



Human-centered design and evaluation of a workplace for the remote assistance of highly automated vehicles

Andreas Schrank¹ · Fabian Walocha¹ · Stefan Brandenburg² · Michael Oehl¹

Received: 14 September 2023 / Accepted: 8 January 2024
© The Author(s) 2024

Abstract

Remotely operating vehicles utilize the benefits of vehicle automation when fully automated driving is not yet possible. A human operator ensures safety and availability from afar and supports the vehicle automation when its capabilities are exceeded. The remote operator, conceptualized as remote assistant, fulfills the legal requirements in Germany as a Technical Supervisor to operate highly automated vehicles at Society of Automotive Engineers 4. To integrate the remote operator into the automated driving system, a novel user-centered human–machine interface (HMI) for a remote assistant’s workplace was developed and initially evaluated. The insights gained in this process were incorporated into the design of a workplace prototype for remote assistance. This prototype was tested in the study reported here by 34 participants meeting the professional background criteria for the role of Technical Supervisor according to the German law. Typical scenarios that may occur in highly automated driving and require remote assistance were created in a simulation environment. Even under elevated cognitive load induced by simultaneously engaging in a secondary task, participants were able to obtain sufficient situation awareness and quickly resolve the scenarios. The HMI also yielded favorable usability and acceptance ratings. The results of this study inform the iterative workplace development and further research on the remote assistance of highly automated vehicles.

Keywords Remote operation · Remote assistance · Highly automated driving · Human–machine interaction · Human–computer interaction · Technical supervisor

1 Introduction

The automation of driving technology advances quickly. It is associated with benefits concerning safety, reliability, and passenger comfort, as well as the reduction of economic and environmental costs of mobility (Litman 2020; Schoitsch 2016). It is considered key for the fundamental shift in transportation away from individual mass motorization to flexible on-demand mobility solutions, for instance, shuttle vehicles (Iclodean et al. 2020). However, highly automated driving (Society of Automotive Engineers level of automation 4; Society of Automotive Engineers 2021) in urban mixed-traffic environments will remain challenging for automated

vehicles. On this level of automation, situations may occur that the vehicle’s automated driving system cannot handle (Kalisvaart 2021). In the worst case, automated vehicles can cause situations in which goods or even people are harmed. For example, in one incident, automated vehicles operated by the robotaxi company Cruise blocked an ambulance, delaying a patient’s urgent transportation to the hospital (New York Times 2023). In another incident, a Cruise vehicle hit a pedestrian and pulled them several meters forward, inflicting serious injuries (Guardian 2023). In some cases, human operators may be able to free automated vehicles from situations characterized by ambiguity and uncertainty by tackling even unforeseen situations with creativity and ingenuity, thereby helping avoid situations like described above, and can be fruitfully included into automated transportation systems consisting of a highly automated vehicle (HAV) and a remote operator (RO).

In remote operation systems, an RO oversees vehicle operations from a control center. The RO overviews and analyzes traffic situations that automated vehicles encounter.

✉ Andreas Schrank
andreas.schrank@dlr.de

¹ Institute of Transportation Systems, German Aerospace Center (DLR), Lilienthalplatz 7, 38108 Brunswick, Germany

² Cognitive Psychology and Human Factors, Chemnitz University of Technology, Chemnitz, Germany

They provide guidance to the vehicle automation on how to tackle difficult situations. However, since the interaction between RO and HAV is essentially a shared control problem between human and machine, potential conflicting decisions between these two actors have to be identified and prevented (Abbink et al. 2018). A helpful approach to discover conflicting decisions is the heuristic-based CAP method that Vanderhaegen (2021) proposed. For service remote operation, it needs to be clear at any time which actor is responsible for which tasks. In order to determine the RO's tasks, this paper refers to industry standards and legal frameworks (see Sect. 1.1).

Remote operation is conceivable for any vehicles with high driving automation (Society of Automotive Engineers 4 or higher), shuttle buses, and other vehicles, e.g., personal vehicles, transportation vehicles such as vans and trucks, as well as larger buses. Remote operation could, therefore, help overcome situations that the automation alone cannot handle, resulting in safer and smoother operations of HAVs. A pivotal component of a safe and smooth HAV remote operation system is the RO's workplace. The following paper will describe the design for a conceptual prototype for a workplace for remote assistance, a variant of remote operation, and its user evaluation focusing on the central indicators performance, situation awareness, and workload.

1.1 Workplaces for remote operators

ROs will be a core component of HAV remote operation systems. The human-machine interface (HMI) of the RO workplace is essential for safe, effective, and efficient operations. Remote operation can mainly be implemented in two different ways. First, in the *remote driving* approach, also known as direct or teleoperated driving, the RO executes the dynamic driving task (DDT) including braking, steering, and accelerating in real time (Society of Automotive Engineers 2021). The input given resembles manual driving, requiring the RO's continuous attention. Second, the *remote assistance*, or indirect, approach is defined as "event-driven provision, by a remotely located human, of information or advice to [... a] vehicle in driverless operation in order to facilitate trip continuation when the ADS [automated driving system] encounters a situation it cannot manage" (Society of Automotive Engineers 2021, p. 18). The HMI presented and evaluated in this paper aims to enable remote assistance at Level 4 Automation (Society of Automotive Engineers, 2021). Since remote assistance, unlike remote driving, does not include the execution of the dynamic driving task (DDT), i.e., the longitudinal and lateral control of the vehicle, the proposed HMI does not enable the remote operator to complete the DDT (Society of Automotive Engineers, 2021, p. 18). In contrast, the focus is to provide assistance to the HAV in assessing a traffic situation and proposing

how to proceed. In accordance with the definition of Level 4, the remote assistant who oversees an HAV does not serve as a fallback for the automation. The HAV must be able to transfer itself into a minimal-risk state, posing the least danger possible to itself, its passengers, as well as surrounding road users. Remote assistance in this implementation is the only permissible way of implementing remote operation of vehicles on public roads in Germany to this date (StVG § 1e, 2021/12.07.2021).

During remote assistance, the RO's main task is the processing of requests for assistance coming from the supervised HAV (see Fig. 2). According to the German Autonomous Driving Act (StVG § 1e, 2021/12.07.2021), ROs specified as Technical Supervisors ("Technische Aufsicht") are responsible to check and assist an HAV based on evidence that it requires support ("Evidenzkontrolle"). This means that an RO becomes involved only when the vehicle detects an event that it cannot handle autonomously and thus submits a request for assistance to the RO (StVG § 1e, 2021/12.07.2021). In this case, the HAV must be able to conduct a minimal-risk maneuver (MRM) independently, i.e., bring itself to a halt in a safe manner and at a safe position. The RO can intervene only after the successful completion of the MRM. The RO's intervention must not be time critical, i.e., does not need to be completed in a specified amount of time. The RO has the following responsibilities: (1) giving clearance to alternative driving maneuvers, (2) deactivating the autonomous driving function, (3) assessing the HAV's signals regarding its functioning and initiate measures for ensuring safety, as well as (4) getting in contact with the HAV's passengers in the event of an MRM (StVG § 1f). In addition, the RO can propose driving maneuvers themselves if the HAV is unable to do so. The presented user study investigates some of these responsibilities using the proposed HMI for the RO's workplace, including giving clearance to driving maneuvers proposed by the HAV (Scenario 1), suggesting driving maneuvers themselves by specifying waypoints to define a pathway that the HAV needs to follow (Scenario 2), as well as by selecting an alternative route (Scenario 3).

Even though German law demands that interventions by the RO cannot be time critical, (a) *task reaction time*, i.e., the duration passed from the request's appearance on the RO's workplace HMI to the RO's acceptance of the request, is still considered a key performance indicator as it is essential for efficient operations and therefore relevant for the economically feasible implementation of RO systems. In addition, (b) *task completion time*, i.e., the time passed from the RO's acceptance of the request to the resolution of the task, is an indicator to measure how long it took an RO to resolve a task.

The literature on workplace HMIs for remote operation is scarce. Following a human-centered design process,

Kettwich et al. (2021) designed and evaluated a click prototype for a remote operation workplace HMI. It was tailored to the remote assistance of Society of Automotive Engineers 4 shuttle buses from a public transport control center. Apart from this research, although software and hardware solutions for the remote operation of vehicles already exist (e.g., DriveU.auto 2023; Herger 2023; T-Systems 2023; Vay 2022), no systematic research has been conducted in a highly controlled laboratory environment to develop and evaluate a prototypical HMI for HAV *remote assistance* to the authors' knowledge. Remote assistance here is defined in accordance with Society of Automotive Engineers J3016 as an "event-driven provision, by a remotely located human [...], of information or advice to an ADS-equipped vehicle in driverless operation in order to facilitate trip continuation when the ADS encounters a situation it cannot manage." This definition is similar to the task of the Technical Supervisor according to the current German Autonomous Driving Act. In particular, there is a gap in research on workplace HMIs for remote operation of vehicles in the contexts of public transport, logistics, and individual mobility that are tailored to the needs, expectations, and operation styles of control centers in these areas. Therefore, the goal of this work is the user-centered design of a prototypical workplace HMI for a concrete implementation of remote operation, remote assistance, and its evaluation regarding performance, situation awareness, and workload in routine remote assistance tasks. Also, we want to assess the operator's subjective experience by assessing their ratings of usability, user experience, and acceptance.

1.2 Situation awareness

Similar to a driver, a remote operator (RO) needs to perceive and identify relevant elements of a traffic situation. They must integrate them to a coherent understanding of the situation and be able to predict how relevant elements will change in the future. These operator tasks can be described by situation awareness (SA). The hierarchical SA model of Endsley (1995) proposes three levels of SA. A lower level of SA needs to be fulfilled in order to reach a higher one. On SA Level 1, a RO has to *perceive* characteristics of the traffic environment like road layout and condition, traffic signs, and other road users. On SA Level 2, the RO has to analyze and integrate these elements in accordance with their goals to "form a holistic picture of the environment, *comprehending* the significance of objects and events" (Endsley 1995, p. 37). For example, a pedestrian crossing an HAV's lane is relevant to the RO's goal to continuously drive on this lane. On SA Level 3, the RO *predicts* how the situation will unfold. A result of high SA is that the RO commands the HAV change lanes in order to avoid the predicted collision.

In a remote setting, it may be difficult to achieve high levels of SA because ROs cannot perceive the elements of the driving situation directly and without delay (SA Level 1), or react immediately to them (based on SA Levels 2 and 3). Also, there is no direct link between a RO and the surrounding traffic environment. Information of the driving situation is sensed via technology, transmitted to the RO's workplace, and displayed to them through the interface. Similarly, the RO's reaction is mediated through data transmission, in-vehicle processes, and execution by actuators, causing delays between operator inputs and vehicle reactions as well as vehicle actions and presented status. Decoupling action, perception, decision, and reaction by inserting intermediate steps of deconstruction, transmission, and reconstruction into the process has important implications: distortions may occur in any of these steps, negatively impacting the RO's SA (Tittle et al. 2002). For instance, Darken et al. (2001) stated that participants performed poorly in spatial orientation as well as object identification tasks when video feedback was supplied to remote observers. Thus, the HMI design of the RO workplace concerning the selection of information modes (visual, auditory, etc.) and the way information is displayed to the RO affects their level of SA (Endsley 1995; Endsley et al. 2003; Hollands et al. 2019). As a result, the RO's workplace needs to ensure high levels of SA.

Specifically, the RO's tasks investigated in this study require the RO to generate and keep up SA on all three levels. In Scenario 1, for example, in order to give clearance to the HAV to conduct the proposed driving maneuver, the RO first needs to recognize the relevant objects in the scenario accurately, including the correct perception of the buildings along the street and the puddle on the street (SA Level 1). Second, the RO needs to be able to integrate the perceived information, i.e., identify the buildings appearing in the puddle as mere reflections rather than actual obstacles (SA Level 2). Third, the RO needs to draw conclusions from the integrated information (SA Level 3). In this case, the RO can conclude that there is no obstacle on the street ahead, so they can give clearance to continue the HAV's ride on the planned pathway.

1.3 Workload

Workload is the experienced difference between required and supplied information processing capability (Hart and Staveland 1988). It is associated with task performance. Therefore, a workplace for remote operation should balance the requirements of tasks to avoid overload, which leads to stress, or underload, which is associated with boredom (Wickens 1984). A good overview of the tasks that need to be completed is therefore essential for the RO to balance

their workload across time. The proposed HMI fulfills this requirement by presenting every request for assistance in a table with the most vital pieces of information, including its status, i.e., whether it still needs to be accepted, is currently processed, or already completed. This view helps the RO realize the current situation, including the number of open requests and which ones need to be prioritized, and therefore facilitates balancing their workload.

In workplace design, all tasks, be they primary or secondary, need to be considered in workload assessment. For primary tasks, this study uses scenarios that are assumed to cause varying levels of workload, for instance, the task of giving clearance to a maneuver that the HAV proposed by itself (Scenario 1) is expected to generate less workload than the task of determining waypoints on a map view (Scenario 2). Secondary tasks pose additional cognitive load on operators (Sweller 1988), thereby increasing perceived workload. These tasks can be (a) directly relevant to fulfill the primary task, for example, when additional pieces of information need to be gathered from other sources. They can also be (b) indirectly relevant as part of other responsibilities of an operator, e.g., an incoming request for support by an HAV while already processing another HAV's request. However, they can also be (c) irrelevant to an operator's responsibilities, i.e., distractions. An example of the fatal consequences of being distracted from job-related tasks is the rail disaster of Bad Aibling in Bavaria, Germany. A train controller distracted himself from his rail traffic management task by playing a game on his phone, leading to a collision of two trains on a single-track stretch, killing twelve (British Broadcast Corporation 2016).

In Human Factors research, examining the impacts of a secondary task on an operator's workload has a long tradition (e.g., Ogden et al. 1979). In the case of a RO's task set, generic cognitive secondary tasks such as the n-back task (Kirchner 1958) can be used as proxies for cognitive load that might result from tasks that the RO could have, like the RO's parallel assistance of several HAVs. The n-back task is widely used in driving-related studies to systematically vary workload (Pfanmüller et al. 2015; Reimer and Mehler 2011; Wu et al. 2019).

Hence, workplaces for ROs should be designed so that primary and secondary tasks do not lead to an increase of the operator's workload that would severely deteriorate performance. This is especially important as processing multiple tasks at the same time affects the operators' workload and their SA. In these situations, operators need to keep multiple pieces of information in their working memory, leaving less cognitive resources for gaining high levels of SA (cf. Baumann et al. 2008).

To summarize, an HMI for remote operation needs to be designed to enable effective and efficient operations, to

balance the RO's workload, and to ensure their SA. In addition, user-focused variables need to be considered.

1.4 Usability, user experience, and acceptance

The user's subjective usability is crucial for their smooth interaction with technical systems. The perceived usability is relevant because it determines how well the user is able to access information from the system and interact with it. High subjective usability is achieved when the interaction between user and system is effective, efficient, and satisfying (International Organization for Standardization 2018). User experience is a concept that assesses how satisfied users are when interacting with a system (Hassenzahl 2008; Minge et al. 2017). It is inevitable for developing successful user-centered products (Schrepp et al. 2017a). Finally, user acceptance is imperative for the success of newly introduced technology as it determines whether a new technology will be adopted by its designated user group (van der Laan et al. 1997).

All these concepts are of utmost importance when it comes to workplace design as they directly influence efficient, effective, and safe operations. The HMI for remote operation needs to be designed in a way that enables the RO to quickly obtain an overview of the HAVs' requests for assistance, to be presented any information on the requests that is needed to answer the requests, and to enter the advice to the HAV on how to behave in the given situation. The quality, ease, and efficacy of the direct interaction with the HMI is captured in the construct usability. Further, the repeated interaction with the HMI constitutes its emotional valence, as represented by user experience. We aimed to achieve the repeated interaction through an extensive training phase and the repetition of trials using a limited set of routine tasks. In order for the participants to experience the interaction with the HMI in different states of perceived workload, we administered a secondary task to simulate additional cognitive load. This paradigm ensures that the measured user experience is also valid in more cognitively challenging situations, capturing a more diverse range of interactions.

1.5 Human-machine interface (HMI)

The structure and components of the HMI for remote assistance strongly resembled the click prototype presented and positively evaluated in Kettwich et al. (2021). To the knowledge of the authors, it is the first workplace HMI for the remote assistance of HAVs in the literature. Particularly, it follows the currently valid legal requirements for highly automated driving in Germany, rendering it a legally compliant approach to implement remote assistance. To achieve this, the initial click prototype was further iterated following

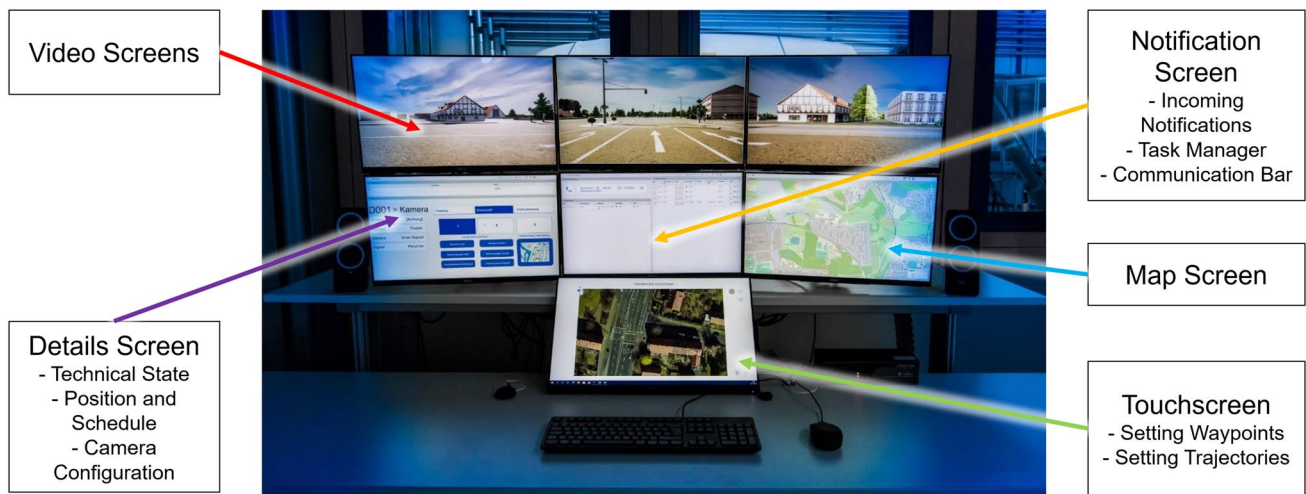


Fig. 1 The HMI of the prototypical workplace for remote assistance

the user-centered design process, incorporating the qualitative feedback from the initial evaluation of the click prototype. Particularly, a higher degree of immersion was implemented by translating the click prototype on one screen to a full prototypical setup using seven screens that is very close the final setup of an RO's workplace. The resulting prototypical workplace for remote assistance is depicted in Fig. 1.

The workplace consists of seven screens of which six were regular computer monitors (24" Dell, 16:9 ratio), set up in two rows with three monitors each, and another monitor with the same specifications but including a touch feature ("Touchscreen," see Fig. 1). The basic elements of the HMI as well as the interaction design are described in detail by Kettwich et al. (2021). The following screens are parts of the workplace:

- **Video screens:** On the three top screens, the live video stream from the supervised HAV is displayed. For the study, simulated video sequences were created in the Unreal Engine for each scenario.
- **Details screen:** On this screen, the RO can view information on the status of the fleet of HAVs, the technical status of each HAV, and its exact position and schedule, and can select various camera configurations.
- **Notification screen:** Here, the RO is shown new incoming requests (left column), the status of accepted requests (right column), and a communication bar to initiate a voice connection with actors of interest, including other departments of the remote operation center and the operator of the HAV service, police, and rescue services.
- **Map screen:** A global map presents the currently assisted HAV in its center as well as the surrounding HAVs that are supervised by the remote operation

center. Additionally, layers such as current load closures, stops, and other points of interest can be activated.

- **Touchscreen:** It presents a highly detailed view of the immediate area around the HAV and enables the RO to interact with the vehicle by giving clearance to suggested driving maneuvers (Scenario 1), setting waypoints to create pathways for the HAV to follow (Scenario 2), and to select alternative routes (Scenario 3).

The steps of the interaction between RO and the supervised HAV are depicted in Fig. 2. In addition, they are pointed out in detail for each scenario in Table 15 in the appendix.

In all three scenarios (see Fig. 3), the HAV drove in highly automated mode before noticing that it needed the RO's support. Subsequently, it submitted a request to the RO's workplace. The operator received the request for support on the central screen of the second row from the top in the section for incoming notifications that also included some core information such as the HAV's ID, the issue that requires the RO's support, and the spatial position of the HAV. By clicking on "Accept," the RO could allocate the task to themselves, transferring the request to a table containing current tasks. Here, further details, such as the latest video stream from the HAV, its position on a map, and details regarding its technical state, were displayed. Furthermore, a suggestion for an action, such as "Give Clearance," "Set Waypoints," or "Select Alternative Route," were provided. This information supported the RO's decision on how to assist the HAV. Finally, the RO's input was transmitted to the HAV and executed before it returned to the highly automated driving mode.



Fig. 2 Interaction between remote operator and the supervised HAV between phases of autonomous driving



Fig. 3 Screenshots from video clips of simulated scenarios created with the Unreal Engine and used in this study. The differences in illumination appear more pronounced here than on the prototypical workplace since separate screens were used to display the images,

putting less emphasis on these differences. **a** Scenario 1: detected situation unclear, **b** Scenario 2: blocked lane, and **c** Scenario 3: rerouting

1.6 Research objectives and hypothesis

The goal of this work is the user-centered design of a novel HMI for remote assistance following the established guidelines and its evaluation regarding the key Human Factors outcome variables performance, situation awareness (SA),

and workload. Also, we want to assess the participants' perceived usability, ratings usability, user experience, and acceptance when interacting with the HMI. To achieve this goal, three research objectives were examined in this study.

The first objective examined whether participants show lower performance at increasing levels of cognitive

demand in routine remote assistance tasks using the proposed workplace HMI for remote assistance. The overall hypothesis was as follows:

- H1 (performance): When the level of induced cognitive demand increases while completing tasks using the designed workplace for remote assistance, participants' performance decreases.

It separates into three sub-hypotheses:

- H1.1 (task reaction time): When the level of induced cognitive demand increases, participants require more time to react to an incoming notification which manifests in more time passed from the appearance to the acceptance of the notification.
- H1.2 (task completion time): When the level of induced cognitive demand increases, participants require more time to process a task which manifests in more time passed from the acceptance of the notification to the completion of the task.
- H1.3 (number of correct n-back comparisons): When the level of induced cognitive demand increases, participants' number of correct n-back comparisons decreases.

The second objective tested whether participants report lower SA at increasing levels of cognitive demand routine while processing remote assistance tasks using the proposed workplace HMI for remote assistance. The corresponding hypothesis was as follows:

- H2 (subjective SA): When the level of induced cognitive demand increases while completing tasks using the designed workplace for remote assistance, participants' reported SA ratings decrease.

The third objective examined whether participants report higher workload with increasing levels of cognitive demand while processing remote assistance tasks using the proposed workplace HMI for remote assistance. Here, the hypothesis was as follows:

- H3 (subjective workload): The participants' reported ratings of workload increase with increasing levels of induced cognitive demand while completing tasks using the designed workplace for remote assistance.

In addition to these objectives, we examined the participants' ratings of usability, user experience, and acceptance. Thereby, we wanted to gain first insights on the participants' subjective experience with the remote assistance workplace. Our analysis examined how participants assess the usability

of the presented HMI for remote operation and how they rate their satisfaction.

2 Method

2.1 Sample

Participants were acquired through postings in buildings and online platforms of engineering departments of universities and research centers in Germany. The participants volunteered but were compensated monetarily for their participation with 25 euros. This study was conceptualized and realized in accordance with the Declaration of Helsinki. The institutional review board of the research institution in which this study was conducted approved this study. Informed consent was obtained from all participants before the experiment. The participants were allowed to stop the study at any point without justification or consequence.

Of the $N=41$ participants who took part in this study, seven had to be excluded due to issues in the data collection process. Technical issues in the tools used for the collection of either questionnaire or performance data rendered some data unusable in these participants. Only participants with complete datasets ($N=34$) were included in the analysis. The final sample analyzed consisted of 34 participants (four female). Participants' ages ranged from 23 to 31 years ($M=26.2$, $SD=2.31$). 62% of the participants had experience in monitoring technical systems such as airplanes, automated vehicles, wind channels, agricultural robots, pumps, and machines. Their affinity for technology (Franke et al. 2019) was high ($M=4.94$, $SD=0.48$; scale poles 1: low to 6: high). All participants had normal or corrected-to-normal vision and possessed a valid driver's license for passenger vehicles. Only participants with a university or state-certified technician degree in the following disciplines were accepted: mechanical, automotive, electrical, aerospace, and aviation engineering. The reason behind this criterion was our objective to closely model the group of participants on the requirements posed to the Technical Supervisor, the German equivalent of the RO as specified in the German Autonomous Driving Act (StVG, 2021). This law demands that a Technical Supervisor have a degree in the listed engineering disciplines. This criterion therefore ensured that only participants that were deemed qualified for this work by the law, at least regarding their education background, were included in this study. 21 participants (62%) held a Bachelor's degree as their highest academic degree, thirteen a Master's degree. More than a third (35%) of the participants stated to drive a vehicle multiple times per month and about 29% reported to drive several times a week. All participants had heard about HAVs in the past. 91% expressed interest or strong interest in

HAV technology indicated by responding with values 4 or 5 on a Likert scale on interest in AVs (1: “not interested at all” to 5: “very interested,” $M=4.29$, $SD=0.72$). 28 participants (82%) indicated not to have used HAVs so far.

2.2 Experimental design

The experimental design was a 3×3 within-subject design. The independent variables were the primary task (Scenarios 1–3) and the secondary task to induce additional cognitive load (none, 1-back, 2-back). Dependent variables were performance in primary and secondary task, workload, SA, usability, user experience, and acceptance (see Sects. 2.3.3 and 2.3.4).

2.3 Materials

2.3.1 Primary task (scenarios)

Three scenarios were used as primary tasks in this study. Figure 3 displays a screenshot of each scenario. The scenarios that target highly automated driving (Society of Automotive Engineers Level 4) were extracted from a previously compiled catalog of scenarios in remote operation (Kettwich et al. 2022) because they were considered typical for routine tasks in remote assistance. This catalog is based upon in-depth interviews, observation studies, and video analyses with control center employees. Also, the scenarios are representative as similar scenarios are already used by leading operators of HAV fleets on public roads. For instance, automated vehicle operator Waymo utilizes remote operation when an automated vehicle finds a closed road on its way, requiring rerouting (Amador et al. 2022). Similar tasks were also confirmed to be used by the robotaxi service Cruise in California, USA (CNBC 2023). The scenarios were implemented in the Unreal Engine (Epic Games 2019) and extracted as video clips. These video clips were played to the participants before and after they interacted with the remote operation workplace.

Detailed steps of the interaction between RO and workplace are listed comprehensively in Table 15 in the appendix.

Scenario 1: Detected Situation Unclear. In this scenario, the supervised HAV detects an obstacle on the road, stops, and reports the incident to the RO. The detected obstacle is a puddle on the road which reflects the surrounding buildings, so the automation is uncertain whether the vehicle can continue its ride. The RO observes the situation via the supervised HAV’s on-board cameras (transmission of video images) and gives clearance, so the vehicle can continue its journey. After assessing the situation, the primary task for the RO therefore is giving clearance to continue driving. The task of giving clearance resembles the fulfillment

of “confirmation requests” by ROs as stated to be used by Cruise (CNBC 2023).

Scenario 2: Blocked Lane. A vehicle is parking on the lane that the supervised HAV uses, blocking the lane and disabling the HAV from continuing its ride. The HAV stops and provides a corresponding message to the RO. The RO checks the situation on site via the HAV’s cameras and sets waypoints for an alternative trajectory using the lane for oncoming traffic to bypass the parking vehicle. The primary task for the RO is to set waypoints to calculate a new trajectory. The task of giving clearance resembles the ROs’ “guiding the AV through tricky situations” as stated to be used by Cruise (CNBC 2023).

Scenario 3: Rerouting. Because of a road closure, the supervised HAV needs to change its route. The RO views the road closure via the HAV’s cameras and suggested alternative routes and chooses one of them. Thus, the RO’s primary task is selecting one of the several proposed routes presented on the touchscreen. The task of rerouting is an RO’s responsibility in the event of road closures, as confirmed by Waymo (Amador et al. 2022).

2.3.2 Secondary task

As a secondary task, the n-back task (Kirchner 1958) was included. Its purpose was to modulate the RO’s cognitive load in order to simulate phases of elevated workload that are likely to occur in the RO’s work. In this task, participants had to compare a presented digit with the digit presented n steps before the current one. The higher the n , the more digits had to be retained in the participants’ working memory, increasing their workload. The n-back task was presented visually and auditorily on a tablet computer distinct from the investigated workplace HMI (Fig. 4, bottom right). However, participants were instructed to listen to the auditory presentation only and give their response verbally. The experimenter assured that participants followed these instructions. From a list of 30 digits plus n , a single digit from one to nine was played auditorily using the tablet computer’s speaker and displayed visually on the screen of the tablet computer every five seconds, so that a total of 30 n-back comparisons had to be made per trial. The order of digits was determined for each trial by randomly assigning one out of four lists that contained a specific order of digits. Participants were instructed to respond verbally with “correct” if the digit presented n steps before the current one was identical with the current one, and to respond with “incorrect” if the digit presented n steps before the current one differed from the current one. Participants were asked to respond before the next digit was presented, i.e., they had to respond within less than five seconds. The experimenter logged the participants’ responses.

Fig. 4 The prototypical workplace for remote assistance and a tablet computer to present the n-back task's stimuli (small dark screen on the right-hand side of the touchscreen). The top row of screens shows the video streams from the on-board cameras of the simulation. The second row of screens presents information on the technical state of the supervised HAV (left screen), an overview of vehicle requests and their respective status (center), and a map of the environment surrounding the supervised vehicle (right). The touchscreen on the bottom is used by the RO to interact with the supervised vehicle, e.g., by setting waypoints or selecting alternative routes on a map, depending on the scenario



2.3.3 Objective measures

We collected three measures that quantified the participants' performance, two of them regarding the primary task and one regarding the secondary task.

Measures of Primary Task. Regarding the primary task, the objective was to examine how fast participants were able to react to incoming notifications. Even though remote operation must not be time critical by law, from an economic point of view, a speedy reaction is still favorable to enable a business case built on remote operation. In addition, the duration participants spent for completing the task was measured to investigate if the HMI is suitable to fulfill the RO's task in a timely manner. Hence, we measured the participants' performance in the primary task using two variables: (a) the time that passed from appearance to the acceptance of the support request in seconds, hereafter called *task reaction time*, and (b) the time that passed from acceptance to completion of the support in seconds, called *task completion time*.

Measure of Secondary Task. To measure how much cognitive load was induced, we measured the participants' performance in the secondary task using the number of correct n-back comparisons (max. 30) in the n-back task (see Sect. 2.3.2 for further details).

2.3.4 Questionnaires

In addition, we collected self-report data using six questionnaires.

NASA-TLX. The NASA Task Load Index (Hart and Staveland 1988) was used to measure subjective workload after each trial. It is an established multi-dimensional measure for participants to report how taxing they experienced a task. The questionnaire distinguishes between six dimensions of workload: mental demand, physical demand, temporal demand, performance, effort, and frustration. Responses to each item were collected on a 7-point Likert scale ranging from the poles 1: "low" to 21: "high."

SART. The Situation Awareness Rating Technique (SART; Taylor 1990) assessed the participants' situation awareness (SA) post-trial. It was originally developed to determine pilots' SA and consists of three subscales: demands on attentional resources, supply of attentional resources, and understanding of the situation. Responses to each item were collected on a 7-point Likert scale. The poles depended on the specific item but ranged from a low to a high degree on a specific construct. For instance, the poles of the item "instability of the situation" were 1: "The scenario is entirely stable" to 7: "The scenario is entirely unstable." An overall SART score was calculated by deducting the difference of attentional demand and attentional supply from understanding.

SUS. The Systems Usability Scale (SUS; Brooke 1996) measures perceived usability. Originating from the need to quickly evaluate usability in software development, it is an economic solution to assess the construct robustly and across a wide range of domains (Bangor et al. 2008). Responses to each item were collected on a 5-point Likert scale ranging from the poles 1: “I do not agree at all” to 5: “I totally agree.” A single indicator value between 0 and 100 summarizes the status of the investigated HMI regarding the participants’ impression how well it was suited to execute a particular task. As a global assessment tool, SUS was administered at the end of the study.

UEQ-S. The User Experience Questionnaire Short Version (UEQ-S; Schrepp et al. 2017b) assesses user experience. It consists of two subscales, the pragmatic and the hedonic subscale. While the pragmatic one captures a construct that leans toward usability, the hedonic one focuses on the emotional quality of the interaction. The UEQ-S is a condensed version of the standard UEQ (Schrepp et al. 2017a), compressing the six subscales from the standard UEQ to the beforementioned two subscales. Using the UEQ-S ensured a more economic collection of subjective data on participants’ emotional experience with a system. Responses to each item were collected on a 7-point Likert scale. The poles were semantically opposed statements on a construct each, e.g., 1: “obstructive” to 7: “supportive.”

Acceptance scale. The Acceptance Scale (van der Laan et al. 1997) was developed as a standard tool to measure driver acceptance of new technology. With nine items divided in two scales, it measures the usefulness of a system, associated with usability, and the user’s satisfaction with said system, similar to user experience. Thus, it is conceptually related to the structure of UEQ-S but adds the dimension of user acceptance when both subscales are considered holistically. Responses to each item were collected on a 5-point Likert scale. The poles were

semantically opposed statements on a construct each, e.g., 1: “useful” to 5: “useless.”

ATI. The Affinity for Technology Interaction Scale (ATI; Franke et al. 2019) assessed the participants’ affinity for technology. This construct entails a person’s tendency to engage in interaction with technology. With its satisfying psychometric characteristics, the ATI scale measures the affinity for technology with nine items. In this study, the construct was used to describe the sample. Responses to each item were collected on a 6-point Likert scale ranging from the poles 1: “I completely disagree” to 6: “I completely agree.”

2.4 Procedure

An overview of the procedure is given in Fig. 5. First, the experimenter welcomed participants, briefed them about the objectives of the study, and asked them to sign an informed consent, a non-disclosure agreement, and a data protection declaration. Subsequently, participants filled in the sociodemographic questionnaire and completed the ATI scale (Franke et al. 2019) before they received a detailed explanation of the research context. Participants were instructed to imagine being a remote operator who assists HAVs that operate as shuttle buses in public transport. An image of an exemplary HAV was presented to the participants. They were also informed about the setup and features of the RO’s prototypical workplace for remote assistance. Next, the experimenter invited participants to take a seat in front of the workplace.

Subsequently, participants were asked to adjust their swivel chair, so they could see every screen well and were able to reach all input devices, i.e., mouse, keyboard, and the tablet computer for administering questionnaires. The experimenter described the features of the workplace screen by screen. Afterward, participants were encouraged to familiarize themselves with the workplace HMI independently by

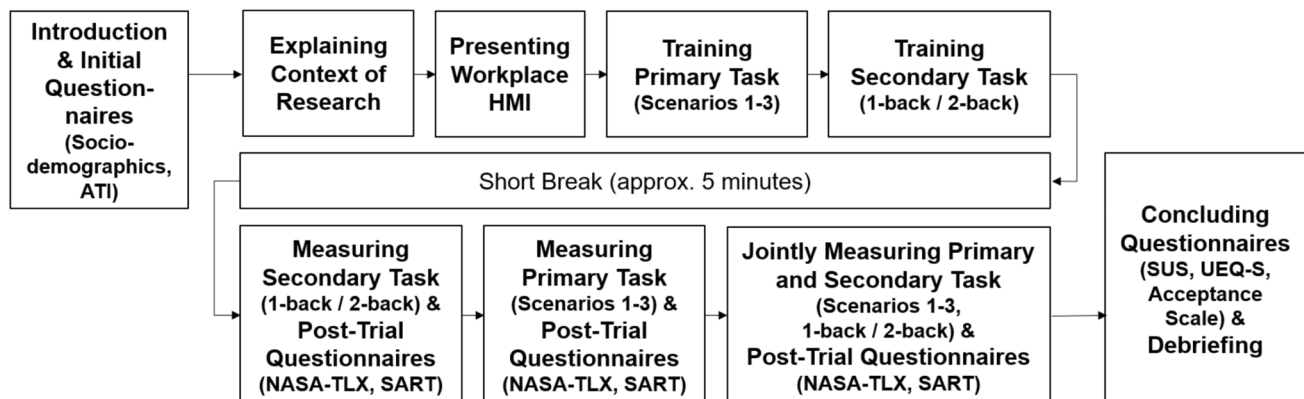


Fig. 5 Overview of the study procedure

closely looking at all the screens, clicking around, and learning about the implemented features. Once they were confident to have acquired a general understanding of the structure and features of the workplace, they were instructed to notify the experimenter. All participants did this within 5 to 10 min. Next, they were guided through the task completion process of the three scenarios (primary tasks, see also Sect. 2.3.1) by the experimenter who commented on each step. The experimenter ensured that participants were able to understand the sequence of actions and possible interactions in all scenarios. Participants were invited to ask questions. Next, they were familiarized with the secondary task, the n-back task (Kirchner 1958). A tablet computer located right next to the touchscreen (see Fig. 4) visually and auditorily presented a digit from 1 to 9 in intervals of 5 s. The participants were given an example each for both variations of the task (1-back and 2-back) and underwent a trial each to familiarize themselves with the task until they felt confident with it. A short break of approx. 5 min concluded the training block.

In the first experimental phase that measured the baseline performance in the secondary task, participants completed one trial for both variants of the secondary task (1-back, 2-back) in a balanced order. After each of the two trials, they filled in the NASA-TLX questionnaire. In the second experimental phase that measured the baseline performance in the primary task, participants completed each of the three scenarios of the primary task (1, 2, 3) in a balanced order. Before carrying out the tasks, participants were instructed to make the journey as smooth, quick, and seamless as possible for the passengers of the HAV. Also, they were reminded of their responsibility for the passengers' safety that can only be fulfilled appropriately if close attention was directed to all the information presented on the screens and recommendations with utmost care. Participants completed the NASA-TLX and SART after each trial. Subsequently, participants completed the joint data collection of the primary and secondary task with the same n-back variant. Two blocks (1-back, 2-back) of three trials each (Scenarios 1, 2, 3) were administered in counterbalanced order. Again, participants completed the NASA-TLX and SART after each of the six trials. Participants were instructed to give priority to completing the primary task but, at the same time, not to neglect the secondary task because failure to do so would disable the supervised HAV, resulting in passenger dissatisfaction. In each trial, participants performed the secondary task alone in the first 25 s of automated driving before the primary task was presented and had to be resolved by the participants. After completing the primary task, the secondary task lasted until 30 n-back comparisons were carried out.

Finally, participants filled in the questionnaires assessing usability, user experience, and acceptance, and were encouraged to provide remarks on the HMI and the study overall. The whole procedure took about 2.5 h.

3 Results

As this study was conducted using a within-subject design, a repeated measures analysis of variance (RM-ANOVA) was applied to determine the influence of primary and secondary task condition on the outcome variables presented above. Since the Mauchly (1940) sphericity test indicated a violation of sphericity in some cases, the Greenhouse–Geisser (1959) correction was applied in all reported RM-ANOVA results. In addition to the RM-ANOVA, post hoc pairwise comparisons with Bonferroni (1936) correction were performed to identify significant differences between specific groups.

3.1 Performance (H1)

To test Hypotheses 1.1 to 1.3, multiple statistical procedures were used. To test whether participants required more time to react to an incoming notification under varying levels of cognitive demand (H1.1), a 3×3 RM-ANOVA was computed. The descriptive statistics regarding task reaction time are presented in Table 1. The main effect of primary task condition (scenario) on task reaction time was not significant, $F(2, 66) = 0.798$, $p = 0.448$, $\eta^2 = 0.024$ (Fig. 6). There was also no main effect of the secondary task condition on task reaction time, $F(2, 66) = 3.178$, $p = 0.063$, $\eta^2 = 0.088$. Thus, induced cognitive load did not affect reaction times to incoming notifications. The respective hypothesis (H1.1) could not be accepted. There was no significant interaction effect between primary task condition and secondary task condition on task reaction time either, $F(4, 132) = 0.597$, $p = 0.612$, $\eta^2 = 0.018$. As shown in Table 2, no significant differences were yielded by post hoc pairwise comparisons.

Another 3×3 RM-ANOVA examined whether participants needed more time from the acceptance of the

Table 1 Descriptive statistics of task reaction time in seconds by primary task (scenario) and condition of secondary task

Primary task (scenario)	Condition of secondary task	<i>M</i>	<i>SD</i>	CI (LL) ^a	CI (UL) ^a
1	None	7.08	3.93	5.71	8.45
	1-back	6.18	2.37	5.36	7.01
	2-back	6.16	2.16	5.41	6.91
2	None	7.55	4.63	5.94	9.17
	1-back	6.22	1.81	5.59	6.85
	2-back	6.18	2.44	5.33	7.03
3	None	6.36	2.54	5.47	7.25
	1-back	6.30	2.71	5.36	7.25
	2-back	5.92	2.12	5.18	6.66

^aCI 95% confidence interval, LL lower limit, UL upper limit

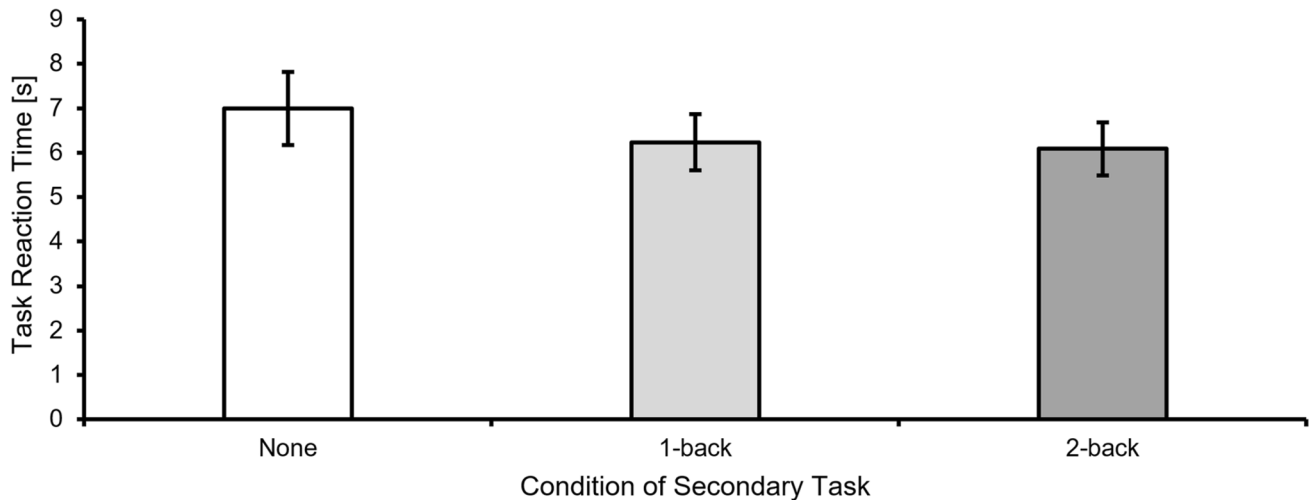


Fig. 6 Means of task reaction time by condition of secondary task. Bars indicate 95% confidence intervals

Table 2 Pairwise comparisons of means between conditions of secondary task regarding task reaction time in seconds (Bonferroni correction applied)

Condition of Secondary Task	M_{diff}^a	SE^b	p	CI (LL) ^c	CI (UL) ^c
1-back–none	−0.76	0.43	0.254	−1.84	0.32
2-back–none	−0.91	0.45	0.149	−2.04	0.22
2-back–1-back	−0.15	0.26	1.000	−0.80	0.51

^a M_{diff} difference of means between respective conditions

^bSE Standard error

^cCI 95% confidence interval, LL lower limit, UL upper limit

Table 3 Descriptive statistics of task completion time in seconds by primary task (scenario) and condition of secondary task

Primary task (scenario)	Condition of secondary task	M	SD	CI (LL) ^a	CI (UL) ^a
1	None	33.42	15.71	27.94	38.90
	1-back	29.43	8.46	26.48	32.38
	2-back	33.96	12.66	29.54	38.38
2	None	50.76	13.79	45.95	55.57
	1-back	47.16	11.38	43.19	51.13
	2-back	54.25	17.44	48.17	60.33
3	None	46.29	12.34	41.98	50.59
	1-back	45.32	9.18	42.12	48.52
	2-back	54.05	16.61	48.25	59.84

^aCI 95% confidence interval, LL lower limit, UL upper limit

notification to the completion of the task at increasing levels of cognitive demand (H1.2). The descriptive statistics regarding task completion time are presented in Table 3. A

Table 4 Pairwise comparisons of means between primary tasks regarding task completion time in seconds (Bonferroni correction applied)

Primary task (scenario)	M_{diff}^a	SE^b	p	CI (LL) ^c	CI (UL) ^c
1–2	−18.45	1.54	<0.001	−22.34	−14.57
1–3	−16.28	1.45	<0.001	−19.93	−12.64
2–3	2.17	1.71	0.635	−2.13	6.47

^a M_{diff} Difference of means between respective conditions

^bSE Standard error

^cCI 95% confidence interval, LL lower limit, UL upper limit

significant main effect of primary task condition on task completion time was found, $F(2, 66) = 82.814$, $p < 0.001$, $\eta^2 = 0.715$. Post hoc pairwise comparisons (Table 4) revealed that Scenario 1 took participants significantly less time to complete than Scenarios 2 and 3. This is a direct consequence of the task design, particularly due to the number of steps required to resolve it, and the kind of input the RO had to provide (see Sect. 4.1 for details). Also, a significant main effect of secondary task condition on task completion time was found, $F(2, 66) = 7.663$, $p = 0.002$, $\eta^2 = 0.188$ (Fig. 7). The more cognitive load was induced by the secondary task, the longer it took participants to complete the task. Therefore, hypothesis H1.2 was accepted. There was no significant interaction effect between primary task condition and secondary task condition on task completion time, $F(4, 132) = 1.784$, $p = 0.152$, $\eta^2 = 0.051$. As shown in Fig. 7, post hoc pairwise comparisons yielded a significant difference between the 2-back and the 1-back secondary task conditions ($M_{diff} = 6.78$, $p < 0.001$) but not between 2-back and

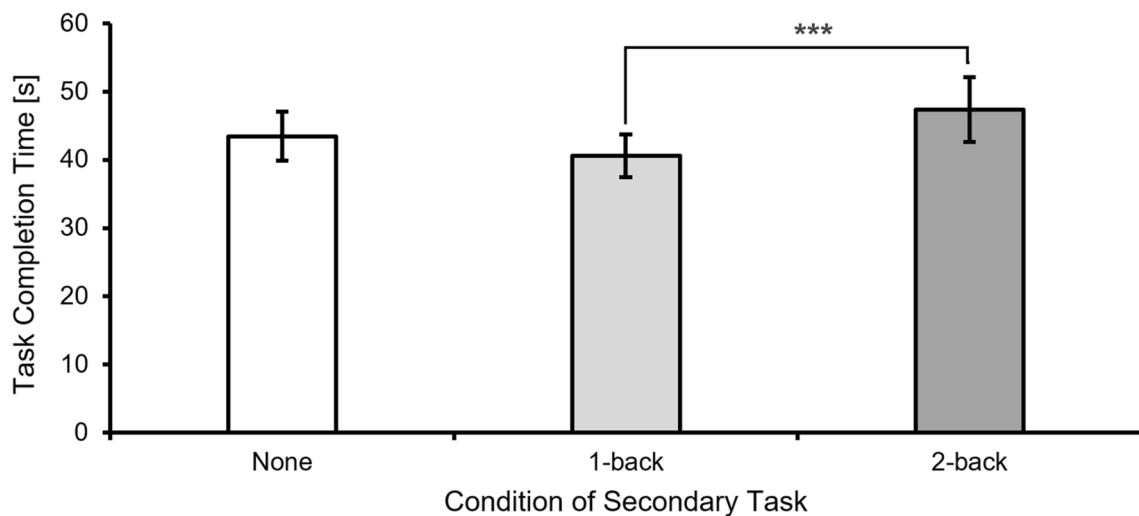


Fig. 7 Means of task completion time of task by condition of secondary task. Bars indicate 95% confidence intervals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5 Pairwise comparisons of means between conditions of secondary task regarding Task completion time in seconds (Bonferroni correction applied)

Condition of secondary task	M_{diff}^a	SE^b	p	CI (LL) ^c	CI (UL) ^c
1-back–none	-2.85	1.57	0.235	-6.81	1.11
2-back–none	3.93	2.12	0.217	-1.41	9.27
2-back–1-back	6.78	1.46	<0.001	3.10	10.46

^a M_{diff} Difference of means between respective conditions

^bSE Standard error

^cCI=95% confidence interval, LL lower limit, UL upper limit

Table 6 Descriptive statistics of numbers of correct n-Back comparisons in seconds by primary task (scenario) and condition of secondary task

Primary task (scenario)	Condition of secondary task	M	SD	CI (LL) ^a	CI (UL) ^a
1	1-back	29.82	0.39	29.69	29.96
	2-back	28.50	1.13	28.10	28.90
2	1-back	29.68	0.59	29.47	29.88
	2-back	28.38	1.60	27.83	28.94
3	1-back	29.71	0.68	29.47	29.94
	2-back	27.91	2.05	27.20	28.63

^aCI 95% confidence interval, LL lower limit, UL upper limit

no secondary task. Refer to Tables 4 and 5 for all post hoc comparisons.

Finally, a third 3 × 3 RM-ANOVA examined whether participants’ numbers of correct n-back comparisons decreased at increasing levels of cognitive demand (H1.3).

The descriptive statistics regarding number of correct n-back comparisons are presented in Table 6. As shown in Fig. 8 and Table 7, a significant main effect of secondary task condition on number of correct n-back comparisons was found, $F(1, 33) = 63.440, p < 0.001, \eta^2 = 0.658$. That means that in the secondary condition that induced a higher cognitive load, significantly less correct n-back comparisons were made. Therefore, hypothesis H1.3 was accepted. There was no significant main effect of primary task condition on number of correct n-back comparisons, $F(2, 66) = 1.885, p < 0.160, \eta^2 = 0.054$. In addition, the interaction effect between primary task condition and secondary task condition on number of correct n-back comparisons was not significant, $F(2, 55) = 1.250, p = 0.283, \eta^2 = 0.036$.

3.2 Situation awareness (H2)

Hypothesis 2 examined whether ratings of situation awareness (SA) on the SART scale decrease at increasing levels of cognitive demand. Again, a 3 × 3 ANOVA with repeated measures was conducted. The descriptive statistics regarding SART scores are presented in Table 8. Primary task did not significantly affect participants SART score, $F(2, 66) = 0.639, p = 0.531, \eta^2 = 0.019$. However, there was a main effect of secondary task condition on participants’ SART score, $F(2, 66) = 27.819, p < 0.001, \eta^2 = 0.457$. Globally, a higher induced cognitive load was therefore associated with a lower SART score. Consequently, the respective hypothesis that subjective situation awareness degraded as workload increased was accepted (Fig. 9). The interaction effect between primary task condition and secondary task condition on SART score was not significant, $F(4, 132) = 1.116, p = 0.349, \eta^2 = 0.033$. Pairwise post hoc

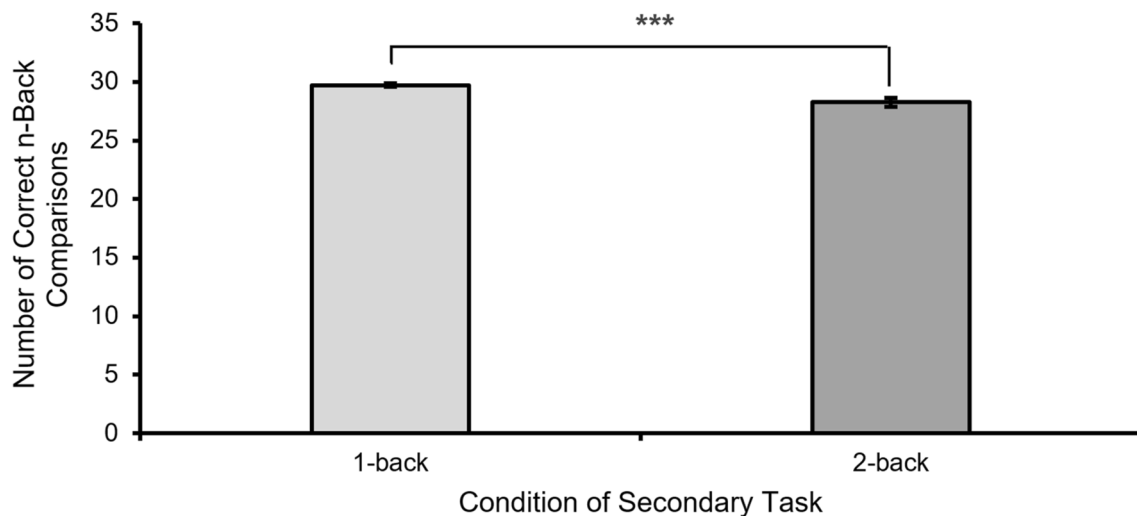


Fig. 8 Means of number of correct n-back comparisons by condition of secondary task. Bars indicate 95% confidence intervals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7 Pairwise comparisons of means between conditions of secondary task regarding number of correct n-back comparisons (Bonferroni correction applied)

Condition of secondary task	M_{diff}^a	SE^b	p	CI (LL) ^c	CI (UL) ^c
2-back–1-back	-1.47	0.18	<0.001	-1.85	-1.09

^a M_{diff} Difference of means between respective conditions

^b SE Standard error

^c CI =95% confidence interval, LL lower limit, UL upper limit

Table 8 Descriptive statistics of SART scores (Taylor 1990) for subjective situation awareness (low: -5 to high: 13) by primary task (scenario) and condition of secondary task

Primary task (scenario)	Condition of secondary task	M	SD	CI (LL) ¹	CI (UL) ¹
1	None	7.78	2.00	7.09	8.48
	1-back	7.24	1.83	6.60	7.87
	2-back	5.93	2.08	5.20	6.66
2	None	7.67	1.91	7.00	8.33
	1-back	7.04	2.02	6.34	7.74
	2-back	6.10	1.85	5.45	6.74
3	None	7.42	1.93	6.75	8.09
	1-back	7.15	1.82	6.51	7.78
	2-back	5.89	2.02	5.19	6.60

¹ CI 95% confidence interval, LL lower limit, UL upper limit

comparisons yielded significant differences between the 2-back and the 1-back secondary task conditions as well as between the 2-back and the no secondary task condition but not between 1-back and no secondary task (Table 9).

3.3 Workload (H3)

Hypothesis 3 tested whether participants' reported ratings of workload on the NASA-TLX questionnaire increase at increasing levels of induced cognitive demand. The descriptive statistics regarding NASA-TLX scores are presented in Table 10. A 3×3 RM-ANOVA revealed a main effect of primary task on NASA-TLX score, $F(2, 64) = 3.748$, $p = 0.041$, $\eta^2 = 0.105$. That means that among the primary tasks, subjective workload differed significantly. Post hoc pairwise comparisons (Table 11) revealed, however, that only between Scenario 1 and Scenario 2 workload was experienced significantly differently, not between any of the other pairs of scenarios. Additionally, the main effect of secondary task condition on NASA-TLX score reached significance, $F(2, 64) = 72.767$, $p < 0.001$, $\eta^2 = 0.695$ (see Fig. 10). Hence, the higher the cognitive load induced by the secondary task was, the higher the perceived workload was. The hypothesis that an elevated induced cognitive load leads to increased perceived workload was therefore accepted. Post hoc comparisons were significant between all conditions (Table 12). No effect of the interaction between primary task condition and secondary task condition on NASA-TLX score was found, $F(4, 128) = 0.257$, $p = 0.877$, $\eta^2 = 0.008$.

3.4 Questionnaires

In order to evaluate the workplace HMI overall, questionnaires regarding user-related variables including usability, user experience, and acceptance were administered after all task trials had been completed. This part of the study had exploratory character to understand if the designed HMI was

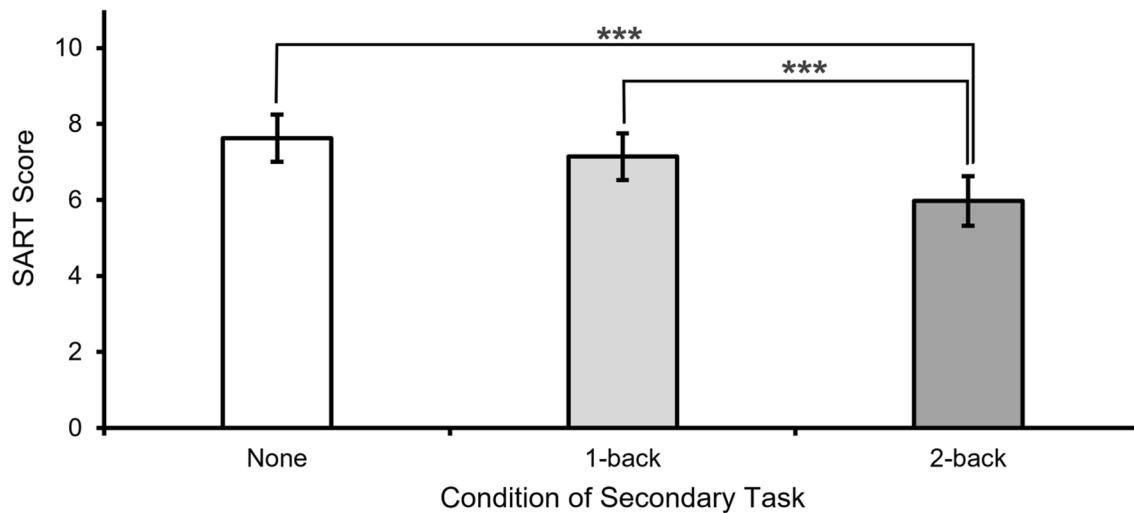


Fig. 9 Means of SART scores (Taylor 1990) for subjective situation awareness by condition of secondary task (low: -5 to high: 13). Bars indicate 95% confidence intervals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 9 Pairwise comparisons of means between conditions of secondary task regarding SART scores for subjective situation awareness (Bonferroni correction applied)

Condition of Secondary Task	M_{diff}^a	SE^b	p	CI (LL) ^c	CI (UL) ^c
1-back–none	-0.48	0.23	0.124	-1.058	0.091
2-back–none	-1.65	0.25	<0.001	-2.283	-1.017
2-back–1-back	-1.17	0.20	<0.001	-1.674	-0.660

^a M_{diff} Difference of means between respective conditions

^b SE Standard error

^c CI 95% confidence interval, LL lower limit, UL upper limit

Table 11 Pairwise comparisons of means between primary tasks (scenarios) regarding NASA-TLX scores for subjective workload (low: 1 to high: 21; Bonferroni correction applied)

Primary task (scenario)	M_{diff}^a	SE^b	p	CI (LL) ^c	CI (UL) ^c
1–2	-0.27	0.09	0.013	-0.09	1.06
1–3	-0.33	0.14	0.086	1.02	2.28
2–3	-0.06	0.15	1.000	0.66	1.67

^a M_{diff} Difference of means between respective conditions

^b SE Standard error

^c CI 95% confidence interval, LL lower limit, UL upper limit

Table 10 Descriptive statistics of NASA-TLX scores (Hart & Staveland 1988) for subjective workload (low: 1 to high: 21) by primary task (scenario) and condition of secondary task

Primary task (scenario)	Condition of secondary task	M	SD	CI (LL) ^a	CI (UL) ^a
1	None	6.35	1.98	5.66	7.04
	1-back	8.52	2.44	7.67	9.38
	2-back	10.11	2.38	9.28	10.94
2	None	6.72	2.00	6.02	7.41
	1-back	8.78	2.38	7.95	9.61
	2-back	10.30	2.42	9.46	11.14
3	None	6.57	1.80	5.93	7.21
	1-back	8.81	2.36	7.99	9.63
	2-back	10.55	2.36	9.73	11.38

^a CI 95% confidence interval, LL lower limit, UL upper limit

user-friendly beyond the scope of specific interactions investigated during the scenarios. Consequently, no hypotheses were stated a priori.

First, the System Usability Scale (SUS; Brooke 1996) yielded very good usability ratings ($M = 76.25$, $SD = 11.87$, on a scale from 0: very poor usability to 100: flawless usability). This score rates between the adjective ratings “good” ($M = 72.8$) and “excellent” ($M = 85.6$) as resulting from Bangor et al.’s (2008) empirical validation of the verbal interpretation of SUS scores. The interaction design between users and the investigated HMI was therefore regarded positive.

Second, user experience was measured with the User Experience Questionnaire Short Version (UEQ-S). The questionnaire consists of two subscales, the pragmatic and the hedonic scale. While the former scale pertains to a concept similar to usability, the latter focuses on the emotional component of user experience. Both subscales and the complete scale were tested against the arithmetic center of the scale, 0. This approach was based on the assumption that the

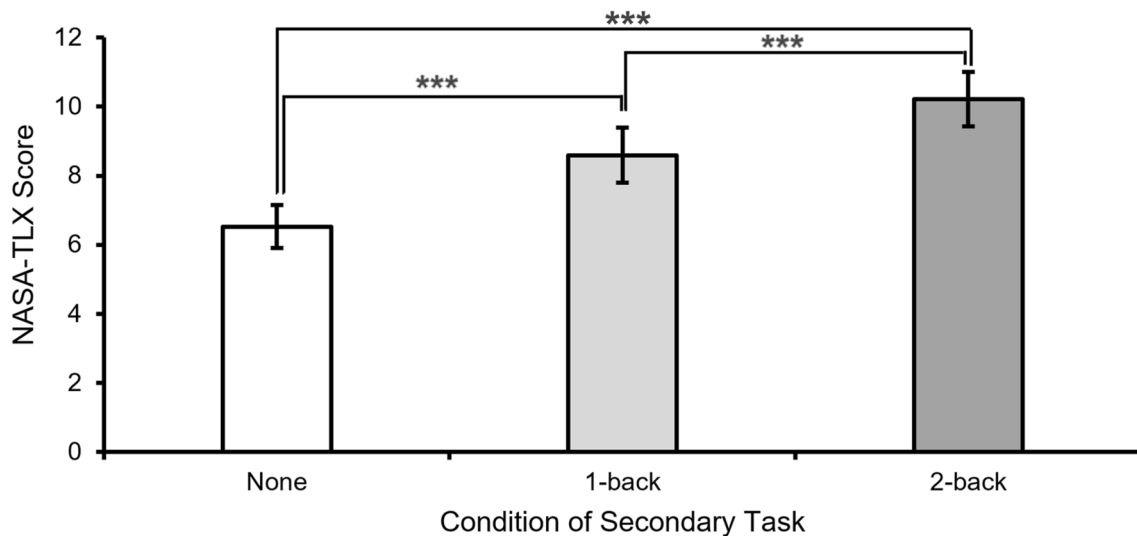


Fig. 10 Means of NASA-TLX scores for subjective workload (low: 1 to high: 21) by condition of secondary task. Bars indicate 95% confidence intervals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 12 Pairwise comparisons of means between conditions of secondary task regarding NASA-TLX scores for subjective workload (low: 1 to high: 21; Bonferroni correction applied)

Condition of secondary task	M_{diff}^a	SE^b	p	CI (LL) ^c	CI (UL) ^c
1-back–none	2.07	0.33	<0.001	–1.06	0.09
2-back–none	3.69	0.33	<0.001	–2.28	–1.02
2-back–1-back	1.62	0.25	<0.001	–1.67	–0.66

^a M_{diff} Difference of means between respective conditions

^b SE Standard error

^c CI 95% confidence interval, LL lower limit, UL upper limit

center of a scale represents a conceptual average value, e.g., a medium extent of usability. Table 13 shows the descriptive and test results. User experience is rated significantly higher than the arithmetic scale mean on both the complete scale and the pragmatic scale. This finding indicates that participants were satisfied with their interactions with the workplace HMI. Results on the hedonic scale did not differ

Table 13 Subjective user experience of the prototypical remote operation workplace measured with UEQ-S

Scale	M_{emp}	SD_{emp}	CI (LL) ^a	CI (UL) ^a	M_{crit}^b	t	p
Overall	0.38	0.50	0.21	0.56	0	4.502	<0.001
Pragmatic	0.63	0.46	0.47	0.78	0	7.966	<0.001
Hedonic	0.14	0.95	–0.19	0.47	0	0.861	0.395

User Experience Questionnaire—Short (UEQ-S; Schrepp et al. 2017a, b); low: 1 to high: 7 (center: 4)

^a CI 95% confidence interval, LL lower limit, UL upper limit

^b M_{crit} critical mean to test empirical mean (M_{emp})

significantly from the mean, suggesting average emotional experiences with the HMI.

Third, acceptance of the HMI was measured with Van der Laan et al.'s (1997) Acceptance Scale. Similar to the UEQ-S, it consists of two subscales, one of which focuses on the usability, or usefulness, to use the term that the authors of the questionnaire used, while the other one centers around the emotional quality of the HMI. As shown in Table 14, it was found that both the overall scale mean and the means of the subscales usefulness and satisfaction were significantly higher ($p < 0.001$) than the center of the scale, 0, indicating that acceptance, usefulness, and satisfaction with the prototype were considered to be above average. These findings are very similar to the UEQ-S results, with the difference that satisfaction was rated more favorably on the satisfaction subscale of the Acceptance Scale than the pragmatic quality in the UEQ-S.

Table 14 Subjective acceptance of the prototypical remote operation workplace measured with the Acceptance Scale

Scale	M_{emp}	SD_{emp}	CI (LL) ^a	CI (UL) ^a	M_{crit} ^b	t	p
Overall	1.02	0.58	0.82	1.22	0	10.181	<0.001
Usefulness	0.98	0.60	0.77	1.19	0	9.565	<0.001
Satisfaction	1.06	0.58	0.86	1.26	0	10.639	<0.001

Van der Laan Acceptance Scale (Van der Laan et al. 1997); low: -2 to high: 2 (center: 0)

^aCI 95% confidence interval, *LL* lower limit, *UL* upper limit

^b M_{crit} = critical mean to test empirical mean (M_{emp})

4 Discussion

The goal of this study was to evaluate a novel prototypical workplace for the remote assistance of highly automated vehicles (HAVs) regarding performance, situation awareness (SA), workload, and other user-related outcomes. To the knowledge of the authors, this is the first study that does not only design a comprehensive HMI for the remote assistance of highly automated vehicles (Society of Automotive Engineers Level 4) but also systematically evaluates it in a controlled laboratory study, measuring outcome variables that are considered key in the field of Human Factors. Serving as representatives of the future user group of remote operators, participants resolved scenarios that were considered relevant routine tasks in HAV remote operation as listed by Kettwich et al. (2022) using the prototypical workplace's HMI. Furthermore, a secondary task was added to elevate the participants' workload, thus simulating the execution of parallel tasks or distractions. Three hypotheses were postulated regarding performance, SA, and workload of participants while resolving three scenarios with the workplace HMI.

4.1 Results on H1 (performance)

The first hypothesis assumed that participants showed lower performance at increasing levels of cognitive demand. This hypothesis was partially accepted: a significant main effect of secondary task condition, which was used as a proxy to systematically vary cognitive demand, was reached for task completion time but not for task reaction time. This finding suggests that induced cognitive load has a negative effect on *processing* a task in a timely manner (H1.2) but not on the response time that a participant needs to *react* to an incoming notification (H1.1). This finding can be explained with Wickens' (2002) multiple resource theory: the process of resolving the relatively simple primary task poses a considerable cognitive demand onto the operator, hence consuming a share of the pool of cognitive resources that is available to them. Cognitive demand induced by the secondary task draws from the same pool and therefore competes with the primary task's execution for cognitive resources. This diminishes the supply of cognitive resources for completing the primary task, resulting in a longer task completion

time. Since this effect occurs already in simple and well-trained routine tasks that participants were subjected to in this study, it is probable that the RO's workload will increase as tasks and interactions become more complex and novel. It is therefore questionable whether additional tasks beyond remotely supporting HAVs can be assigned to ROs. It can be concluded that in order to design the RO's workplace in a manner that does not create overload, the number and cognitive load of tasks that are to be executed simultaneously need to be kept at a minimum level. It seems advisable to limit the RO's responsibilities to tasks that are immediately related to providing remote assistance. Additional tasks, such as dispatching or passenger information beyond the level of driving maneuvers, might deplete the RO's cognitive resources and thus diminish safety and performance. A sophisticated system for task allocation and prioritization could help balance the RO's workload, particularly if the RO's work is embedded in a remote operation center in which tasks can be divided among ROs.

In contrast to completing a task, accepting a task is assumed to be cognitively less demanding. The additional demand posed by the secondary task does not deplete the pool of resources as strongly as processing it as it is a simple procedure that does not require abundant cognitive resources. Moreover, the pattern to accept the request is identical across tasks. Thus, instead of high-level cognitive mechanisms like working memory, more basal reaction times to visual stimuli might influence the results regarding task reaction time. They do not depend on a common pool of cognitive resources but are separate cognitive processes that might be explained as phenomena of attention distribution, e.g., by Wickens et al.'s (2003) SEEV model. According to this model, both bottom-up and top-down mechanisms of cognitive processing come into play when directing attention to stimuli. Specifically, the SEEV model claims that a stimulus's likelihood to attract attention is influenced by its *saliency*, the *effort* it takes to perceive it, the *expectancy* to perceive it, and the *value* the perceiver assigns to it. Furthermore, reaction time to a stimulus may be influenced by factors such as intensive training (Barrett et al. 2020). Consequentially, reacting to incoming notifications is not compromised by inducing more cognitive load, resulting in similar reaction times between the secondary task conditions.

Concerning the secondary task, it was hypothesized that the number of correct n-back comparisons decreases at increasing levels of cognitive demand (H1.3). This hypothesis was accepted, demonstrating the successful induction of cognitive load that diminished the participants' performances at the secondary task. The finding indicated that increasing the level of n does indeed result in a higher reported workload. The n-back task can therefore be considered valid at least as an approximation for additional work-related or unrelated tasks that occur in a real-world setting. An example for an additional task is being responsible for services other than core remote operation, such as communicating with passengers on board of the assisted HAV. This finding is in favor of splitting up core remote assistance tasks, such as ensuring the HAV's onward travel, from surrounding tasks, such as passenger communication, into different roles. Separate roles could help avoid cognitive overload in ROs in situations where several tasks would have to be executed simultaneously, ensuring safe and efficient operations.

The main effect of primary task on task completion time results from the finding that Scenario 1 took participants significantly less time to complete than Scenarios 2 and 3. This effect can be traced back to the differing lengths of the displayed scenarios and the kind of required interaction between participant and workplace HMI. For instance, in Scenario 1 only clearance had to be given, whereas in Scenarios 2 and 3 longer, multi-step interactions were required to resolve the task.

4.2 Results on H2 (situation awareness)

It was postulated in the second hypothesis that participants' ratings of situation awareness (SA) decrease at increasing levels of cognitive demand. This was also found in the collected data: the level of reported SA slightly deteriorated as cognitive load increased. The finding means that a higher cognitive load degrades subjective SA, even in routine and well-trained tasks in which participants could be expected to maintain SA due to learning effects. The finding has implications on the design of the interaction between RO and HAV: keeping the RO's workload at a manageable level may help them maintain a sufficient degree of SA. To ensure this, one way could be for the RO to focus their cognitive capabilities on resolving the HAV's request solely, without attending to other tasks, at least while processing the request. Second, the system for task management could be further improved to help generate and maintain SA by providing a status display on the requests that are to be processed and those that are currently being processed. The action that the RO is required to do next could be highlighted to improve the RO's overview of open tasks, boosting SA. Third, visual aids could be added to the HMI design, particularly to the video screens,

to draw the RO's attention to important stimuli such as relevant road users that are likely to interact with the supervised HAV. This could improve not only the perception of elements (SA Level 1) and their integration into a coherent situational representation (SA Level 2) but also the prediction of how the situation will unfold (SA Level 3).

The result that executing other tasks in parallel to a main task lowers SA is consistent with existing literature. Merat et al. (2010) found in a driving simulator study that SA, measured by participants' responses to critical incidents, was negatively affected by an auditory secondary task. In a similar vein, drivers who had to navigate the menu of a driver information system as a visual secondary task while driving in a simulator had a significantly lower SA (Wulf et al. 2013). Comparable findings that show an association between workload and SA were also made in other domains: in the aviation domain, Endsley and Rodgers (1998) found a significant relationship between workload and several indicators of SA.

Regardless of the condition of primary or secondary task, SA scores ranged in a medium to high area. They were significantly above the arithmetic center of the scale in all levels. Therefore, it can be assumed that the HMI design is robust against SA degradation even when additional workload is induced, regardless of the scenario—at least for scenarios similar to the ones investigated in this study, i.e., routine and rather well-trained scenarios. However, whether this will hold true in more complex, less trained scenarios is for future research to examine. Finally, it is noteworthy that no significant difference was found between the 1-back condition and the no secondary task condition for the SART score. This result gives rise to the notion that a floor effect showed in the 1-back condition, with this secondary task condition not impeding SA more strongly than the condition without any secondary task.

4.3 Results on H3 (workload)

The third hypothesis assumed that reported ratings of workload increase at rising levels of cognitive demand. The conducted RM-ANOVA confirmed this hypothesis. The finding indicates that the cognitive demand induced by a secondary n-back task fulfilled its purpose, actually increasing the perceived workload in participants. Thus, the task proved effective to reach the intended goal of artificially elevating workload to emulate working on multiple tasks simultaneously. Workload means did not transgress the center of the scale in any condition, ranging between 6 and 11 on a scale from 1 to 21. This finding suggests a low-to-medium perceived workload in all conditions of the secondary task. The simplicity and routine character of the primary task may have contributed to a low workload baseline. Nevertheless, effects

of the additional cognitive load induced via a secondary task on perceived workload showed statistical significance.

The identified effect of cognitive secondary task load is in line with research in automotive HMI research that used the n-back task as a secondary task to induce cognitive load. Liang and Pitts (2019), for instance, reported that workload measured by the NASA-TLX questionnaire was significantly elevated when the difficulty level of the n-back task was increased. The study measured participants' performance in a simulated driving task supported by a lane keeping system.

In addition to the significant main effect of secondary task, a significant main effect of primary task on reported workload was revealed. This finding is a logical consequence of the varying degree of complexity and demand inherent in the different scenarios, for instance, when participants only had to give clearance to a maneuver that the driving automation suggested as in Scenario 1 workload was low, while Scenarios 2 and 3 required more complex interactions, such as drawing waypoints and selecting a new route on a map and resulted in higher workload. Designing an HMI for the RO therefore needs to take into account how the respective tasks are structured and how much complexity and workload they entail in order to determine whether and which additional tasks may be assigned to the RO. Since even simple tasks with a moderate level of induced additional cognitive load led to increased perceived workload in this study, it needs to be carefully and critically analyzed whether more workload is tolerable. It is advisable to follow a conservative approach, entrusting the RO with a small task set initially and adding new tasks incrementally and cautiously under constant observation of their impact on safety and performance until an ideally balanced task load is established. However, routine tasks such as those investigated in this study do not seem to be overly mentally taxing to the RO. From this finding, it can be cautiously concluded that assisting several ROs might be possible as long no simultaneous interaction with several HAVs is required. This conclusion is backed by experience in the industry. For instance, Cruise requires a single remote assistant for about 15–20 HAVs (CNBC 2023).

4.4 Results on questionnaires

In addition to the variables that were directly linked to hypotheses, the three user-related outcome measures usability, user experience, and acceptance were collected to evaluate the workplace HMI overall. Intentionally, no hypotheses were specified beforehand as the goal was to capture the users' impressions with and experiences of the HMI in an exploratory fashion.

First, usability was found to be in a good to excellent range as measured with SUS. In a similar vein, both the usefulness dimension of the Acceptance Scale and the

pragmatic scale of UEQ-S showed results significantly above the arithmetic centers of the scales. Thus, all the indicators suggest a good level of usability. This is in line with the objectives of the designed prototype: the workplace HMI was supposed to implement basic functionalities and interactions between the workplace and the RO. These seem to be sufficient to yield positive ratings of usability. The results show that participants valued the approach to focus on the main features, keeping the HMI clear of an abundance of information that is not required for the tasks they had to execute. This finding underlines the need of a task-related HMI meaning that the HMI is limited to resolving an issue, which is the RO's task, and does not provide information unnecessary for task completion. Hence, even though a lot of data may be collected in the supervised HAV, surrounding vehicles and perhaps also in intelligent infrastructural units that is potentially accessible to the RO, one needs to constantly assess and reassess whether the information is actually instrumental for the specific task of the RO.

Second, user experience scales were administered to capture the emotional aspect of user interaction. Average assessments were given by participants on the hedonic subscale of UEQ-S but positive assessments on the satisfaction dimension of the Acceptance Scale. These more moderate-to-positive ratings relative to usability suggest that the emotional quality of the interaction is on a good way in the evaluated HMI but may be further refined. Consequently, after sufficient usability of the remote operation workplace has been demonstrated, future iterations of its HMI design should focus on improving user experience even more beyond the phase of the user's immediate interaction with it. However, the somewhat mixed results regarding user experience are comprehensible in the context of the development of the prototype, which did not have the priority to develop a system with outstanding user experience. The authors prioritized generating a prototype as a "proof of concept" to demonstrate its capability to process basic scenarios and tasks. Creating particularly positive experiences while interacting with the workplace is a quest for further refinement.

Third, the acceptance rating, operationalized as the overall Acceptance Scale, was above average, suggesting that participants could imagine to work with the prototypical workplace in general. This is an important finding as it demonstrates the willingness of participants to use the workplace HMI, a fundamental prerequisite for its deployment.

4.5 Limitations

Even though the approach taken in this study pursued ecological validity, it comes with five limitations.

First, the sample consisted mostly of male participants, with a percentage of 12 percent of participants reporting to be female. While this an objectively low figure, it can be

explained by the educational requirements posed to the role of the Technical Supervisor that informed the inclusion and exclusion criteria of this study. As only engineers in certain disciplines were eligible to participate, the gender balance in these fields of engineering need to be considered in order to compare the sample distribution with the relevant population. For example, among all 2020 Bachelor graduates in electrical engineering in Germany, one of the eligible fields for this study, only 12.5 percent were female (Kompetenzzentrum Technik-Diversity-Chancengleichheit 2023). Thus, the gender balance among study participants is close to the balance in the population of reference.

Second, the scenarios used in this study aimed at representing typical routine scenarios and the tasks related to resolving them from a RO's perspective. They are limited in number (only three scenarios were used throughout this study) and so is the range of tasks included. Inevitably, participants habituated to the tasks, rendering potential effects of novelty on performance improbable. However, the objective of the reported study was to evaluate the basic features and interaction patterns between RO and workplace. This was achieved by training and performing a set of standard tasks and scenarios deemed to be executed on a daily basis. Since acquiring new skills is based on general learning mechanisms, honing the skill to effectively assist HAVs with the evaluated workplace HMI is assumed to be gradual, hierarchical, and based on gaining experience: only after basic core tasks can be executed successfully, like those examined in this study, more complex and novel tasks may be feasible for trained RO. The authors thought it to be beneficial to pursue the same approach in user testing by initiating user studies with a limited set of routine scenarios and tasks and gradually extend this set. Relating to this, the particular HMI design used in this study influences the outcomes regarding performance, SA, workload, and other dependent variables. However, this is the case for any HMI evaluation. The examined HMI is based on a thorough user-centered design process has been partially evaluated in a previous usability test, yielding positive results. It also is the first HMI of a prototypical workplace for remote assistance to the knowledge of the authors. Thus, the authors deemed it justifiable to use this HMI for an extensive evaluation as a proxy for a typical workplace for HAV remote assistance.

Third, only a specific variant of HAV remote operation was investigated since the workplace HMI is tailored to it: remote assistance. Hence, the results reported in this study may not be directly transferable to other variants of remote operation, specifically, remote driving. It can be assumed, however, that the Human Factors investigated here, workload and SA, will be equally or even more critical in light of as-if real-time situational representations and immediate interventions, as required for remote driving. If remote driving becomes technically feasible and legally permissible,

assessing it from a Human Factors perspective and proposing HMIs for it will become more relevant. In a similar vein, combining different roles with divergent tasks within a remote operation center, as, for example, suggested by Schrank and Kettwich (2021), will require modifications of the evaluated workplace HMIs as tasks are likely to change and diversify.

Forth, the current legal situation in Germany demands that the Technical Supervisor's interaction with the supervised HAV be not time critical. It could therefore be argued that the outcome variables task reaction time and task completion time that both measure durations of RO HMI interactions are not an adequate measure to examine the workplace HMI's performance. However, even though this variable does not reflect the current legal status, the variables were deemed key performance indicators as efficiency is vital for systems to be economically viable. Only if handling an incident with an HAV via remote assistance is favorable time wise, a business case that involves this technology may emerge.

Fifth, it can be argued that the secondary task used in this study, the n-back task, is somewhat artificial as it does not occur in the natural environment of an RO. Additionally, it does not directly interfere with the primary task as different sensory modalities are used, implying less interference (Wickens 2002): while the visual modality is used for processing the primary task at the workplace HMI, the auditory modality is used for the secondary task. However, the n-back task is a reliable and commonly used method that ensures internal validity and enables systematic variation of induced cognitive load, enabling drawing steadfast conclusions about the effects of modulating cognitive load. Regarding different modalities, presenting the primary and secondary tasks on dissimilar modalities may actually be an advantage for determining the lower threshold of workload that is to be expected while conducting remote assistance tasks: since differences in workload were measured even in different modalities at rather low levels of induced cognitive load and simple routine tasks, it is to be expected, according to Wickens' theory, that cognitive load that is induced via a secondary task on the same modality may impede the performance at the primary task even more, particularly when complex and novel tasks come into play. The outcome observed here may therefore be considered a lower boundary of workload that can be further increased. What the upper boundary of workload may be is subject to future research. Following from these findings, imposing even more responsibilities onto a RO is likely not to be beneficial for their performance and ought to be considered with caution.

Nevertheless, in order to increase ecological validity on top or in lieu of internal validity, future research could use more natural secondary tasks. An example for such a task is supporting passengers on board of the assisted vehicle over intercom, including auditory interfaces, to provoke same-modality

interferences. This is argued to deplete ROs' cognitive resources even more, making it less likely they are still capable to resolve their tasks using the proposed workplace HMI.

4.6 Conclusions and future research

This study has shown that the presented user-centered HMI for a remote assistance workplace helps execute routine tasks by potential users—even when additional moderate cognitive load was induced via a secondary task. Thus, the remote assistance workplace HMI appears to be a feasible way to design the interaction between RO and HAV for supporting HAVs in routine scenarios, utilizing human cognitive skills. The proposed workplace HMI for remote operation proved capable of enabling a remote operator to provide support. This claim was supported by four central outcomes, relating to performance, SA, workload, and global evaluation measures.

First, even though induced cognitive load did have a significant impact on one of the performance indicators (task completion time but not task reaction time), perceived workload did not surpass a medium level. This finding indicated that at least for simple routine scenarios and related tasks, workload was maintained at a manageable level. It is for future research to examine whether the same holds true for more complex scenarios and interactions, particularly when they have not been encountered before.

Second, the degree of induced cognitive load had a negative main effect on the participants' perceived SA when processing routine tasks using the presented workplace HMI. That means that in this context, SA degrades as cognitive load increases. It must be noted, however, that the absolute differences among SA scores between conditions were moderate, with all scores ranging in a medium to upper level. Nevertheless, this finding bears significance as it gives rise to the expectation that in complex, hardly trained or entirely novel scenarios these effects may be more pronounced. Thus, higher cognitive load resulting from more challenging tasks is likely to impede SA even more. Indirectly, a diminished level of SA may affect performance negatively as the RO may make wrong assumptions about the current status of the traffic environment, which in turn might entail drawing wrong conclusions on how the situation will unfold. Future research may therefore shed light on the generation and maintenance of SA in more complex, novel scenarios. A systematic variation of scenario complexity may help elucidate how SA develops across varying levels of complexity.

Third, the main effect of the secondary task condition on perceived workload can be considered as passing the "manipulation check": inducing cognitive load indeed resulted in increasing perceived workload. Thus, the n-back task in the conditions 1-back and 2-back proved to have an impact on how much workload participants experienced. However, similar to SART scores, the differences between

conditions were generally low. In all conditions, workload means did not exceed the arithmetic center of the scale. The variance was therefore narrow. Hence, a takeaway for future research may be increasing the spectrum of the secondary task's difficulty levels to inquire the effects of high induced cognitive load on primary task performance.

Fourth, the global evaluation of the workplace HMI produced favorable results. This is particularly true regarding the variable that pertains to the direct interaction with the HMI, usability. On three questionnaires (SUS, UEQ-S, Acceptance Scale), this variable or related concepts consistently yielded above-average ratings. The hedonic-emotional quality of the interaction was assessed slightly more modest but still sufficient in an average-to-positive range. This finding aligns with the authors' objective to give priority to a proof of concept that works for basic interactions before focusing on user experience-related aspects. Finally, the HMI's acceptance ratings were positive. This result is meaningful because the sample was tech-savvy, receiving high scores on the Affinity for Technology Scale. The HMI lived up to the potentially higher expectations that may be posed by technologically experienced and invested participants. In addition, the sample fulfilled the educational requirements of the legally specified role of the Technical Supervisor. It is of utmost importance for the utilization of an HMI that the future group of users adopts it. The results in the Acceptance Scale provide at least hints for this claim. However, future iterations of the workplace HMI need to examine critically whether groups of people beyond the Technical Supervisor may also be capable of remotely assisting HAVs. Including a wider group of participants may thus be a goal for future research.

To summarize, this paper presented the evaluation of a workplace HMI for the remote assistance of HAVs by inducing a secondary task to systematically modulate cognitive load. The results show that in accordance with the hypotheses, cognitive load did have a negative impact on task completion time, perceived SA, and workload. This finding demonstrates the importance of considering cognitive load when designing an HMI for remote assistance. However, across all conditions, workload was low to medium, SA was medium to high, and the global evaluation of the workplace HMI regarding usability, user experience, and acceptance yielded favorable results as well. These results hint at the general suitability of the tested workplace HMI for fulfilling certain routine remote assistance tasks.

To conclude, when designing a workplace for the remote operator, bearing in mind the Human Factors involved and their interaction with the remote operation technology are essential to ensure safety and feasibility of the system overall. Only by applying user-centered methods, using workplaces for remote operation can become a successful approach to booster automated driving technologies and thus lay the groundwork for a more sustainable mobility.

Appendix (Table 15)

Table 15 Steps of interaction between remote operator (RO) and workplace HMI by scenario

Scenario	Steps
Scenario 1: Detected situation unclear	<ol style="list-style-type: none"> (1) A message appears on the disturbance ticker under “Incoming messages”: “Detection situation unclear” including more detailed information on this disturbance (1) The RO clicks on “Accept” and thus assigns himself the responsibility for this incident (2) The message disappears under “Incoming messages” and is now displayed under “Tasks” with the status “active” (3) The RO clicks on “Details” under “Tasks” and receives more detailed information about the task in a pop-up window. At the same time, the camera screens display the current video streams, the details screen displays the details of the corresponding AV, and the map screen displays an environment map with the AV centered (4) In the dialogue menu, the RO is asked the following question: “Can vehicle proceed?” (5) The RO clicks on the answer options “Yes, allow further travel” (6) A pop-up window appears for checking the prerequisites (7) The status of the task changes to “completed” (8) An according entry is made in the operating log
Scenario 2: Blocked lane	<ol style="list-style-type: none"> (1) The following entry appears on the fault ticker under “Incoming messages”: “Lane blocked” including more detailed information on this (2) The RO clicks on “Accept” and thus assigns himself the responsibility for this incident (3) The message disappears under “Incoming messages” and is now displayed under “Tasks” with the status “active” (4) The RO clicks on “Details” under “Tasks” and receives more detailed information about the task in a pop-up window. At the same time, the camera screens display the current video streams, the details screen displays the details of the corresponding AV, and the map screen displays a map of the environment with the AV centered on it (5) In a dialog menu, the RO is prompted: “Draw waypoints for bypass.” At the same time, a detailed map of the AV’s immediate surroundings appears on the touchscreen. Using the “Draw waypoints” tool, the RO now sets waypoints that the automation converts into a drivable alternative trajectory (6) In the dialogue menu, the RO is asked the following question: “Can the alternative pathway be used?” (7) The RO clicks on the answer options “Give clearance now” (8) The status of the task changes to “completed” (9) An according entry is made in the operating log
Scenario 3: Rerouting	<ol style="list-style-type: none"> (1) The following entry appears under “Incoming messages”: “Road closed” including more detailed information on this (2) The RO clicks on “Accept” and assigns themselves the responsibility for this incident (3) The message disappears under “Incoming messages” and is now displayed under “Tasks” with the status “active” (4) The RO clicks on “Details” under “Tasks” and receives more detailed information about the task in a pop-up window. At the same time, the camera screens show the current video streams, the details screen shows the details of the corresponding AV, and the map screen shows an environment map with the AV centered (5) A dialogue menu prompts the RO: “Select an alternative route to bypass from the suggested routes on the touchscreen” (6) At the same time, a detailed map of the AV’s surroundings appears on the touchscreen showing alternative routes avoiding the closed road. In a dialogue menu on the touchscreen, the RO is prompted: “Select an alternative route to bypass from the suggested routes.” In the details screen, the “Location” tab is automatically invoked. In the table of next stops, canceled stops are indicated with “CANCELLED: [name stop]” in red font. New stops are displayed with “NEW: [name stop]” in green font (7) In the dialogue menu, the RO is asked the following question on the touch screen: “Select an alternative route to bypass from the suggested routes.” (8) The RO clicks on the response options “Select route now” (9) At the same time, the location is displayed in the details screen and the stops to be served and the stops to be canceled are shown (10) The status of the task changes to “completed” (11) An according entry is made in the operating log

Acknowledgements The authors would like to express their gratitude to Hoai Phuong Nguyen, Florian Rudolph and Nils Wendorff for their steadfast support in providing the technical framework for this study, Thorben Brandt for his valuable comments to the manuscript, as well as student assistants Sarah Helweg and Fabian Reese for their assistance with the data collection.

Author contributions All authors contributed to the study conception and design. Material preparation, data collection, and analysis were

performed by AS and FW. The first draft of the manuscript was written by AS and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. This research received funding as part of the project GAIA-X 4 ROMS—Support and Remote Operation of Automated and Networked Mobility Services (FKZ: 19S21005C). The joint project is funded by the German

Federal Ministry of Economics and Climate Protection (BMWK). Solely the authors are responsible for the content of this article.

Data availability The dataset is sensitive due to protection of personal data and thus not openly available.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abbinck DA, Carlson T, Mulder M, de Winter JCF, Aminravan F, Gibo TL, Boer ER (2018) A topology of shared control systems—finding common ground in diversity. *IEEE Trans Human-Mach Syst* 48(5):509–525. <https://doi.org/10.1109/THMS.2018.2791570>
- Amador O, Aramrattana M, Vinel A (2022) A survey on remote operation of road vehicles. *IEEE Access* 10:130135–130154. <https://doi.org/10.1109/ACCESS.2022.3229168>
- Bangor A, Kortum PT, Miller JT (2008) An empirical evaluation of the system usability scale. *Int J Human-Comput Interact* 24(6):574–594. <https://doi.org/10.1080/10447310802205776>
- Barrett BT, Cruickshank AG, Flavell JC, Bennett SJ, Buckley JG, Harris JM, Scally AJ (2020) Faster visual reaction times in elite athletes are not linked to better gaze stability. *Sci Rep* 10(1):13216. <https://doi.org/10.1038/s41598-020-69975-z>
- Baumann M, Petzoldt T, Groenewoud C, Hogema K (2008) The effect of cognitive tasks on predicting events in traffic. In: C. Brusque (Ed) Proceedings of the European Conference on Human Interface Design for Intelligent Transport Systems. Humanist Publications. <https://www.humanist-vce.eu/fileadmin/contributeurs/humanist/Lyon2008/Proceedings-HUMANIST-S1.1.pdf>. Accessed on Sep 2023
- Bonferroni CE (1936) *Teoria statistica delle classi e calcolo delle probabilità*. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze.
- British Broadcast Corporation (2016) Germany train crash: Controller 'distracted by computer game'. <https://www.bbc.com/news/world-europe-36025951>. Accessed on Sep 2023
- Brooke J (1996) Sus: A "quick and dirty" usability scale. In P. W. Jordan (Ed.), Usability evaluation in industry: Based on the International Seminar Usability Evaluation in Industry that was held at Eindhoven, The Netherlands, on 14 and 15 September 1994]. Taylor & Francis. http://www.tbistafftraining.info/smartphones/documents/b5_during_the_trial_usability_scale_v1_09aug11.pdf. Accessed on Sep 2023
- CNBC. (2023). Cruise confirms robotaxis rely on human assistance every four to five miles. <https://www.cnbc.com/2023/11/06/cruise-confirms-robotaxis-rely-on-human-assistance-every-4-to-5-miles.html>. Accessed on Sep 2023
- Darken RP, Kempster K, Peterson B (2001) Effects of streaming video quality of service on spatial comprehension in a reconnaissance task. In Proceedings of IITSEC, Orlando, FL, USA.
- DriveU.auto (2023) Deploy robot and autonomous vehicle fleets with confidence, using teleoperation. <https://driveu.auto/>. Accessed on Sep 2023
- Endsley MR (1995) Toward a theory of situation awareness in dynamic systems. *Hum Factors* 37(1):32–64. <https://doi.org/10.1518/001872095779049543>
- Endsley MR, Rodgers MD (1998) Distribution of attention, situation awareness and workload in a passive air traffic control task: implications for operational errors and automation. *Air Traffic Control Quarterly* 6:21–44
- Endsley MR, Bolté B, Jones DG (2003) Designing for situation awareness: an approach to user-centered design. Taylor & Francis, New York
- Franke T, Attig C, Wessel D (2019) A personal resource for technology interaction: development and validation of the Affinity for Technology Interaction (ATI) Scale. *Int J Human-Comput Interact* 35(6):456–467. <https://doi.org/10.1080/10447318.2018.1456150>
- Epic Games (2019) Unreal engine. <https://www.unrealengine.com>. Accessed on Sep 2023
- Gesetz zur Änderung des Straßenverkehrsgesetzes und des Pflichtversicherungsgesetzes—Gesetz zum autonomen Fahren, Bundesgesetzblatt (2021). <https://www.gesetze-im-internet.de/stvg/BJNR004370909.html>. Accessed on Sep 2023
- Greenhouse SW, Geisser S (1959) On methods in the analysis of profile data. *Psychometrika* 24:95–112
- Guardian (2023) Cruise recalls all self-driving cars after grisly accident and California ban. <https://www.theguardian.com/technology/2023/nov/08/cruise-recall-self-driving-cars-gm>. Accessed on Sep 2023
- Hart SG, Staveland LE (1988) Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Hancock PA, Meshkati N (eds) Human mental workload. Elsevier, New York
- Hassenzähl M (2008) User experience (UX). In: É. Brangier (Ed) ACM Other conferences, Proceedings of the 20th International Conference of the Association Francophone d'Interaction Homme-Machine. ACM. pp. 11–15. <https://doi.org/10.1145/1512714.1512717>
- Herger M (2023) How Driverless Waymo and Cruise Robotaxis Deal With Double Parked Cars. The Last Driver License Holder. <https://thelastdriverlicenseholder.com/2023/03/17/how-drive-ress-waymo-and-cruise-robotaxis-deal-with-double-parked-cars/>. Accessed on Sep 2023
- Hollands JG, Spivak T, Kramkowski EW (2019) Cognitive load and situation awareness for soldiers: effects of message presentation rate and sensory modality. *Hum Factors* 61(5):763–773. <https://doi.org/10.1177/0018720819825803>
- Iclodean C, Cordos N, Varga BO (2020) Autonomous shuttle bus for public transportation: a review. *Energies* 13(11):2917. <https://doi.org/10.3390/en13112917>
- International Organization for Standardization. (2018). ISO standard no. 9421–11:2018. Ergonomics of human-system interaction—Part 11: usability: definitions and concepts.
- Kalisvaart, S. (2021). Introducing edge cases. TNO. https://www.connectedautomateddriving.eu/wp-content/uploads/2021/05/Kalisvaart_Introducing_Edge_Cases.pdf. Accessed on Sep 2023
- Kettwich C, Schrank A, Oehl M (2021) Teleoperation of highly automated vehicles in public transport: user-centered design of a human-machine interface for remote-operation and its expert usability evaluation. *Multimod Technol Interact* 5(5):26. <https://doi.org/10.3390/mti5050026>

- Kettwich C, Schrank A, Avsar H, Oehl M (2022) A helping human hand: relevant scenarios for the remote operation of highly automated vehicles in public transport. *Appl Sci* 12(9):4350. <https://doi.org/10.3390/app12094350>
- Kirchner WK (1958) Age differences in short-term retention of rapidly changing information. *J Exp Psychol* 55(4):352–358. <https://doi.org/10.1037/h0043688>
- Kompetenzzentrum Technik-Diversity-Chancengleichheit (2023) Datentool MINT-Bereich. <https://www.komm-mach-mint.de/service/datentool>. Accessed on Sep 2023
- Liang N, Pitts BJ (2019) The effect of secondary cognitive task difficulty on headway maintenance and perceived workload while using lane keeping systems. *Proc Hum Factors Ergonom Soc Ann Meeting* 63:2011–2015
- Litman TA (2020) Autonomous vehicle implementation predictions: implications for transport planning. Victoria Transport Policy Institute.
- Mauchly JW (1940) Significance test for sphericity of a normal S_n -variate distribution. *Ann Math Stat* 11(2):204–209. <https://doi.org/10.1214/aoms/1177731915>
- Merat N, Jamson A, Lai F, Carsten O (2010) Automated driving, secondary task performance and situation awareness. *Human factors: a system view of human, technology, and organization*. p 41–53. https://www.researchgate.net/publication/285976484_Automated_driving_secondary_task_performance_and_situation_awareness. Accessed on Sep 2023
- Minge M, Thüring M, Wagner I, Kuhr CV (2017) The meCUE Questionnaire: a modular tool for measuring user experience. In: Soares M, Falcão C, Ahram TZ (Eds) *Advances in Intelligent Systems and Computing Ser: v.486. Advances in Ergonomics Modeling, Usability & Special Populations: Proceedings of the AHFE 2016 International Conference on Ergonomics Modeling, Usability & Special Populations, July 27–31, 2016, Walt Disney World®, Florida, USA vol. 486*. p 115–128. Springer International Publishing, New York. https://doi.org/10.1007/978-3-319-41685-4_11
- New York Times (2023) Driverless Taxis Blocked Ambulance in Fatal Accident, San Francisco Fire Dept. Says. <https://www.nytimes.com/2023/09/02/technology/driverless-cars-cruise-san-francisco.html>. Accessed on Sep 2023
- Ogden GD, Levine JM, Eisner EJ (1979) Measurement of workload by secondary tasks. *Hum Factors* 21(5):529–548. <https://doi.org/10.1177/001872087902100502>
- Pfannmüller L, Walter M, Bengler K (2015) Lead me the right way?! The impact of position accuracy of augmented reality navigation arrows in a contact analog head-up display on driving performance, workload, and usability. In: G. Lindgard & D. Moore (Chairs), *Proceedings of the 19th Triennial Congress of the International Ergonomics Association (IEA)*.
- Reimer B, Mehler B (2011) The impact of cognitive workload on physiological arousal in young adult drivers: a field study and simulation validation. *Ergonomics* 54(10):932–942. <https://doi.org/10.1080/00140139.2011.604431>
- Schoitsch E (2016) Autonomous Vehicles and Automated Driving Status, Perspectives and Societal Impact. In *Schriftenreihe Informatik: vol. 45, IDIMT 2016 Proceedings*. p 405–423. Trauner Verlag. https://www.researchgate.net/profile/Erwin-Schoitsch/publication/306373756_AUTONOMOUS_VEHICLES_AND_AUTOMATED_DRIVING_STATUS_PERSPECTIVES_AND_SOCIETAL_IMPACT/links/57bb27b008ae48f0f444beb/AUTONOMOUS_VEHICLES-AND-AUTOMATED-DRIVING-STATUS-PERSPECTIVES-AND-SOCIETAL-IMPACT.pdf. Accessed on Sep 2023
- Schrank A, Kettwich C (2021) Roles in the teleoperation of highly automates vehicles in public transport. In *Humanist VCE (Chair), Proceedings of the 7th Humanist Conference, Rhodes Island, Greece*. https://www.humanistvce.eu/_files/archives/ace9ed_64d1f0cf8a164dd185a50e52238fa1a4.zip?dn=Rhodes-2021.zip. Accessed on Sep 2023
- Schrepp M, Hinderks A, Thomaschewski J (2017a) Construction of a BENCHMARK for the user Experience Questionnaire (UEQ). *International Journal of Interactive Multimedia and Artificial Intelligence* 4(4):40. <https://doi.org/10.9781/ijimai.2017.445>
- Schrepp M, Hinderks A, Thomaschewski J (2017b) Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S). *Int J Interact Multimed Artif Intell* 4(6):103. <https://doi.org/10.9781/ijimai.2017.09.001>
- Society of Automotive Engineers (2021) Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. (SAE J 3016–202104). Washington, D.C. SAE. https://www.sae.org/standards/content/j3016_202104. Accessed on Sep 2023
- Straßenverkehrsgesetz: Betrieb von Kraftfahrzeugen mit autonomer Fahrfunktion; Widerspruch und Anfechtungsklage (2021 & rev. 12.07.2021). <https://dejure.org/gesetze/StVG/1e.html>. Accessed on Sep 2023
- Sweller J (1988) Cognitive load during problem solving: effects on learning. *Cogn Sci* 12(2):257–285. https://doi.org/10.1207/s15516709cog1202_4
- Taylor RM (1990) Situation awareness rating technique (SART): the development of a tool for aircrew systems design (AGARD-CP-478). In: *Situational Awareness in Aerospace Operations*. NATO-AGARD, Neuilly Sur Seine, France. pp3/1 –3/17
- Tittle JS, Roesler A, Woods DD (2002) The Remote Perception Problem. *Proc Hum Factors Ergonom Soc Ann Meeting* 46(3):260–264. <https://doi.org/10.1177/154193120204600309>
- T-Systems (2023) Teleoperated driving: safely controlling cars remotely. <https://www.t-systems.com/de/en/industries/automotive/connected-mobility/teleoperated-driving>. Accessed on Sep 2023
- van der Laan J, Heino A, de Waard D (1997) A simple procedure for the assessment of acceptance of advanced transport telematics. *Transp Res Part C: Emerg Technol* 5(1):1–10
- Vanderhaegen F (2021) Heuristic-based method for conflict discovery of shared control between humans and autonomous systems—a driving automation case study. *Robot Auton Syst* 146:103867. <https://doi.org/10.1016/j.robot.2021.103867>
- Vay (2022) A huge step for us and for the future of teledriving: We are the first company in Europe to be allowed on the road without a safety driver [Press release]. <https://vay.io/a-huge-step-for-us-and-for-the-future-of-teledriving-we-are-the-first-company-in-europe-to-be-allowed-on-the-road-without-a-safety-driver/>. Accessed on Sep 2023
- Wickens CD (1984) Processing resources in attention. In: Parasuraman R, Davies R (eds) *Varieties of attention*. Academic Press, pp 63–101
- Wickens CD (2002) Multiple resources and performance prediction. *Theor Issues Ergon Sci* 3(2):159–177. <https://doi.org/10.1080/14639220210123806>
- Wickens CD, Goh J, Helleberg J, Horrey WJ, Talleur DA (2003) Attentional models of multitask pilot performance using advanced display technology. *Hum Factors* 45(3):360–380. <https://doi.org/10.1518/hfes.45.3.360.27250>
- Wu C, Wu H, Lyu N, Zheng M (2019) Take-over performance and safety analysis under different scenarios and secondary tasks in conditionally automated driving. *IEEE Access* 7:136924–136933. <https://doi.org/10.1109/ACCESS.2019.2914864>
- Wulf F, Zeeb K, Rimini-Doring M, Arnon M, Gauterin F (2013) Effects of human-machine interaction mechanisms on situation awareness in partly automated driving. *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*. 2012–2019.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.