction (Scenario UH)

**Figure 1: Example trials in the RA setup. The upper displays show synchronized real-world driving videos from multiple viewpoints, while the lower touch display presents contextual information for the dashboard and planning interface.**

## 4 Methods

An a priori power analysis conducted with G*Power [4] indicated a minimum sample size of 34 participants (two-tailed t-test with dependent-means, $d = .5$, $\alpha = .05$, $1 - \beta = .80$). In the study, 37 participants took part. Due to missing data in the paper-based questionnaires, only 35 participants are included in the analysis presented. The participants identified mainly as male (m = 24, f = 11) with a mean age of 25.1 (SD = 6.3). Driving experience varied across the sample with $n = 13$ participants (37.1%) reported driving several times per month, $n = 9$ (25.7%) less than once per month, $n = 7$ (20.0%) several times per week, and $n = 4$ (11.4%) daily. Two participants (5.7%) reported not having a driver's license.

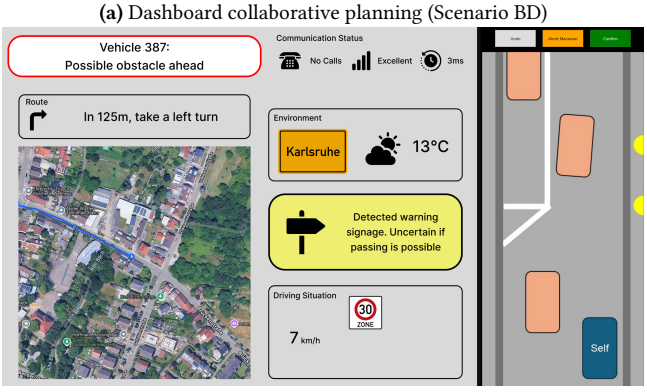
Given the formative focus on setup validation rather than professional operator performance, participants were recruited as lay users with prior touch and stylus experience (80.0% in private contexts, 34.3% in employment contexts), which limits generalizability to trained RA operators. Additional sample information is included in Appendix C.1. The study followed a within-subjects design with one factor with two levels (interaction methods: trajectory-based and collaborative planning). The order of the interaction methods was randomized.

For each interaction method, participants first completed two familiarization trials, followed by eight assistance scenarios based on real-world driving videos (see Appendix Table 1). Scenario sets were not identical across interaction methods but were balanced at the level of problem clusters, with four unclear perception and four blocked road scenarios per method. Thus, the evaluation allows an exploratory comparison of implemented conditions under cluster-balanced scenario sets, but cannot isolate interaction-method effects from scenario variance.

Participants rated the perceived difficulty of each scenario on a five-point scale ranging from 1 (*very easy*) to 5 (*very difficult*). Difficulty ratings are reported in Appendix C.3 to increase transparency regarding perceived scenario demands as one facet of comparability across the cluster-balanced scenario sets. Each scenario had at least one predefined feasible solution or abort option, depending on whether the situation could be resolved safely through RA. The submitted interactions were stored as drawings to enable subsequent inspection of selected maneuvers, trajectory patterns, and potentially ambiguous inputs.




**(a)** Dashboard collaborative planning (Scenario BD)




**(b)** Dashboard trajectory-based interaction (scenario UH)

**Figure 2: Enlarged screenshots of the dashboard and interactive planning space for the two example scenarios shown in Figure 1. Panel (a) shows the collaborative planning interface for scenario BD, and Panel (b) shows the trajectory guidance interface for scenario UH.**

The complete study assessed subjective workload, user experience, and overall system usability. Subjective workload was measured with the NASA-TLX, including pairwise weighting of the six workload dimensions and ratings on an adapted scale from 1 to 10 [6]. User experience was assessed using the 26-item User Experience Questionnaire (UEQ) [15], and usability with a non-standard seven-point version of the System Usability Scale (SUS) [3]; SUS scores are therefore reported descriptively rather than interpreted against standard benchmarks. Participants could additionally provide open feedback on the setup and the interaction methods.

The present paper focuses on subjective usability and user experience because its primary aim is to present the RA setup and provide an initial formative evaluation of the implemented interaction conditions. Additional interaction, attention, workload, and submitted input data were recorded but are not reported here, as they require separate analyses beyond the scope of this short paper. The complete procedure is outlined in Figure 4 and further details on the complete study design are provided in the supplementary material within the OSF project.

## 5 Results

User experience was compared for collaborative planning and trajectory guidance descriptively (see Figure 3(a)) and by using paired-samples tests. Internal consistencies of the UEQ subscales are reported in Table 2. Cronbach's $\alpha$ ranged from .54 to .79, with lower values for perspicuity and dependability; results for these subscales should therefore be interpreted cautiously. Based on the Shapiro–Wilk test results for each subscale, paired-samples $t$-tests or Wilcoxon signed-rank tests were computed. No statistically significant differences between the two implemented conditions were found for any UEQ subscale after Holm correction. The paired-samples $t$-tests were non-significant for attractiveness, $t(34) = 1.25$, $p = .219$, $p_{\text{Holm}} = 1.00$; efficiency, $t(34) = 1.46$, $p = .154$, $p_{\text{Holm}} = .922$; perspicuity, $t(34) = -0.20$, $p = .847$, $p_{\text{Holm}} = 1.00$; and dependability, $t(34) = 1.05$, $p = .303$, $p_{\text{Holm}} = 1.00$. Similarly, Wilcoxon signed-rank tests showed no significant differences for stimulation, $V = 169.00$, $p = .870$, $p_{\text{Holm}} = 1.00$, or novelty, $V = 147.00$, $p = .795$, $p_{\text{Holm}} = 1.00$. Given the lower internal consistency of the perspicuity and dependability subscales, these subscales are interpreted with caution.

System usability was analyzed using a modified SUS total score because the questionnaire used a seven-point response scale. After recoding the items to a 0–6 range, item scores were summed and multiplied by 2.5, yielding a possible score range of 0 to 150, with higher values indicating greater perceived usability. Therefore, the scores were not interpreted against conventional SUS benchmarks. Descriptively, SUS scores were very similar for collaborative planning ($M = 114.50$, $SD = 19.51$) and trajectory guidance ($M = 115.14$, $SD = 19.45$; see Figure 3(b)). A Shapiro–Wilk test did not indicate a deviation from normality for the paired difference scores, $W = .983$, $p = .838$. Therefore, a paired-samples $t$-test was conducted. The test showed no statistically significant difference, $t(34) = -0.20$, $p = .842$, $d_z = -0.03$.

## 6 Discussion and Future Work

The results showed no statistically significant differences between the two implemented conditions in perceived usability or user experience. Both collaborative planning and trajectory guidance received similar subjective evaluations in cluster-balanced, scenario-specific assistance situations. This finding is relevant because the two concepts represent different levels of operator involvement: collaborative planning restricts interaction to selecting predefined maneuver options, whereas trajectory guidance allows operators to specify a spatial path more directly [2].

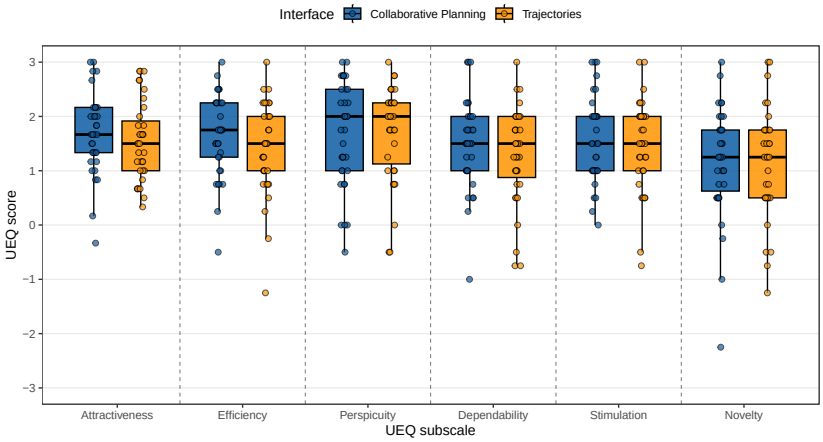
The absence of statistically significant differences should not be interpreted as evidence of equivalence or as a definitive comparison of the two RA concepts. Rather, it indicates that the implemented conditions were not clearly differentiated by the global usability and user experience measures used in this study. This may reflect that both concepts provided high-level planning support rather than continuous remote driving. It may also indicate that global UX and usability scales are not sufficiently sensitive to capture RA-specific differences, such as perceived responsibility, decision confidence, perceived safety, risk perception, situation awareness, or trust in the vehicle's execution of the submitted input. Because the scenarios were not identical across conditions, the findings should therefore be interpreted as formative evidence for the usability of the implemented prototypes.

Future research should investigate flexible RA interfaces that combine multiple interaction methods and systematically vary scenario characteristics. Predefined maneuver options may be useful for routine cases with clearly distinguishable alternatives, whereas trajectory input may be more appropriate when spatial constraints or detailed path information need to be communicated. Intermediate functions could complement both approaches, for example, by allowing operators to confirm free space [5] or by enhancing transparency through signaling uncertainty or marking relevant objects in the scene [16, 20]. Subsequent studies should use scenarios that are solvable using all investigated input methods, systematically vary the complexity of the situation, and combine global usability measures with RA-specific indicators, objective interaction data, and qualitative analyses of submitted decisions or trajectories.

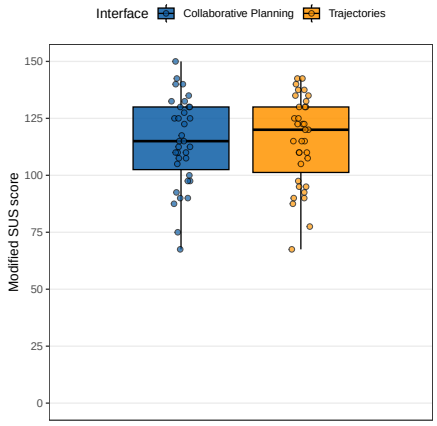
## 7 Limitations and Conclusion

Several limitations should be acknowledged. The study was conducted in a laboratory setting and used recorded real-world driving videos rather than dynamically coupled vehicle behavior. Thus, the setup increased visual realism compared with abstract simulations, but did not reproduce key ecological demands of operational RA, such as real-time responsibility, time pressure, safety consequences, or feedback from an executing vehicle. In addition, the scenario sets differed between interaction methods, although they were selected from the same two problem clusters. This limits internal validity because observed similarities or differences cannot be attributed solely to the interaction method, but rather to scenario variance.

Only subjective usability and user experience results are reported here. Some UEQ subscales showed low internal consistency, particularly perspicuity and dependability, which calls for cautious interpretation of these subscale results. Furthermore, the formative lay-user sample limits generalizability to trained RA operators.




**(a) UEQ subscale scores.** Boxplots show the distributions of participant-level scores for each UEQ subscale, ranging from −3 to +3, with higher values indicating more positive ratings. Individual data points are overlaid, and dashed vertical lines separate the subscales.




**(b) Modified SUS scores.** Boxplots show the distribution of participant-level scores for both interfaces on a possible scale from 0 to 150, with higher values indicating higher perceived usability. Individual data points are overlaid.

**Figure 3: User experience and system usability ratings for collaborative planning and trajectory guidance. Panel (a) shows UEQ subscale scores, and Panel (b) shows modified SUS scores. Boxplots display participant-level score distributions, with individual data points overlaid.**

Overall, this paper contributes a lab-based RA setup for studying planning-oriented interaction with real-world driving stimuli. The formative evaluation showed similar subjective ratings for collaborative planning and trajectory guidance across cluster-balanced assistance scenarios. Future work should test the suitability of scenario-specific methods under more controlled conditions and operational pressure, while including RA-specific subjective and objective measures.

## References


[1] Arbeitsgruppe Forschungsbedarf Teleoperation. 2025. *Abschlussbericht der Arbeitsgruppe „Forschungsbedarf Teleoperation“.* Berichte der Bundesanstalt für Straßen- und Verkehrswesen, Fahrzeugtechnik F 166a. Bundesanstalt für Straßen- und Verkehrswesen, Bergisch Gladbach. doi:10.60850/bericht-f166a Koordination: Tom Michael Gasser and Alexander Frey.

[2] David Brecht, Nils Gehrke, Tobias Kerbl, Niklas Krauss, Domagoj Majstorovic, Florian Pfab, Maria-Magdalena Wolf, and Frank Diermeyer. 2024. Evaluation of Teleoperation Concepts to Solve Automated Vehicle Disengagements. *IEEE Open Journal of Intelligent Transportation Systems* 5 (2024), 629–641. doi:10.1109/OJITS.2024.3468021

[3] John Brooke. 1996. SUS: A 'Quick and Dirty' Usability Scale. In *Usability Evaluation in Industry*, Patrick W. Jordan, Bruce Thomas, Ian Lyall McClelland, and Bernard Weerdmeester (Eds.). CRC Press. doi:10.1201/9781498710411-35

[4] Franz Faul, Edgar Erdfelder, Axel Buchner, and Albert-Georg Lang. 2009. Statistical Power Analyses Using G*Power 3.1: Tests for Correlation and Regression Analyses. *Behavior Research Methods* 41, 4 (2009), 1149–1160. doi:10.3758/BRM.41.4.1149

[5] Gaetano Graf, Yomna Abdelrahman, Hao Xu, Yasmeen Abdrabou, Dmitrij Schitz, Heinrich Hußmann, and Florian Alt. 2020. The Predictive Corridor: A Virtual Augmented Driving Assistance System for Teleoperated Autonomous Vehicles. In *ICAT-EGVE 2020 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*, Ferran Argelaguet, Ryan McMahan, and Maki Sugimoto (Eds.). The Eurographics Association. doi:10.2312/egve.20201260

[6] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, Peter A. Hancock and Najmedin Meshkati (Eds.). Advances in Psychology, Vol. 52. North-Holland, 139–183. doi:10.1016/S0166-4115(08)62386-9

[7] Carmen Kettwich, Andreas Schrank, Hüseyin Avsar, and Michael Oehl. 2021. What if the Automation Fails? A Classification of Scenarios in Teleoperated Driving. In *Proceedings of the 13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '21).* Association for Computing Machinery, New York, NY, USA, 92–96. doi:10.1145/3473682.3480271

[8] Carmen Kettwich, Andreas Schrank, Hüseyin Avsar, and Michael Oehl. 2022. A Helping Human Hand: Relevant Scenarios for the Remote Operation of Highly Automated Vehicles in Public Transport. *Applied Sciences* 12, 9 (2022), 4350. doi:10.3390/app12094350

[9] Domagoj Majstorovic, Simon Hoffmann, Florian Pfab, Andreas Schimpe, Maria-Magdalena Wolf, and Frank Diermeyer. 2022. Survey on Teleoperation Concepts for Automated Vehicles. In *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC).* 1290–1296. arXiv:2208.08876 doi:10.1109/SMC53654.2022.9945267

[10] Miriam Alexandra Nick, Nicole Damm, Julian Elias Reiser, Martin Baumann, and Barbara Deml. 2025. Understanding teleoperation: A human-centered framework for workplace design. *Zeitschrift für Arbeitswissenschaft* 79, 4 (2025), 581–593. doi:10.1007/s41449-025-00492-3

[11] Miriam Alexandra Nick, Marco Käppler, and Barbara Deml. 2025. Interaction Methods for Teleoperation: A Usability Study on Remote Assistance in Highly Automated Vehicles. In *IEEE International Automated Vehicle Validation Conference (IAVVC), Baden-Baden, Germany, 30 September 2025 - 02 October 2025.* Institute of Electrical and Electronics Engineers (IEEE). doi:10.1109/IAVVC61942.2025.11219475

[12] On-Road Automated Driving (ORAD) Committee. 2021. *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles.* doi:10.4271/J3016_202104

[13] Alexander Richter, Teo Puig Walz, Mayur Dhanani, Ivo Häring, Georg Vogelbacher, Fabian Höflinger, Jörg Finger, and Alexander Stolz. 2023. Components and Their Failure Rates in Autonomous Driving. In *Proceedings of the 33rd European Safety and Reliability Conference (ESREL 2023)*, Mário P. Brito, Terje Aven, Piero Baraldi, Marko Čepin, and Enrico Zio (Eds.). Research Publishing, Singapore, 223–240. doi:10.3850/978-981-18-8071-1_P398-cd

[14] Andreas Schrank, Fabian Walocha, Stefan. Brandenburg, and Michael Oehl. 2024. Human-centered design and evaluation of a workplace for the remote assistance of highly automated vehicles. *Cognition, Technology & Work* 26 (2024), 183–206. doi:10.1007/s10111-024-00753-x

[15] Martin Schrepp, Jörg Thomaschewski, and Andreas Hinderks. 2017. Construction of a Benchmark for the User Experience Questionnaire (UEQ). *International*

*Journal of Interactive Multimedia and Artificial Intelligence* 4, 4 (06/2017 2017), 40–44. doi:10.9781/ijimai.2017.445

[16] Monica Tatasciore, Luke Strickland, and Shayne Loft. 2024. Transparency improves the accuracy of automation use, but automation confidence information does not. *Cognitive Research: Principles and Implications* 9, 1 (Oct. 2024), 67. doi:10.1186/s41235-024-00599-x

[17] Felix Tener and Joel Lanir. 2022. Driving from a Distance: Challenges and Guidelines for Autonomous Vehicle Teleoperation Interfaces. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 250, 13 pages. doi:10.1145/3491102.3501827

[18] Felix Tener and Joel Lanir. 2024. Investigating Intervention Road Scenarios for Teleoperation of Autonomous Vehicles. *Multimedia Tools and Applications* 83, 21 (2024), 61103–61119. doi:10.1007/s11042-023-17851-z

[19] Felix Tener and Joel Lanir. 2025. Guiding, Not Driving: Design and Evaluation of a Command-Based User Interface for Teleoperation of Autonomous Vehicles. *International Journal of Human–Computer Interaction* (2025). doi:10.1080/10447318.2025.2594751 Advance online publication.

[20] Koen van de Merwe, Steven Mallam, Salman Nazir, and Øystein Engelhardtsen. 2024. The Influence of Agent Transparency and Complexity on Situation Awareness, Mental Workload, and Task Performance. *Journal of Cognitive Engineering and Decision Making* 18, 2 (June 2024), 156–184. doi:10.1177/15553434241240553

[21] Maria-Magdalena Wolf, Niklas Krauss, Arwed Schmidt, and Frank Diermeyer. 2025. Control Center Framework for Teleoperation Support of Automated Vehicles on Public Roads. In *2025 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2538–2545. doi:10.1109/IV64158.2025.11097634

[22] Maria-Magdalena Wolf, Henrik Schmidt, Michael Christl, Jana Fank, and Frank Diermeyer. 2025. A User-Centered Teleoperation GUI for Automated Vehicles: Identifying and Evaluating Information Requirements for Remote Driving and Assistance. *Multimodal Technologies and Interaction* 9, 8 (2025), 78. doi:10.3390/mti9080078

# A Disengagement Scenarios

In addition to the scenario overview in Table 1, the OSF project provides a visual scenario table with anonymized screenshots of the final video frames, the corresponding dashboards, and the interaction spaces.

# B Procedure

The complete experimental procedure is illustrated as a flowchart in Figure 4. Figure 5 shows screenshots from the trial phases: (a) the setup screen where the participant ID and trial order were entered based on the randomization table, (b) the ready screen for accepting the incoming vehicle request, accompanied by an alert sound, (c) the assistance task with video feed, dashboard, and interaction space adapted to the participant's dominant hand, and (d) the post-submission difficulty rating.

**Figure 4: Overview of the study procedure, including preparation, two experimental blocks, post-block evaluations, and final feedback.**

# C Additional Analysis Information

## C.1 Additional Sample Information

More detailed description of the sample: The highest educational attainment was a high school graduate qualification for $n = 15$ participants (42.9%), a bachelor's degree for $n = 15$ participants (42.9%), some college education without a degree for $n = 2$ participants (5.7%), and a master's degree for $n = 3$ participants (8.6%). Most participants were students ($n = 31$, 88.6%), while $n = 3$ participants (8.6%) reported full-time employment and $n = 1$ participant (2.9%) reported part-time employment. Regarding field of work or study, most participants were affiliated with manufacturing and engineering ($n = 27$, 77.1%), followed by education and research ($n = 4$, 11.4%), other fields ($n = 2$, 5.7%), finance, insurance and real estate ($n = 1$, 2.9%), and IT and telecommunications ($n = 1$, 2.9%). Across participants, ATI scores indicated a mean of $M = 4.23$ ($SD = 1.04$), with observed values ranging from 1.89 to 6.00 (Cronbach's $\alpha = .90$).

## C.2 Scale Reliability and Descriptives

Table 2 reports the internal consistencies and descriptive values of the UEQ subscales for each interaction method. Most subscales showed acceptable internal consistency, whereas perspicuity and dependability showed lower Cronbach's $\alpha$ values in at least one condition. Results for these subscales should therefore be interpreted cautiously, particularly because the present evaluation focuses on formative usability insights rather than scale validation.

## C.3 Perceived Difficulty Across Scenarios

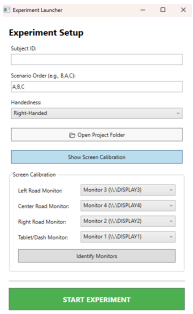
As an additional descriptor of scenario variance, Table 3 reports perceived difficulty ratings for each experimental scenario, and Table 4 summarizes these ratings by interaction method and problem cluster. In addition, difficulty ratings were averaged across the eight experimental scenarios for each participant and interaction method. These participant-level averages were similar for collaborative planning ($M = 2.19$, $SD = 0.45$) and trajectory guidance

**Table 1: Overview of the assistance scenarios used in the study. Scenario codes indicate the problem cluster: U = unclear perception, and B = blocked road. Scenarios marked with (F) were used as familiarization trials. T denotes trajectory guidance, and CP denotes collaborative planning.**

| Interaction method | Short code | Description |
|---|---|---|
| T | UH (F) | AV detects a cyclist ahead, but the system is uncertain whether the cyclist can be passed safely within the available road space. |
| T | BF (F) | AV approaches a possible blockage (car) ahead. The system is uncertain whether the blockage will remain and whether it can be passed safely. |
| T | UA | AV reports an unclear road layout and is uncertain whether continuation is possible (yellow road marks, many traffic signs). |
| T | UB | AV reports an unclear road layout and is uncertain whether following the planned route is possible alongside construction barriers. |
| T | UI | AV reports an unclear road layout and is uncertain whether the planned route can be followed (leaves, lack of road marking). |
| T | UD | AV reports a lack of road markings and is uncertain whether the planned route is possible (offset roundabout, residential area). |
| T | BB | AV reports a possible obstacle ahead because a parked car has been detected in the street and the system is uncertain whether passing is possible (fog, yellow road marks). |
| T | BE | AV reports a cyclist ahead and the system is uncertain whether passing is possible (city street, tram lines left). |
| T | BH | AV reports a cyclist ahead and the system is uncertain whether passing is possible (residential area, no road marking). |
| T | BI | The AV reports an obstacle (parked car) ahead and is uncertain whether the blockage will persist or whether passing is necessary (busy two-lane road). |
| CP | UC (F) | AV reports impaired sensory perception and is uncertain whether the road ahead is open. |
| CP | BA (F) | AV detects a delaying vehicle ahead and is uncertain whether an overtaking maneuver is possible. |
| CP | UE | AV reports that people have been detected on or near the road and is uncertain whether a detected person may enter the roadway. |
| CP | UF | AV reports a vehicle possibly crossing the road (driveway) and is uncertain whether continuation is safe. |
| CP | UG | AV reports an unclear road layout because the observed traffic flow does not appear to match the expected road layout (roundabout, missing markings), creating uncertainty about whether the planned route can be followed. |
| CP | UJ | AV reports an unclear road layout (shifted roundabout) because no road markings are detected and is uncertain whether the planned turning maneuver can be followed (residential area). |
| CP | BC | AV reports a possible obstacle ahead because a car stopping in the street has been detected and the system is uncertain whether passing is possible. |
| CP | BD | The AV reports an obstacle ahead because a car is stopping in the street (right turn of several vehicles ahead) has been detected, and the system is uncertain whether passing is possible (changed road layout ahead with tram lines to the left). |
| CP | BG | AV reports an obstacle ahead and is uncertain whether a detected blockage will persist or whether passing is possible (intersection). |
| CP | BJ | AV reports an obstacle ahead and is uncertain whether a detected blockage will persist or whether passing is necessary (no road markings). |

**Table 2: UEQ subscale reliability and descriptives for each interaction method.**

| Subscales | Collaborative Planning | | | Trajectories | | |
|---|---|---|---|---|---|---|
| | Cronbach's $\alpha$ | Mean | SD | Cronbach's $\alpha$ | Mean | SD |
| Attractiveness (6 items) | .77 | 1.69 | 0.74 | .79 | 1.54 | 0.72 |
| Perspicuity (4 items) | .54 | 1.65 | 0.76 | .60 | 1.41 | 0.85 |
| Efficiency (4 items) | .75 | 1.68 | 0.95 | .74 | 1.71 | 0.87 |
| Dependability (4 items) | .58 | 1.49 | 0.81 | .67 | 1.29 | 0.89 |
| Stimulation (4 items) | .75 | 1.55 | 0.79 | .73 | 1.48 | 0.84 |
| Novelty (4 items) | .75 | 1.16 | 1.02 | .71 | 1.14 | 1.01 |



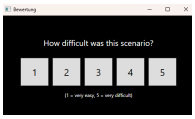

**(a)** Start of setup



**(b)** Ready screen



**(c)** Assistance trial with collaborative planning




**(d)** Difficulty rating

**Figure 5: Sequence of a single assistance trial. Participants first started the setup, then confirmed readiness, completed the assistance task, and finally rated the perceived difficulty of the scenario.**

($M = 2.12$, $SD = 0.46$). Descriptively, difficulty varied across problem clusters: trajectory guidance was rated slightly more difficult for unclear perception scenarios, whereas collaborative planning was rated slightly more difficult for blocked road scenarios. These ratings provide additional transparency regarding perceived scenario demands within the cluster-balanced design.

**Table 3: Perceived difficulty ratings for each experimental assistance scenario by interaction method ( collaborative planning = CP, trajectories = T) and situation cluster (cluster: U = unclear perception, and B = blocked road). Difficulty was rated after each scenario on a five-point scale ranging from 1 (*very easy*) to 5 (*very difficult*).**

| Interface | Cluster | Scenario | n | Mean | SD | Median | Min | Max |
|---|---|---|---|---|---|---|---|---|
| CP | U | UE | 35 | 1.57 | 0.74 | 1 | 1 | 4 |
| CP | U | UF | 34 | 3.00 | 1.02 | 3 | 1 | 5 |
| CP | U | UG | 35 | 2.14 | 1.09 | 2 | 1 | 4 |
| CP | U | UJ | 35 | 2.09 | 1.12 | 2 | 1 | 5 |
| CP | B | BC | 35 | 2.51 | 0.85 | 2 | 1 | 4 |
| CP | B | BD | 35 | 2.31 | 1.02 | 2 | 1 | 5 |
| CP | B | BG | 35 | 2.31 | 0.90 | 2 | 1 | 4 |
| CP | B | BJ | 35 | 1.60 | 0.85 | 1 | 1 | 4 |
| T | U | UA | 35 | 2.14 | 1.06 | 2 | 1 | 5 |
| T | U | UB | 35 | 2.14 | 1.03 | 2 | 1 | 4 |
| T | U | UD | 35 | 2.77 | 1.09 | 3 | 1 | 5 |
| T | U | UI | 35 | 2.69 | 1.16 | 3 | 1 | 5 |
| T | B | BB | 35 | 1.69 | 0.96 | 1 | 1 | 5 |
| T | B | BE | 35 | 2.03 | 0.86 | 2 | 1 | 4 |
| T | B | BI | 35 | 1.80 | 0.76 | 2 | 1 | 4 |
| T | B | BH | 35 | 1.71 | 0.79 | 2 | 1 | 3 |

**Table 4: Perceived difficulty ratings summarized by interaction method ( collaborative planning = CP, trajectories = T) and situation cluster (cluster: U = unclear perception, and B = blocked road). The table provides descriptive information on perceived scenario demands within the cluster-balanced design.**

| Interface | Cluster | n | Mean | SD | Median | Min | Max |
|---|---|---|---|---|---|---|---|
| CP | U | 35 | 2.19 | 0.64 | 2.00 | 1.25 | 3.75 |
| CP | B | 35 | 2.19 | 0.50 | 2.00 | 1.50 | 3.50 |
| T | U | 35 | 2.44 | 0.64 | 2.50 | 1.25 | 3.75 |
| T | B | 35 | 1.81 | 0.46 | 1.75 | 1.00 | 3.25 |