POLITECNICO DI TORINO
Repository ISTITUZIONALE

CARLA-IVR: An Open CARLA Extension for Immersive AR-HUD Evaluation and Reproducible Experimental Design

Original
CARLA-IVR: An Open CARLA Extension for Immersive AR-HUD Evaluation and Reproducible Experimental Design / Vezzani, L., Strada, F., Bottino, A.. - In: IEEE OPEN JOURNAL OF VEHICULAR TECHNOLOGY. - ISSN 2644-1330. - ELETTRONICO. - (2026), pp. 1-16. [10.1109/ojvt.2026.3740837]

Availability:
This version is available at: 11583/3016913 since: 2026-10-23T09:46:09Z

Publisher:
IEEE

Published
DOI:10.1109/ojvt.2026.3740837

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

18 November 2026

This article has been accepted for publication in IEEE Open Journal of Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OJVT.2026.3740837

IEEE Open Journal of Vehicular Technology

Received XX Month, XXXX; revised XX Month, XXXX; accepted XX Month, XXXX; Date of publication XX Month, XXXX; date of current version XX Month, XXXX.

Digital Object Identifier 10.1109/OJVT.XXXX.XXXXXXX

CARLA-IVR: An Open CARLA Extension for Immersive AR-HUD Evaluation and Reproducible Experimental Design

Leonardo Vezzani *, Francesco Strada *, and Andrea Bottino *

[1]Department of Control and Computer Engineering, Politecnico di Torino, Turin, 10129, Italy

CORRESPONDING AUTHOR: Leonardo Vezzani (e-mail: leonardo.vezzani@polito.it). Francesco Strada (e-mail: francesco.strada@polito.it). Andrea Bottino (e-mail: andrea.bottino@polito.it).

ABSTRACT Augmented Reality Head-Up Displays (AR-HUDs) are increasingly studied as next-generation in-vehicle interfaces, yet findings remain difficult to reproduce and compare because evaluations rely on heterogeneous simulation environments and experimental protocols. A shared, openly reusable immersive infrastructure for controlled AR-HUD experimentation is still lacking. We present *CARLA-IVR*, an open extension of the CARLA simulator that adds stereoscopic XR rendering, embodied interaction with the vehicle, and a scene-referenced AR-HUD overlay layer while preserving interoperability with standard CARLA workflows. We technically characterize the closed-loop pipeline through a rendering throughput profile, an input-to-render latency budget, and a scene-population ramp that delimits the real-time operating envelope of our platform, without human participants. We then present the platform through an initial, literature-grounded experimental design comparing AR-HUD and head-down display (HDD) conditions across two simulator setups: an immersive configuration (IVR) and a desktop configuration (DVR). Objective and subjective measures were collected from 34 participants. AR-HUD-HDD differences were descriptively more pronounced under IVR for perceived workload and trajectory excursions, whereas DVR showed weaker differences. However, the available sample size did not provide statistical support for a reliable interaction between interface type (AR-HUD vs. HDD) and simulator setup (IVR vs. DVR), and the IVR and DVR setups differed in several respects beyond visual immersion. The user study is therefore interpreted as exploratory evidence and a reproducible evaluation case rather than as confirmatory evidence of AR-HUD efficacy. Regardless, CARLA-IVR provides an open, interoperable simulation foundation that supports reproducible AR-HUD evaluation and the development of more systematic, comparable immersive AR-HUD evaluation protocols.

INDEX TERMS Augmented Reality Head-Up Display, CARLA Simulator, Human-Machine Interface, Driver information systems, User interfaces, Virtual reality

I. Introduction

Augmented Reality Head-Up Displays (AR-HUDs) project virtual content into the driver's field of view, spatially aligning information with the roadway and surrounding environment [1]. By reducing gaze shifts and anchoring relevant cues to the driving scene, AR-HUDs can support situational awareness, navigation, and driving safety across different levels of vehicle automation [2]–[4]. As a result, they have become a central topic in the design and evaluation of next-generation in-vehicle human-machine interfaces.

Despite this growing interest, AR-HUD evaluation remains methodologically fragmented [4]. Prior studies rely on heterogeneous simulation environments and experimental protocols, ranging from conceptual evaluations based on images or videos to desktop and immersive virtual reality (IVR) driving simulators, and, more rarely, on-road evaluations. Even within simulator-based studies, experiments differ in simulation engines, display technologies, cockpit configurations, interaction devices, task design, and outcome measures [3], [4]. This heterogeneity limits reproducibility and weakens cross-study comparability, making it difficult

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Open Journal of Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OJVT.2026.3740837

to determine whether reported effects arise from AR-HUD design choices, experimental procedures, or characteristics of the simulation platform itself.

This limitation is especially relevant for AR-HUDs because they are spatial interfaces. Their evaluation depends not only on what information is displayed, but also on how visual cues are perceived in relation to the driving scene, including scene-referenced placement, apparent depth, and attentional guidance within the driver's field of view. Testing conditions that differ in display technology, stereoscopy, viewpoint control, field of view, motion cues, cockpit layout, or input hardware may therefore affect how AR-HUD interfaces are experienced and measured. This does not imply that one simulator configuration is inherently superior to another; rather, it makes the simulation setup itself a methodological variable that should be explicitly documented, controlled where possible, and considered when comparing AR-HUD results across studies.

A reusable and openly accessible simulation infrastructure is therefore needed to support more cumulative AR-HUD research. In other automotive research domains, open-source platforms such as CARLA [5] and SUMO [6] have enabled widely adopted baselines for simulation-based research [7], [8]. AR-HUD research, however, still lacks an openly reusable infrastructure that combines driver-in-the-loop vehicle control, low-latency stereoscopic rendering, embodied interaction, scene-referenced AR-HUD overlays, controlled scenario design, and reproducible human factors evaluation.

CARLA is a promising foundation for this purpose because of its open-source ecosystem, broad adoption, and modular client-server architecture. However, standard CARLA does not natively support immersive AR-HUD experimentation. Existing CARLA-based immersive systems demonstrate the feasibility of VR driving simulation [9], [10], but they were not designed to provide an interoperable, scene-referenced AR-HUD layer for controlled evaluation within standard CARLA workflows. The gap addressed in this paper is therefore the lack of an immersive CARLA extension that integrates spatially registered overlays, embodied interaction, and driver-in-the-loop vehicle control within a reusable experimental framework.

Beyond infrastructure, AR-HUD evaluation also lacks a shared methodological foundation. Recent reviews have highlighted substantial variability in representative driving tasks, event structures, subjective instruments, and objective performance measures used for AR-HUD assessment [3], [4]. As a result, even studies investigating similar interface concepts often remain difficult to compare systematically. Nevertheless, recurring elements in the literature can still provide a useful basis for defining controlled evaluation cases. In this work, we therefore draw on recurrent task families, interface comparisons, and measurement choices from prior AR-HUD studies to instantiate an initial, reproducible study design within the proposed simulation environment. This design is intended as a transparent starting point for comparable AR-HUD evaluation rather than a finalized benchmark or general evaluation standard.

To address these issues, this paper presents *CARLA-IVR*, an open immersive extension of CARLA for controlled AR-HUD experimentation. The extension preserves interoperability with standard CARLA workflows while adding stereoscopic XR rendering, embodied vehicle interaction, and a scene-referenced AR-HUD overlay layer. In addition to describing the architecture, we characterize the real-time rendering pipeline through a baseline throughput profile, an input-to-render latency measurement, and a load-scaling screening in which the vehicular and pedestrian population is varied to identify the limit sustainable by the system. We then demonstrate the platform through an initial literature-grounded human-subject study comparing AR-HUD and head-down display (HDD) interfaces across two simulator setups: an immersive virtual reality setup (IVR) and a desktop monitor-based simulator setup (DVR). The study is intended as an exploratory and reproducible evaluation case rather than as a definitive benchmark or confirmatory test of AR-HUD efficacy.

More specifically, this paper makes four contributions:

- We present **CARLA-IVR**, an open immersive extension of CARLA for AR-HUD experimentation that remains consistent with CARLA's modular architecture and interoperable with standard CARLA workflows.
- We provide a **technical characterization of the real-time IVR pipeline**, including (i) a baseline rendering profile of both study configurations, (ii) an input-to-render latency budget of the immersive path, and (iii) a stepwise scene-population ramp that identifies the operating performance within which the platform preserves real-time behavior.
- We introduce an initial **literature-grounded experimental design** that operationalizes recurrent task types, interface comparisons, and measurement choices from prior AR-HUD studies within a controlled and reproducible evaluation case.
- We demonstrate the platform through an **exploratory human-subject study** comparing AR-HUD and HDD interfaces across IVR and DVR setups. The results provide exploratory evidence that selected AR-HUD-HDD differences were descriptively more pronounced under IVR for perceived workload and trajectory excursions; however, the interaction between interface type and simulator setup was not statistically supported at the available sample size, and the two setups differed along several coupled dimensions beyond visual immersion.

II. Background

Recent reviews [3], [4] describe AR-HUD research as methodologically fragmented in both technical platforms and evaluation practices. Although many studies use driving simulation, no shared immersive infrastructure has emerged

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Open Journal of Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OJVT.2026.3740837

for controlled AR-HUD experimentation, and no common methodological basis has been established for comparing results across studies. This section summarizes these two issues to better frame the contribution of the present work.

A. Driving Simulation for AR-HUD Evaluation

AR-HUDs present scene-referenced information directly within the driver's field of view. Their evaluation therefore depends not only on interface content but also on how the testing environment preserves the spatial relationship between visual cues and the driving scene [1]. For this reason, differences in simulation setup are methodologically relevant in AR-HUD research rather than merely technical implementation choices.

Based on the review by Vezzani et al. [4], AR-HUD evaluation studies can be grouped into three broad methodological categories with distinct trade-offs. *Conceptual methods*, including video walkthroughs and structured interviews [11], [12], support low-cost early-stage design iteration but do not allow assessment of active driving behavior. *Real-world deployments* offer the highest ecological validity [13], [14] but are limited by safety constraints, instrumentation cost, and the difficulty of introducing controlled hazard events. *VR-based simulators* have therefore become the dominant evaluation modality, as they provide controllable and repeatable environments with programmable traffic, hazards, and scenario structures.

Within simulator-based studies, however, this heterogeneity manifests along three dimensions: **simulation engines**, **device configurations**, and **evaluation practices**. Regarding **simulation engines**, no dominant platform emerges from the reviewed literature. Unity [15] and SCANeR [16] are among the most frequently cited tools, while a substantial fraction of studies relies on proprietary or insufficiently documented systems [17], [18], which limits reproducibility and cross-laboratory interoperability [1]. For **device configurations**, studies range from projection-based mock-ups and CAVE-like systems [19], [20] to Head-Mounted Displays [21], [22] and hybrid solutions [13]. Across these categories, critical perceptual parameters such as field of view, stereoscopic baseline, and vehicle-motion cues are rarely reported in sufficient detail to support fair comparison. Finally, for **evaluation practices**, subjective instruments (e.g., NASA-TLX, SUS, UEQ) and objective behavioral measures (e.g., gaze dwell time, reaction time, lateral deviation) vary widely across studies, without reference to a shared protocol or agreed minimum measurement set [1], [23].

Despite this heterogeneity, one structural feature recurs across a substantial portion of the literature: AR cues are often implemented as scene-coupled graphical elements, such as bounding boxes, virtual carpets, and spatial arrows, whose interpretation depends on the relationship between the interface and the surrounding driving scene. This recurring design pattern suggests that the fidelity of AR-HUD evaluation depends not only on the interface itself but also on how the simulation environment preserves scene integration, depth relations, and spatial registration. In this sense, differences in simulation setup may influence how reliably AR-HUD effects can be observed and compared across studies.

B. CARLA as a Foundation for Immersive AR-HUD Studies

Among open-source driving simulators, CARLA [5] stands out as a promising foundation for AR-HUD research because it combines broad adoption, an open ecosystem, and a modular client–server architecture. In the standard configuration, the Unreal Engine (UE) server manages physics, world state, actor behavior, and sensor rendering, while external Python or C++ clients issue control commands and consume simulation data. This separation allows control logic and experimental workflows to be developed without modifying the simulator core, which is one reason for CARLA's widespread use in autonomous driving research [7], [8]. In contrast, other open-source alternatives such as AirSim [24] and LGSVL [25] are no longer actively maintained, limiting their value as sustainable shared infrastructures.

CARLA is also attractive from an authoring perspective. Its UE backend provides access to a mature real-time rendering environment, while Blueprint-based scripting lowers the barrier to implementing scenarios, interface logic, and experimental components for researchers who are not engine developers. These features make CARLA a strong candidate for interdisciplinary AR-HUD research, where simulation control, immersive rendering, and interface prototyping must coexist within the same environment.

At the same time, CARLA's standard architecture introduces a specific limitation for IVR integration. In typical CARLA usage, the visual information available to external clients is derived from server-side rendering outputs. This approach is unsuitable for a real-time stereoscopic XR pipeline on the client side, primarily due to latency and synchronization constraints, requiring the direct integration of the immersive support within the UE simulation environment rather than treated as a simple client-side add-on.

Previous efforts to extend CARLA toward immersive VR demonstrate both the feasibility and current limitations of this approach. DReyeVR [10] integrates low-latency IVR rendering and behavioral instrumentation inside a dedicated CARLA/UE4 stack, but remains tied to an older CARLA/UE version and does not provide a scene-referenced AR-HUD interface layer. Drive Right [9] similarly demonstrates immersive CARLA-based simulation, but focuses on trust and education scenarios rather than controlled AR-HUD evaluation, and does not provide an interoperable AR-HUD overlay architecture aligned with current OpenXR-based practice. More recent efforts [26] confirm the usefulness of immersive CARLA simulation without yet providing a reusable foundation explicitly tailored to AR-HUD studies.

Taken together, these observations suggest that CARLA is a strong candidate for immersive AR-HUD experimentation.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Open Journal of Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OJVT.2026.3740837

However, what is still missing is an immersive extension that preserves interoperability with standard CARLA workflows while supporting low-latency stereoscopic rendering, embodied interaction, and explicit control of AR-HUD interface elements. Section III addresses this gap by describing the architecture developed in this work.

C. Limits of Current AR-HUD Evaluation Practices

The lack of a shared methodological foundation makes it difficult to compare studies systematically, even when they investigate similar AR-HUD concepts. This limitation is particularly relevant in immersive simulation, where the evaluation environment itself may influence how interface effects appear. However, the reviewed literature reveals recurring patterns. Studies are most often conducted under SAE Level 0 conditions, in relatively simple road networks, and focus on a small set of recurring task families, such as wayfinding and safety-related driving tasks [4]. Although these recurring choices do not constitute a standardized protocol, they provide a useful basis for defining more controlled and comparable evaluation cases.

Rather than proposing a benchmark or a general evaluation standard, this work uses these recurring elements to establish an initial controlled study design within a shared immersive simulation environment. Section IV describes this design in detail.

III. Methodology

This section describes the proposed immersive CARLA extension, its AR-HUD integration layer, and the current implementation boundaries of the system[1].

A. Open Immersive CARLA Extension

The proposed immersive extension was designed to preserve CARLA's modular workflow while adding immersive rendering, AR-HUD overlays, and embodied vehicle interaction without disruptive changes to the simulator core. It also had to remain compatible with heterogeneous XR hardware and with standard CARLA clients used for scenario control, vehicle command generation, agent logic, data acquisition, and AR-HUD cue generation.

CARLA is organized as a client–server simulator: the UE server manages the virtual world, physics, actor state, and sensor rendering, while external Python/C++ CARLA clients communicate through the TCP/RPC interface to issue commands and consume simulation data (Fig. 1). This architecture supports flexible experimental workflows but also limits immersive integration. In standard CARLA usage, visual information available to external clients is typically obtained from server-rendered streams, which are not suitable for a native, real-time stereoscopic XR pipeline. Immersive AR-HUD support therefore cannot be implemented as a simple external client-side visualization add-on and must instead be integrated within the UE simulation environment.

[1]Publicly available https://github.com/CGVGroup/Carla_ARHUD_VR

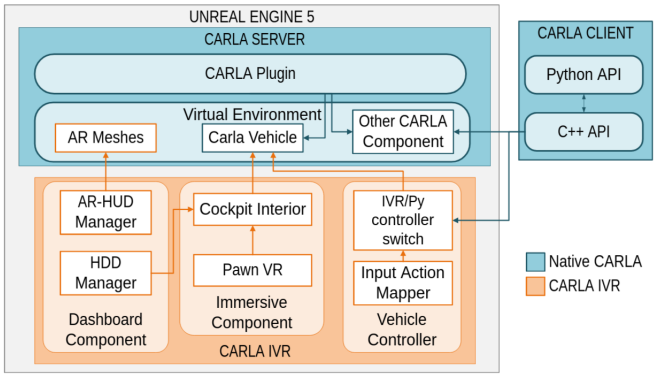


FIGURE 1: Functional architecture of the CARLA-IVR extension. blue: native CARLA; orange: CARLA-IVR; faded boxes group CARLA-IVR modules by function. Arrows denote either attachment to a parent entity or command and state flow, as detailed in the text. All CARLA-IVR modules run inside the Unreal Engine instance hosting the CARLA server, whereas standard Python/C++ clients remain external over TCP/RPC.

Accordingly, CARLA-IVR is implemented as a set of UE-side runtime modules rather than as a standard external CARLA RPC client. The extension does not restructure the CARLA core or replace the conventional CARLA client–server workflow. Instead, it adds dedicated modules within the UE simulation environment: an *Immersive Component* for embodied stereoscopic rendering, a *Dashboard Component* for AR-HUD and HDD display management, and a *Vehicle Controller* for command arbitration. These modules access simulation state through the UE update loop and operate within the same simulation environment, while standard CARLA Python/C++ clients remain external and continue to communicate through the conventional TCP/RPC interface. This placement keeps immersive rendering, AR-HUD overlays, and vehicle actuation coupled to the same simulation state and coordinate frame while preserving interoperability with existing CARLA workflows and concurrent CARLA-side logic.

Figure 1 details this organization. Native CARLA modules (blue) comprise the CARLA plugin, the simulated environment with its actors (traffic vehicles, pedestrians, sensors), and the external Python/C++ client stack while CARLA-IVR modules (orange) are instantiated inside the same environment.

The extension does not replace the CARLA ego vehicle actor. Instead, a *Cockpit Interior* entity is instantiated at runtime as a child of the vehicle and bound to its reference frame. The user viewpoint (*Pawn VR*), the *AR-HUD Manager*, and the *HDD Manager* are attached to this cockpit entity, so vehicle motion, head-coupled rendering, and in-vehicle display content are updated within the same UE coordinate hierarchy. This avoids asynchronous state transfer between the vehicle, cockpit, and display elements. AR overlays are rendered within the same UE scene and depth-

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Open Journal of Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OJVT.2026.3740837

tested against the surrounding geometry, allowing occlusion relationships to be handled by the rendering pipeline.

The *Immersive Component* manages the user's embodied viewpoint inside the simulated vehicle, including stereoscopic rendering, head-coupled viewpoint updates, spatial audio, and, when available, hand-tracking-based interaction. To decouple the implementation from vendor-specific XR stacks, it uses OpenXR as its runtime abstraction interface.

Vehicle actuation is handled through the *Vehicle Controller*, which keeps vehicle control separate from immersive perception. It integrates commands from physical steering wheels and pedals, normalized by the *Input Action Mapper*, with commands issued by conventional CARLA-side algorithms over the RPC interface. The *IVR/Py controller switch* resolves a single command at each update and applies it to the CARLA vehicle actor. This separation supports manual driving and mixed-control configurations without restructuring the immersive rendering module.

The *Dashboard Component* manages in-vehicle information display through two channels. The *AR-HUD Manager* controls the AR overlay by updating the geometry, position, and visual appearance of AR elements rendered ahead of the driver. These elements can be configured as visual cues anchored to the driving scene or as screen-fixed cues, and were used in this study for navigation guidance and hazard warnings. The *HDD Manager* controls conventional dashboard displays placed on cockpit surfaces. Both consume simulation state from the UE update loop and can remain interoperable with standard CARLA-side logic when required. In particular, the AR-HUD Manager derives cue geometry directly from the ground-truth state of the relevant actors (e.g., the 3D position of pedestrians, vehicles, and traffic elements) rather than from a simulated sensing or perception pipeline; this reflects the platform's focus on evaluating AR-HUD interfaces and cue designs under controlled conditions, rather than on sensor-fusion fidelity, which is outside its objective.

B. AR-HUD Interface Integration

Standard CARLA does not provide vehicle interiors or dashboard elements suitable for controlled AR-HUD studies. To address these limitations, we developed dedicated interface components that allow dashboard panels and AR-HUD overlays to be instantiated directly within the UE environment without modifying the CARLA core.

The interface layer supports two complementary representation modes. The first uses 2D widgets placed in the vehicle's cockpit for conventional dashboard elements such as instrument cluster displays, implemented through the Unreal Motion Graphics framework (Fig.2 *bottom*). The second supports scene-referenced AR cues, whose placement is defined relative to the 3D geometry of the simulated environment rather than only in screen space (Fig.2 *top*). This allows AR elements to remain spatially registered to the driving scene and maintain consistent perspective and scale as the viewpoint changes, while also supporting occlusion relationships with scene geometry when required by the cue design. All interface components are updated synchronously with the UE rendering loop.

C. Current Implementation Boundaries

The current implementation was developed during the transition to CARLA 0.10.0 and UE 5.5, which introduced practical constraints for immersive stereoscopic deployment. Initial IVR testing showed that existing CARLA maps and lighting configurations, originally designed for monoscopic desktop rendering, produced visible artifacts under stereoscopic conditions, including aliasing, graininess, and shadow instability. These artifacts become perceptually salient in the headset because they are view-dependent and temporally unstable across frames, conditions that are masked under static monoscopic desktop viewing.

Therefore, stereoscopic rendering quality was treated as a primary criterion in selecting the experimental environment. The present study relied on a dedicated UE5 virtual environment originally developed for technical demonstration purposes. By leveraging the Nanite rendering and adapting rendering settings for IVR compatibility, this environment provided improved frame stability, reduced aliasing, and more consistent lighting under stereo conditions. Minor shadow flickering remained in isolated situations, but it did not interfere with the experimental tasks considered in this work.

Finally, the present instantiation deploys the CARLA-IVR modules within a single engine instance co-located with the CARLA server, which is the configuration described in Section V-A. Because the extension is built on UE's native networking model, detached and multi-client deployments are architecturally admissible; their validation, however, falls outside the scope of this work.

IV. Experimental Study

This section reports the first human-subject use of the proposed immersive CARLA extension, together with a literature-grounded experimental design developed for this purpose. The study is structured around two factors: the *simulator setup* and the *interface*. The simulator setup factor compares an immersive configuration (IVR) with a desktop simulator configuration (DVR), while the interface compares AR-HUD and head-down display (HDD) presentation. The goal is not to provide a definitive validation of the extension or the experimental design, nor to isolate the effect of immersion as a single causal factor. Rather, the study aims to show that the proposed system can support a controlled AR-HUD evaluation, that recurring methodological elements from prior AR-HUD research can be incorporated into a transparent and reproducible study structure, and that AR-HUD-HDD outcomes can be examined across two simulator setups within the same experimental protocol. Since these setups differ in display technology, viewpoint control, motion

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Open Journal of Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OJVT.2026.3740837

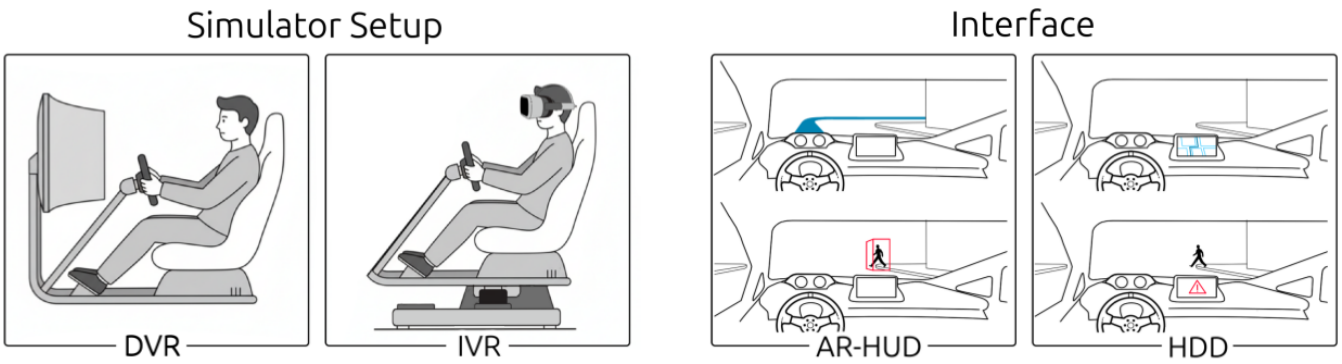


FIGURE 2: Experimental apparatus, defining the two testing dimensions. *Left - simulator setup:* the DVR desktop setup *(left)* and the IVR head-mounted-display setup with force-feedback actuation *(right)*. *Right - interface:* the AR-HUD *(left column)* and HDD *(right column)* interfaces, each shown with wayfinding cues in blue *(top)* and hazard warnings in red *(bottom)*.

feedback, cockpit embodiment, and steering hardware, the IVR-DVR comparison is treated as a setup-level comparison rather than an isolated manipulation of visual immersion.

A. Study Design Overview

The study crossed two factors: *simulator setup* and *interface*. The simulator setup factor compared two complete display-and-interaction configurations: a desktop virtual reality setup (DVR) and an immersive virtual reality setup (IVR) (Fig. 2). The interface factor compared AR-HUD and HDD presentation (Fig. 2). This yielded four experimental conditions: IVR-AR-HUD, IVR-HDD, DVR-AR-HUD, and DVR-HDD.

The IVR setup used a Meta Quest 3 head-mounted display, an Atomic 3 force-feedback motion seat that provided inertial cues during acceleration, braking, and steering, and a Fanatec direct-drive steering wheel and pedal set. The motion seat was included to reduce vestibular-visual conflict, a documented contributor to cybersickness in HMD-based driving simulation without inertial feedback [27]. The DVR setup used a static driving rig with a fixed seat, a Thrustmaster steering wheel and pedal set, and a three-monitor panoramic display providing an approximately 170° horizontal field of view.

The IVR-DVR comparison was included in order to evaluate the same AR-HUD protocol across two simulator configurations rather than within a single one.

B. Literature-Grounded Experimental Design

The experimental design was derived from a systematic review of prior AR-HUD studies [4], selecting studies meeting two criteria: (i) use of simulator-based AR-HUD evaluation, including both IVR setups (such as head-mounted displays and CAVE systems) and desktop simulator configurations; and (ii) implementation of AR-based HUD interfaces.

This process yielded a corpus of 46 studies (Fig. 3), concentrating predominantly on SAE Level 0 across three recurrent task families: *wayfinding* (12 studies), *longitudinal safety* (11), and *non-driving-related tasks* (11). Accordingly, the present study focuses on SAE Level 0 *wayfinding* and *longitudinal safety* tasks, which are most directly tied to scene-referenced perception and hazard response — making them particularly suitable for comparing AR-HUD and HDD conditions across simulation configurations. Non-Driving Related Tasks (NDRTs) were excluded as they emphasize interface-internal interaction procedures that would introduce variability orthogonal to the main comparison.

1) Itinerary Design

The itinerary design was informed by recurring characteristics identified in the reviewed literature, where driving sessions are typically brief, involve limited or no surrounding traffic, and take place in relatively simple urban or peri-urban environments. Consistent with these patterns, we designed two experimental routes with a total length of 1.6 km each (Fig. 4). Two routes were used to support within-subject counterbalancing of the interface: each participant completed one route under the AR-HUD condition and the other under the HDD condition, with the route-interface pairing alternated across participants to prevent learning effects from repeated exposure to the same path.

The two routes were designed to impose comparable navigational demands while remaining challenging enough to reveal interface-related differences in navigation performance. Each route includes the same number of left and right turns and one intentionally ambiguous intersection placed approximately at the midpoint, creating a controlled opportunity for navigational errors to emerge under both AR-HUD and HDD conditions.

2) Scenario and Interface Design

Wayfinding constituted the primary driver task. In the AR-HUD condition, guidance was provided through a virtual carpet [28] (Fig.2) spatially anchored to the road surface and updated turn by turn. In the HDD condition, equivalent navigational information was presented through a conventional dashboard-mounted navigation display (Fig.2). This mapping enabled us to compare AR-HUD and HDD guidance under

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Open Journal of Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OJVT.2026.3740837 IEEE Open Journal of Vehicular Technology

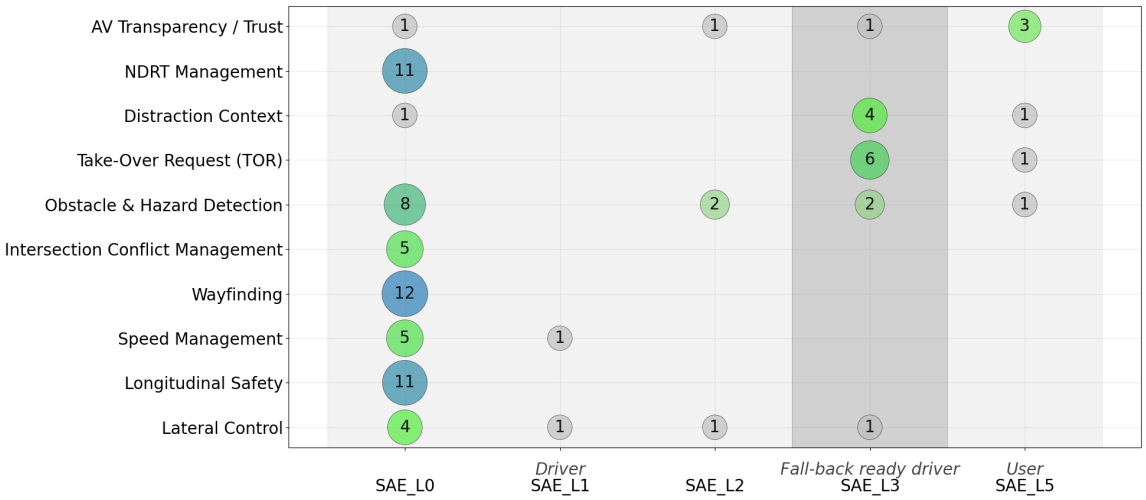


FIGURE 3: The 46 studies selected, sorted by SAE level of automation and task execution. On the X axis SAE levels and the corresponding Driver-User Continuum role [4] are reported; on the Y axis, recurring studied driving tasks. Each cell reports the number of studies at that SAE level-task intersection.

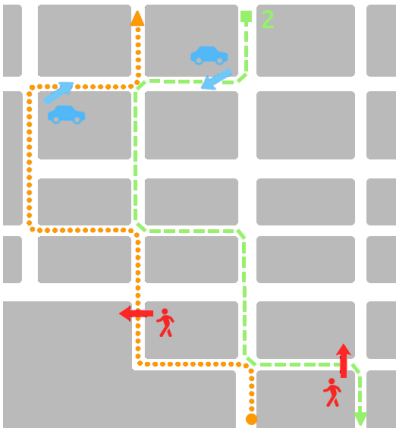


FIGURE 4: Experimental routes used for the wayfinding and safety tasks. Dashed lines indicate the route driven by participants. Pedestrian crossing from the sidewalk events are shown in red, and Vehicle cut-in from a parking spot events in blue.

the same itinerary while evaluating turn compliance, route-following accuracy, and deviations from the intended path, particularly at the ambiguous intersection.

Event-based safety support was implemented through localized hazard events placed at predetermined positions along the routes (Fig. 4). Specifically, we included *pedestrian-crossing events*, in which a pedestrian enters the roadway from a partially occluded position, and *vehicle cut-in events*, in which a parked vehicle abruptly merges into the participant's lane. These events were selected because they recur frequently in the reviewed literature and can be implemented in a controlled and reproducible manner within CARLA. In the AR-HUD condition, warnings were displayed as spatially anchored visual cues triggered at event onset: 3D bounding boxes with red edges highlighted the presence of the obstacle on the road (Fig. 2, bottom-left). The corresponding HDD condition conveyed equivalent warning information through a warning icon displayed on the dashboard (Fig. 2, bottom-right).

3) Measurements

Objective and subjective measures were selected based on two criteria: recurrence in the reviewed AR-HUD literature and suitability for the present controlled comparison. Timing of administration for each instrument is reported in Section D. Custom questionnaires were deliberately avoided to maintain comparability with prior work.

a: Objective measures.

Objective measures were selected on the basis of their direct applicability to the wayfinding and safety tasks under evaluation. We analyzed **trajectory consistency** using mean lateral deviation from the intended route, together with route excursion count and divergence percentage. These measures belong to the position-based vehicle metrics category and provide a frame-level characterization of route-following behavior, offering finer spatial resolution than the aggregated or event-based outcomes typically reported in prior AR-HUD studies. We measured also **reaction times** for both pedestrian-crossing and vehicle cut-in events, and used to quantify participants' temporal responsiveness to safety-critical situations, corresponding to the reaction-time category of the reviewed literature. Finally, we registered the number of **hit events** resulting in collision, included as a discrete safety-oriented outcome complementing reaction time analysis. This measure is complementary to reaction time: whereas reaction time captures the *speed* of the hazard response, hit events capture its *outcome*, and together they provide a more complete characterization of safety behavior than either measure alone.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Open Journal of Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OJVT.2026.3740837



FIGURE 5: Screenshots from the simulator. AR-HUD showing the VC *(top)*; HDD wayfinding device *(bottom)*.

b: Subjective measures.

NASA Task Load Index (NASA-TLX) [29] was selected to assess perceived workload, as it is the most frequently adopted workload instrument in the reviewed AR-HUD literature and has broad validation in driving-related contexts. **User Experience Questionnaire (UEQ)** [30] was selected to assess user experience across interfaces; it was preferred over narrower usability-oriented instruments because it captures both pragmatic and hedonic dimensions - including novelty and stimulation - which are particularly relevant to AR-HUD evaluation. **Simulator Sickness Questionnaire (SSQ)** [31] was included to quantify simulator-induced sickness: although sickness measures are rarely reported in the reviewed AR-HUD literature, SSQ administration is methodologically necessary whenever head-mounted displays are employed, both to enable screening of participants who experience excessive discomfort and to verify that performance differences between simulator setups are not confounded by differential sickness levels. **I-Group Presence Questionnaire (IPQ)** [32] was included to assess perceived presence: if the IVR and DVR conditions do not produce meaningfully different levels of perceived presence, the simulator setup factor loses its primary theoretical justification and any observed AR-HUD $\times$ simulator setup interactions become uninterpretable. IPQ therefore serves as a manipulation check for the between-subject simulator setup factor. Preference was given throughout to standardized instruments with established use in driving- and VR-related research.

C. Interface Implementation

Both interfaces were implemented using the widget architecture described in Section B, instantiating the design choices specified in Section 1. In the AR-HUD condition, the virtual carpet was realized as a scene-referenced UMG widget anchored to the road surface geometry and updated turn-by-turn; safety warnings were rendered as world-space bounding boxes with red outlines drawn around dangerous obstacles (Fig. 5,top). In the HDD condition, a 10-inch dashboard-integrated 2D UMG panel delivered both wayfinding arrows and warning icons, replicating conventional in-vehicle presentation conventions (Fig. 5, bottom). Both interfaces were implemented within the same extensible widget architecture, providing a controlled instantiation of the interface layer for the present study.

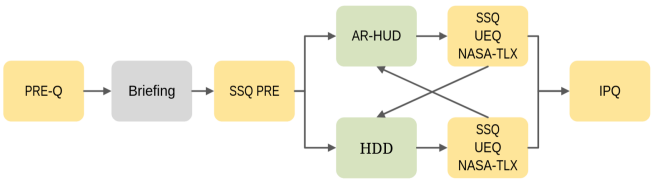


FIGURE 6: Experimental protocol. Colour encodes the block type: self-report instruments in orange, experimenter-delivered procedures in grey, and driving sessions green.

D. Procedure

The experimental protocol, illustrated in Fig. 6, began with a standardized briefing, after which participants signed the informed consent form. A demographic questionnaire covering age, gender, driving experience, familiarity with video games and 3D software, and prior exposure to VR was then administered, followed by a baseline SSQ.

Each participant was assigned to a single simulator setup (IVR or DVR), which remained constant throughout the experiment. Within that configuration, each participant completed two sessions corresponding to the two routes described in Section 1: one session under the AR-HUD interface and one under the HDD interface. The assignment of interface to route was counterbalanced across participants, to minimize order and learning effects attributable to route familiarity.

During each session, the system continuously recorded vehicle position, velocity, steering angle, throttle input, and braking input. After each session, participants completed the post-session NASA-TLX, UEQ, and SSQ. After both sessions, they completed the IPQ once to assess perceived presence under the assigned simulator setup.

E. Statistical Analysis Strategy

The analysis addresses the within-subject effect of interface (AR-HUD vs. HDD) and the between-subject effect of simulator setup (IVR vs. DVR). Normality of each continuous measure was assessed with the Shapiro-Wilk test ($\alpha = 0.05$); parametric procedures were applied when the assumption was satisfied and non-parametric alternatives otherwise.

Mixed-design ANOVAs were computed on NASA-TLX and UEQ subscale and all objective metrics under analysis, with partial eta squared (η_p^2) as effect size; post hoc pairwise comparisons used t-tests with Holm correction applied separately within each instrument family, and effect sizes are re-

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Open Journal of Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OJVT.2026.3740837

ported as Hedges' g. Excursion-based trajectory metrics were compared between interfaces within each simulator setup using Wilcoxon signed-rank tests, without a corresponding between-setup test; the same within-setup procedure was applied to the steering contrasts. SSQ sub-scale and total scores and IPQ sub-scale scores were compared with Mann-Whitney U tests (two-tailed, except one-tailed, IVR $>$ DVR, for IPQ), as Shapiro-Wilk violations were observed in one or more groups. Effect sizes for non-parametric tests are reported as the rank-biserial correlation r_{rb}. Categorical outcomes-hit events and avoidance-strategy frequencies-were analysed with Fisher's exact test on pooled event-level data, given the low event frequency per participant (two hazard events per route); the participant-level braking rate was compared between setups with a Mann-Whitney U test.

IVR and DVR contrasts are planned comparisons on distinct participant groups and are therefore reported uncorrected. Sessions without a detectable avoidance response were excluded listwise from the reaction-time analysis, and hazard-response latency was analysed for the pedestrian-crossing event only, as the cut-in geometry did not permit a reliable latency estimate.

V. Results

This section presents the outcomes of the controlled study described in Section IV. We first describe system performance and sample composition, as platform stability and sample characteristics are relevant for interpreting the subsequent IVR-DVR and AR-HUD-HDD comparisons. We then present the subjective and objective results of the study.

A. Platform Performance

The performance assessment was threefold: rendering throughput in the two study configurations, input-to-render responsiveness of the immersive configuration, and behaviour under increasing scene population in the same configuration. All measurements were taken on the same reference machine (Intel Core i9-13900K, 64 GB RAM, NVIDIA RTX 4080 Super).

1) Baseline rendering

Baseline throughput was measured on the experimental scenario described in Sec. A-1. Frame times are reported in Table 1. The IVR configuration held a stable framerate at the headset's native refresh rate, with a negligible fraction of frames below 60 fps. DVR throughput alone was left uncapped and ran far above the display refresh rate; the isolated scheduling anomalies visible in its $0.1\,\%$ low were not perceptible, as the monitors present at a fixed 144 Hz. Both configurations therefore render every presented frame with margin, sustaining real-time behaviour at their respective level of immersion.

2) Input-to-render latency

Input-to-render delay was measured as well, defined as the interval from an input event on the game thread to render-thread completion of the corresponding frame. To test the immersive pipeline alone, input-to-render was measured in an empty simulation.

TABLE 1: System performance indicators for the DVR and IVR configurations, measured over the full experimental scenario.

Metric	DVR	IVR
Target refresh rate [Hz]	144	72
Mean frame rate [fps]	416.5	68.4
P95 frame rate [fps]	474.0	70.7
1 % Low Average [fps]	88.8	57.3
0.1 % Low Average [fps]	12.9	51.9
Frame time <60 fps [%]	0.5	1.0

This quantity is reported for the IVR configuration alone, as it is the configuration subject to a hard perceptual constraint: given the head-coupled viewpoint the delay between motion and the corresponding visual update is bounded by the inter-frame budget of the headset and is a documented contributor to cybersickness [27]. Additionally, the DVR configuration is the standard CARLA display configuration, unmodified by the proposed extension, requiring therefore no novel assessment.

As it excludes GPU execution and frame transmission to the panel, this metric is a strict lower bound on end-to-end motion-to-photon delay. The resulting distribution is narrow, with a mean of 7.62 ± 0.70 ms and a median (P50) of 7.73 ms; the tail remains equally contained, at 8.59 ms (P95) and 9.43 ms (P99). All these values stay within the inter-frame budget $\Delta t = 13.9$ ms of the target 72 Hz HMD, leaving headroom before presentation: even at the 99th percentile roughly one third of the frame interval is still available. The narrow spread indicates that responsiveness is governed by the engine's render-thread decoupling rather than by scene-dependent render cost. Crucially, this sub-frame latency budget keeps the pipeline well within the perceptual tolerances required for immersive, fast-paced driving scenarios.

3) Scene population and the real-time operating envelope

The two measurements above describe the platform under the load of the present study, which by design involves little surrounding traffic (Sec. IV). To assess the re-usability of the proposed solution in denser scenarios, we measured how far scene population can be pushed before real-time behaviour degrades.

Seven traffic conditions were tested in a dedicated test. In every $90\,$s run, a set of vehicles and pedestrian agents was spawned along the route ahead of the ego vehicle driven in autopilot, so that the camera would actually encounter them. The upper bound of 135 agents was set empirically: below it the requested population was always fully spawned, whereas above it spawning became unreliable and the traffic manager produced congestion that stalled the simulation.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Open Journal of Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OJVT.2026.3740837

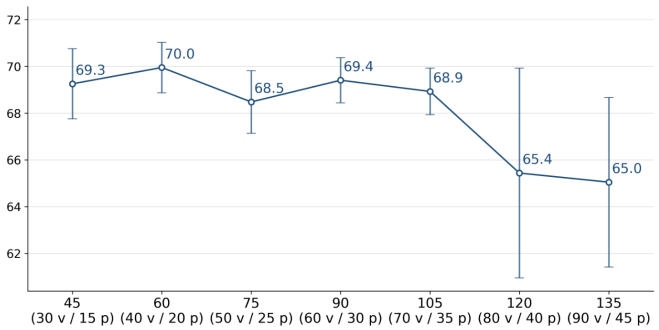


FIGURE 7: Mean frame rate as a function of scene population. Each point corresponds to one of seven independent runs, labelled with the total agent count and its vehicle/pedestrian composition; consecutive points differ by 15 agents. Error bars give ± 1 standard deviation of the per-second frame rate within the run. The HMD refresh rate is 72 Hz.

The frame rate (Fig. 7) is stable up to 105 agents: the per-level mean remains in the 68.5-70.0 fps range, i.e. within 5 % of the 72 Hz reference, with a narrow dispersion ($\sigma \leq 1.5$ fps). Over this population the platform reaches its limit: at 120 agents the mean drops to 65.4 fps and the distribution widens abruptly ($\sigma \approx 4.5$ fps, lower tail below 61 fps), with no recovery at 135 agents (65.0 fps, $\sigma \approx 3.6$ fps). The binding degradation is therefore not the mean throughput, which stays above the 60 fps target, but the widening dispersion: the frame-time spikes it reflects are perceptually salient in a head-mounted display even when infrequent. We therefore take 105 agents as a conservative operating point for IVR applications.

Hardware telemetry recorded alongside the runs indicates that the limit is not in the rendering path: the GPU never saturates (85.5 % on average), while CPU work fills the frame interval at every load level. Therefore, the threshold reflects the cost of agent simulation in CARLA (i.e., actor updates, physics, and world-snapshot exchange) rather than a limit of the IVR rendering pipeline, and below it the immersive condition remains stable.

B. Participants

The study included 34 participants: DVR ($n = 16$) and IVR ($n = 18$). The sample was predominantly male (24 M / 10 F), with a more balanced ratio in IVR (11 M / 7 F) than in DVR (13 M / 3 F). Self-report items were scored on a five-point Likert scale (0-4); age and driving experience are in years.

Shapiro-Wilk tests indicated that most variables were not normally distributed ($p < 0.05$), with generally right-skewed distributions and one high-end outlier in DVR (a 50-year-old with 27 years of experience). Descriptive statistics are therefore reported per group as the mean with the observed min-max range (Table 2), and group comparisons use non-parametric tests.

TABLE 2: Demographic and experiential characteristics of the participant groups. Values are mean with the observed min-max range in parentheses. Age and driving experience are in years; Likert items range 0-4.

Measure	DVR (16)	IVR (18)	Overall (34)
Gender (M/F)	13/3	11/7	24/10
Age, y	26.1 (20-50)	25.4 (21-34)	25.7 (20-50)
Driving exp., y	6.6 (1-27)	6.4 (1-16)	6.5 (1-27)
Use of IVR	0.4 (0-3)	0.7 (0-2)	0.6 (0-3)
Driving familiarity	3.2 (1-4)	2.8 (1-4)	3.0 (1-4)
3D gaming, non-VR	1.6 (0-4)	2.0 (0-4)	1.8 (0-4)
3D software	0.6 (0-3)	1.9 (0-4)	1.3 (0-4)
Motion sickness	1.2 (0-4)	0.9 (0-4)	1.1 (0-4)

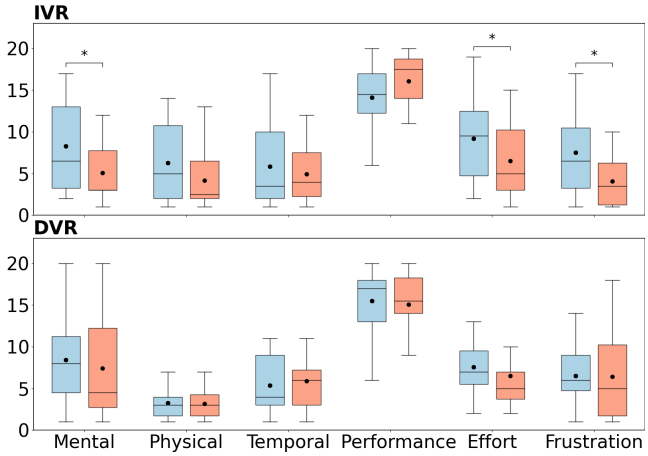


FIGURE 8: NASA-TLX results for IVR *(top)* and DVR *(bottom)* configuration: HDD scores are represented in blue, AR-HUD's in orange; (* $p < 0.05$).

C. Subjective Outcomes

1) Perceived Workload

NASA-TLX results indicate that the AR-HUD tended to reduce perceived workload compared to the HDD, with the clearest effects in the IVR condition (Fig. 8). In the IVR condition, significant reductions were observed for Mental Demand, Effort, and Frustration (all $p < 0.01$), with consistently lower mean scores in the AR-HUD condition. In contrast, no significant AR-HUD-HDD differences were found in the DVR condition.

Overall, perceived workload showed a clearer AR-HUD-HDD separation within the IVR setup, whereas ratings were more similar across interfaces within the DVR setup. This pattern is interpreted descriptively, in line with the exploratory role of the present human-subject study. Across simulator setups, IVR showed lower NASA-TLX scores than DVR on most subscales, most notably Mental Demand and Physical Demand, but these between-setup differences did not reach statistical significance (all $p > 0.05$). These results are therefore treated as descriptive setup-level patterns.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Open Journal of Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OJVT.2026.3740837

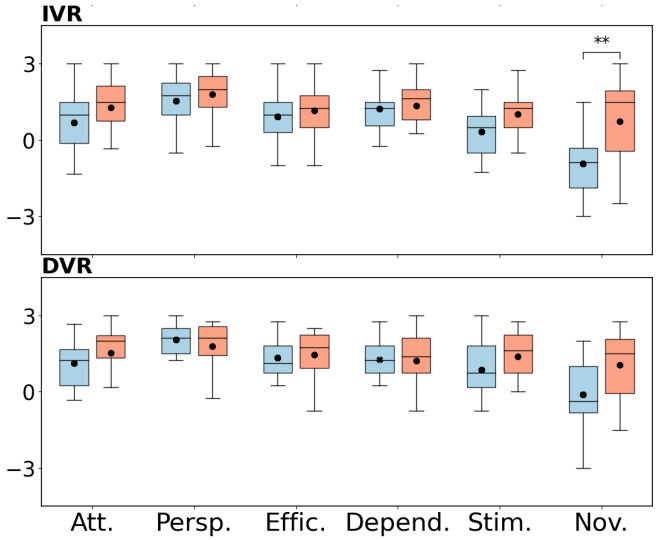


FIGURE 9: *Top*: overall UEQ results *(Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, Novelty)*. IVR and DVR configuration scores are presented respectively in *Top* and *Bottom*. In the chart, HDD scores are represented in blue, AR-HUD's in orange; the central line is the median and the dot the mean; (** $p < 0.01$).

2) User Experience

UEQ results show that the AR-HUD was rated more positively than the HDD primarily on hedonic rather than pragmatic dimensions. In the IVR condition, the only statistically significant advantage was observed for Novelty ($p < 0.01$, $g = 0.93$), the Stimulation difference was in the same direction and of medium magnitude ($g = 0.53$) but did not reach statistical significance ($p = 0.15$) and is reported descriptively only. In the DVR condition, no sub-scale reached statistical significance, with only small descriptive effects for Novelty ($g = 0.43$) and Attractiveness ($g = 0.14$). Differences on pragmatic dimensions remained minimal across both configurations ($|g| < 0.12$).

No Interface × Setup interaction across any UEQ subscale (all $F(1, 32) < 1.7$, all $p > 0.20$; Novelty: $F(1, 32) = 0.851$, $p = .361$, $\eta_p^2 = 0.026$). The concentration of the Novelty difference within the IVR setup is therefore reported as a descriptive within-setup pattern rather than as inferential evidence of a setup-dependent effect on hedonic experience.

To characterize the experiential effects of simulation configuration, we analyzed perceived presence and simulator-induced sickness across the two simulator setups. These measures characterize the experiential conditions under which the interface-level results should be interpreted.

a: Perceived presence.

Consistent with the label-based scoring convention of Melo et al. [33], the IVR group achieved label **A** (Excellent) across all IPQ sub-scales but *Experienced Realism*, where it scored as **B** (Very Good). Conversely, the DVR group received lower labels on every dimension (Fig. 10), providing a qualitative summary of the presence gap between configurations prior to inferential testing. IPQ results confirmed that the IVR condition elicited a stronger sense of presence than the DVR condition. General Presence scores demonstrated a significant advantage for IVR ($M = 5.78$, $SD = 0.65$) compared to DVR ($M = 4.00$, $SD = 1.46$), $U = 32.0$, $p < .001$, indicating a large effect. Similarly, Spatial Presence favored IVR ($M = 5.59$, $SD = 0.74$) over DVR ($M = 4.56$, $SD = 1.02$), $U = 59.5$, $p < .01$, also reflecting a large effect. Involvement ($M_{\text{IVR}} = 5.26$, $SD = 0.96$; $M_{\text{DVR}} = 4.59$, $SD = 1.48$) and Experienced Realism ($M_{\text{IVR}} = 4.01$, $SD = 0.43$; $M_{\text{DVR}} = 3.37$, $SD = 0.50$) did not differ significantly between conditions ($p = .34$ and $p = .26$, respectively). This pattern indicates that the IVR setup was associated primarily with higher general and spatial presence, whereas involvement and experienced realism remained comparatively stable across setups in the context of the current driving task. Taken together, these results support treating IVR and DVR as meaningfully distinct experiential setups.

b: Simulator-induced sickness.

TABLE 3: SSQ sub-scale and total scores across DVR and IVR conditions ($M \pm SD$). The Max row reports the theoretical upper bound of each sub-scale.

Condition	Nausea	Oculomotor	Disorientation	Total Score
Max	200.34	159.18	292.32	235.62
DVR	0.80 ± 8.42	1.90 ± 5.95	1.16 ± 8.92	1.56 ± 6.22
IVR	6.96 ± 10.01	3.79 ± 8.66	8.12 ± 13.44	6.78 ± 8.12

SSQ scores were higher in the IVR setup than in the DVR setup (Table 3), particularly for nausea ($M = 6.96$ vs. $M = 0.80$) and disorientation ($M = 8.12$ vs. $M = 1.16$). Significant between-group differences were confirmed for nausea ($U = 377.5$, $p = .040$, $r_{rb} = +.311$) and disorientation ($U = 377.0$, $p = .035$, $r_{rb} = +.309$), both reflecting medium effects; the oculomotor sub-scale did not differ significantly ($U = 324.0$, $p = .440$). The total score difference was directionally consistent but did not reach conventional significance ($U = 376.5$, $p = .058$, $r_{rb} = +.307$), a result attributable in part to the inclusion of participants reporting zero symptom change, which increased within-group variance while reducing observed means. These results indicate that the immersive setup was more demanding in terms of simulator tolerability, as expected for HMD-based driving simulation [27]. Despite the significant between-setup differences, the absolute magnitude of IVR-induced sickness requires contextualization against established normative benchmarks. The IVR total mean ($M = 6.78$, $SD = 8.12$) falls below the grand mean of $M = 9.8$ reported in the original SSQ calibration dataset [31], and well below the most sickness-inducing simulator therein ($M = 18.8$). Furthermore, when expressed relative to the theoretical maximum of each sub-scale as defined by the Kennedy et al. scoring algorithm [31] (see **Max** row, Table 3), the IVR total mean represents approximately 2.9% of

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Open Journal of Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OJVT.2026.3740837

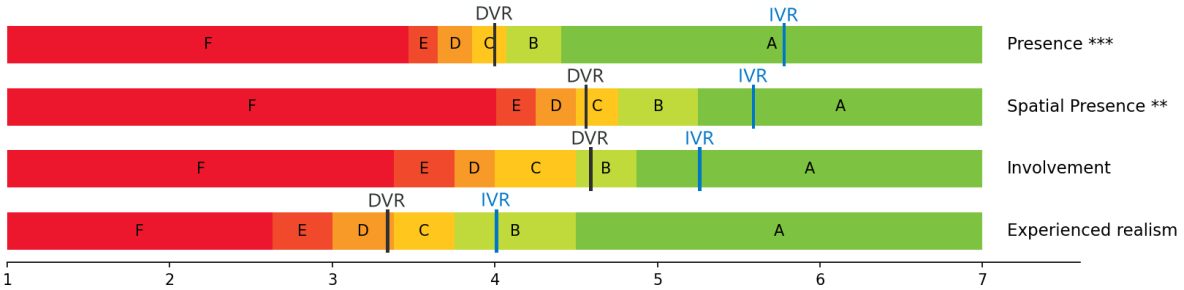


FIGURE 10: IPQ sub-scale scores for the DVR and IVR configurations. Each horizontal bar shows the qualitative score bands of Melo *et al.* [33], colored from red (F, Unacceptable) through orange and yellow to green (A, Excellent); the color encodes the band label only and carries no additional quantitative meaning. Vertical markers give the DVR and IVR group means on each sub-scale. Asterisks denote the significance of the DVR-IVR comparison ($^{**}p < 0.01$, $^{***}p < 0.001$).

TABLE 4: Trajectory consistency per condition ($M \pm SD$); lower is better for all metrics. A route excursion is a contiguous station interval of at least 2 m with lateral deviation above 5 m; SDLP is computed on the excursion-free portion of the profile. **Bold**: better interface within each setup.

Condition	SDLP [cm]	Lat. dev. [cm]	Exc. [n]	Div. [%]
IVR-AR-HUD	101.0±23.6	112.6±25.3	**0.39±0.85**	**0.82±1.87**
IVR-HDD	**95.1±21.7**	**109.5±25.5**	1.00±1.03	1.35±1.90
DVR-AR-HUD	99.3±19.5	100.7±24.6	0.19±0.40	0.18±0.48
DVR-HDD	**93.0±24.7**	**100.3±39.7**	**0.12±0.34**	**0.13±0.45**

the attainable ceiling, placing it within the *minimal* severity band per Bimberg et al. [34]. Critically, no session was excluded due to sickness-related interruption, confirming that symptom levels, while elevated relative to the DVR condition, remained within operationally acceptable limits throughout the experiment.

D. Objective Outcomes

Objective outcomes are organised in three metric families. Trajectory consistency reports how closely drivers followed the route throughout the experiments; hazard response describes how drivers reacted to the two critical events; collisions summarises the safety-critical outcomes of those events.

1) Trajectory Consistency

Trajectory consistency was computed from the lateral deviation profile sampled along the mid-lane of the reference route through four indices (Table 4): Standard Deviation of Lateral Position (SDLP) and mean absolute lateral deviation for continuous lane-keeping precision, route excursions and divergence rate for discrete departures from the ideal path.

Lane-keeping precision was independent of interface: neither SDLP nor mean lateral deviation differed between AR-HUD and HDD in either setup, and no Interface $\times$ Setup interaction emerged (all $p > 0.10$); the main effect of setup on SDLP approached but did not reach significance ($F_{1,32} = 3.38$, $p = .075$). Excursion-based measures differed only in the IVR setup, where AR-HUD guidance produced fewer route excursions than HDD ($V^+ = 8.0$, $r_{rb} = -0.71$, $p = .041$), with divergence rate following the same direction at a comparable effect size ($V^+ = 15.0$, $r_{rb} = -0.62$, $p = .060$). Under DVR neither measure differed between interfaces ($p > 0.10$), excursions being rare under both and leaving little room for an interface effect to emerge.

AR-HUD-HDD differences therefore did not concern continuous lane-keeping precision, but only discrete route excursions, where AR-HUD produced fewer than half as many departures from the intended route under IVR. The between-setup contrast is reported as a descriptive within-setup pattern rather than as inferential evidence of a setup-dependent effect. The two excursion measures are not independent, and their IVR effect sizes are large and mutually consistent while the corresponding p-values lie close to the conventional threshold at the available sample size; this metric family is therefore read as convergent exploratory evidence rather than a confirmatory result, in line with selected subjective workload and user-experience measures.

2) Reaction to Hazardous Events

Hazard response was derived from the controller inputs logged during the two scripted critical events. Reaction time was defined as the time to the first avoidance action, braking or evasive steering, after the obstacle entered the lane, and is reported as median [Q1-Q3] given non-normal distributions. Avoidance strategy and steering dynamics were analysed with the reaction time to separate *when* drivers reacted from *how* they reacted. The pedestrian crossing yielded no timing effect of either factor. Two IVR participants produced no detectable response in one of their sessions and were excluded, leaving N=32 (16 IVR, 16 DVR). The mixed-design ANOVA showed neither main effects nor an Interface $\times$ Setup interaction (all $p > 0.10$), and the within-setup interface contrasts were likewise non-significant (all $p > 0.10$). Median reaction times were descriptively shorter under DVR (0.94 s [0.67-1.29] HDD; 1.04 s [0.85-1.26] AR-HUD) than under IVR (1.65 s [1.05-2.47] HDD; 1.28 s [1.13-1.69] AR-HUD), with an interface ordering that was not

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Open Journal of Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OJVT.2026.3740837

consistent across setups: AR-HUD was faster than HDD under IVR and marginally slower under DVR.

The avoidance strategy differed markedly between setups. Braking was almost universal in DVR (96.9 % of sessions) and no driver avoided by steering alone, on the other hand in IVR steering-only avoidance occurred in 19.4 % of encounters and braking in only 69.4 % (braking: Fisher's exact p=.004; steering-only: p=.012; participant-level braking rate: U=72, p=.003). The difference held across interfaces within each setup, pointing to an effect of the setup rather than of the interface. Steering dynamics provided the only interface-dependent effect of this family: peak steering rate during the pedestrian conflict, where no curve negotiation overlaps the response, showed an Interface $\times$ Setup interaction ($F_{1,32} = 9.51$, $\eta_p^2 = .23$, $p = .004$), with AR-HUD associated with sharper evasive steering than HDD under IVR ($W = 31$, $p = .016$) and with softer responses under DVR ($W = 26$, $p = .029$); steering variability and reversal rates followed the same pattern, while slower braking was only weakly associated with sharper steering peaks ($\rho = 0.25$, partial correlation, $p = .073$). Scene-referenced hazard cues may therefore affect steering-response dynamics differently across setups, although the present design does not isolate which setup component accounts for the interaction. Contrary to pedestrian hazards, vehicle cut-in could not be analysed in the same way. Only 10 of the 32 DVR reactions occurred before the vehicle entered the lane, while in 25 of 36 IVR runs the first detected action occurred while the vehicle was turning around the corner, as the cutting vehicle became visible; in IVR the first channel engaged was steering in 35 out of 36 runs, making steering correction indistinguishable from obstacle avoidance. One interpretation is that the head-tracked viewpoint let users see the parked vehicle in advance and anticipate the manoeuvre during the curve, leaving no separable reaction, whereas under the fixed monitor-based view the need to avoid became apparent later and more often elicited a separate, delayed response. The data cannot adjudicate this account, since the absence of eye tracking (Sec. VI-C) does not establish when the vehicle was actually seen in either setup. The ambiguity is specific to the cut-in geometry, since the pedestrian crossing occurs on a straight segment where steering input is not confounded with path following, which is why the steering measures above are restricted to that event.

The two setups thus differed in the type of avoidance response adopted, but the reaction time data did not reach significance to support a differential effect of either simulator setup or interface on hazard-response timing.

3) Collisions

Collision rates remained low across all conditions: they showed no significant difference between interfaces in the IVR group ($p = .289$), the DVR group ($p = 1.000$), or the pooled sample ($p = .294$). Vehicle collision rates were negligible across all conditions (pooled Fisher's exact: $p = 1.000$).

VI. Discussion

The technical characterization shows that CARLA-IVR can sustain immersive AR-HUD operation under controlled real-time conditions. Our tests indicate that the rendering pipeline remains stable within the identified operating conditions. This supports the role of CARLA-IVR as reusable infrastructure for driver-in-the-loop AR-HUD experimentation, independent of the specific human-subject results discussed below.

Those human-subject observations, in contrast, come from a small initial evaluation, and the interaction between interface and simulator setup was not statistically supported at the available sample size. Consequently, the results should be viewed as exploratory rather than confirmatory, whereas the platform-level findings stand on their own.

First, the clearest AR-HUD-HDD separation emerged in perceived workload within the IVR setup. Specifically, the AR-HUD condition was associated with lower Mental Demand, Effort, and Frustration scores on the NASA-TLX, indicating a lower subjective workload burden, whereas no corresponding differences were observed in the DVR condition. Within the scope of the present study, perceived-workload differences between AR-HUD and HDD were more apparent within the IVR setup, whereas the DVR setup produced more similar workload ratings across interfaces.

Second, AR-HUD advantages for user experience were more limited and concentrated primarily on hedonic rather than pragmatic dimensions. The strongest effect was observed for Novelty, especially under IVR, while pragmatic UEQ dimensions remained largely comparable across interfaces. This pattern suggests that, in the present study, the AR-HUD was not perceived as substantially easier to use than the HDD, but rather as more engaging and experientially distinctive, particularly within the IVR setup.

Third, objective performance outcomes showed a more mixed pattern. Trajectory-consistency results indicated that AR-HUD-HDD differences did not involve continuous lane-keeping precision, but were instead limited to discrete route-excursion measures, with the clearest separation observed in the IVR setup. Reaction time varied more clearly across simulator setups than across interfaces, with faster responses observed overall in the DVR setup but did not reach statistical significance. By contrast, steering dynamics during the pedestrian conflict showed a setup-dependent AR-HUD-HDD pattern, suggesting that scene-referenced hazard cues may affect corrective steering behavior differently across simulator setups. These objective results therefore do not support a broad performance advantage of AR-HUD, but they indicate that different metric families capture different aspects of driver response.

A. Relation to Prior AR-HUD Literature

Both workload and user-experience findings are consistent with prior AR-HUD literature reporting that spatially aligned information reduces subjective demand and improves expe-

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Open Journal of Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OJVT.2026.3740837

riential quality [1], [3], [4]. The present results refine this picture by showing that, in this initial evaluation case, differences between AR-HUD and HDD on selected workload dimensions and UEQ Novelty were more evident within the IVR setup than within the DVR setup. This pattern suggests that simulator setup may affect how strongly certain AR-HUD-related outcomes are expressed, without identifying which setup component accounts for the difference.

These findings are consistent with methodological concerns identified in recent systematic reviews. Vezzani et al. [4] document that AR-HUD evaluations relying on virtual environments frequently fail to replicate critical hardware constraints and explicitly call for progressive, multi-stage testing frameworks spanning conceptual evaluation through on-road validation. The present study provides a concrete example of this methodological issue: when AR-HUD and HDD interfaces were evaluated in two different simulator setups, some subjective and trajectory-related outcome patterns differed across setups. This supports the broader argument that AR-HUD evaluations should report simulator characteristics in sufficient detail and should avoid assuming that results obtained in one setup necessarily generalize unchanged to another.

In this sense, the contribution of this study lies less in identifying entirely new AR-HUD effects and more in refining how they should be interpreted across simulator setups. The results indicate that the direction of several AR-HUD-HDD patterns was broadly consistent across IVR and DVR, while the magnitude and statistical expression of selected outcomes differed between setups. This illustrates how an open simulator framework can support controlled comparisons of interface outcomes across distinct display-and-interaction configurations.

B. Methodological Implications for AR-HUD Evaluation

The present results suggest that simulator setup requires explicit methodological consideration in AR-HUD evaluation. In the controlled case examined here, AR-HUD-HDD differences were not expressed identically across the two setups, indicating that setup characteristics can shape the measurement context in which interface outcomes are observed.

The scope of this implication is bounded by the design of the present study. The experiment was limited to SAE Level 0 and a restricted set of scene-referenced tasks (see Section VI-C), which limits generalizability but does not invalidate the observation. The tasks considered, namely wayfinding and event-related safety cues, rely on the spatial relationship between interface elements and the surrounding driving scene. They are therefore useful cases for examining AR-HUD interfaces across simulator setups that differ in stereoscopy, viewpoint control, motion feedback, and cockpit embodiment.

From a methodological perspective, comparability in AR-HUD evaluation depends not only on consistent task definitions and measurement choices but also on the simulation environment in which those tasks are implemented. Some simulator setups may therefore provide more appropriate measurement contexts for specific AR-HUD questions, particularly when the interfaces under study rely on scene integration, depth relationships, or viewpoint-dependent spatial registration.

The present study examined one AR-HUD implementation under SAE Level 0 conditions, using scene-referenced wayfinding and warning cues. Other AR-HUD designs may place stronger demands on depth layering, perspective-correct spatial anchoring, or occlusion-based cue integration. Larger and more targeted studies using the proposed infrastructure can assess whether AR-HUD-HDD outcome patterns remain stable across simulator setups or vary with specific display-and-interaction characteristics.

C. Limitations

This study should be interpreted as an initial controlled evaluation case. It was limited to SAE Level 0, a restricted task set, and a single AR-HUD/HDD instantiation pair, with a modest between-subjects sample for the IVR-DVR comparison. This sample size limits the statistical sensitivity of the mixed-design analysis, particularly for detecting Interface $\times$ Setup interactions. Non-significant interaction terms should therefore be interpreted as inconclusive for small or moderate interaction effects rather than as evidence of their absence. These constraints delimit the empirical generalizability of the human-subject study to other driving contexts, automation levels, and interface designs.

A second limitation concerns the interpretation of the IVR-DVR comparison. The two simulator setups differed along several coupled dimensions, including display technology, viewpoint control, motion feedback, cockpit layout, and steering hardware. The observed between-setup differences are therefore best interpreted as differences between two complete display-and-interaction configurations rather than as the isolated effect of visual immersion, stereoscopy, motion feedback, or any other single component.

The objective measurement space remains incomplete. Eye-tracking measures could not be collected because the HMD used in the IVR setup did not provide integrated eye tracking, and comparable gaze data were not available across simulator setups. This prevents direct analysis of attentional allocation, visual search behavior, and glance patterns, which are important for stronger claims about AR-HUD safety and attentional guidance. Future iterations should therefore incorporate gaze-based measures.

Reaction times were slower overall in IVR than in DVR, but results did not reach statistical significance and the mechanism underlying this difference remains unclear. A more demanding display-and-interaction context, a more conservative response threshold under immersive conditions, differences in control hardware, and broader characteristics of the IVR setup may all have contributed, and the present

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Open Journal of Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OJVT.2026.3740837

study does not allow these alternatives to be disentangled reliably. At the same time, IVR-induced sickness remained within operationally acceptable bounds: the total SSQ mean fell below the grand mean reported in the original calibration dataset [31], within the minimal severity band defined by Bimberg et al. [34], and no session was interrupted due to discomfort. This suggests that the immersive setup achieved an adequate tolerability profile despite elevated scores relative to DVR, making it less likely that the observed between-setup differences can be explained solely by simulator-induced sickness.

Finally, the current implementation reflects technical constraints of the present platform instantiation. As discussed in Section III, the immersive extension was developed under specific CARLA and Unreal Engine conditions that imposed restrictions on environment configuration and content integration. These should be understood as current operational boundaries rather than intrinsic architectural limitations of the proposed approach and may be relaxed as the CARLA ecosystem evolves.

VII. Conclusion

This paper presented *CARLA-IVR*, an open immersive extension of the CARLA simulator designed to support controlled AR-HUD experimentation while preserving interoperability with standard CARLA workflows. The system integrates stereoscopic XR rendering, embodied vehicle interaction, driver-in-the-loop vehicle control, and scene-referenced AR-HUD overlays within a reusable simulation framework.

Beyond the architectural contribution, the paper assesses the real-time rendering capabilities of the system. The analysis of the frame rate stability, input-to-render latency, and a stress test under increasing traffic-agent density were used to identify the operating limits of the IVR condition. While these findings are hardware and scenario bound they show that the framework could be used to evaluate AR-HUD interfaces in IVR under reproducible yet dynamically varying traffic conditions within the identified limits.

The paper also introduced an initial, literature-grounded evaluation case focused on recurrent AR-HUD tasks and measures, providing a first controlled use of the proposed infrastructure for comparative study design. The human-subject study compared AR-HUD and HDD interfaces across two simulator setups: an immersive IVR setup and a desktop DVR setup. The results showed that selected AR-HUD-HDD outcome patterns differed across setups. Subjective workload differences were more pronounced within the IVR setup, particularly for NASA-TLX Mental Demand, Effort, and Frustration, while UEQ differences were mainly concentrated on Novelty. Objective outcomes were more mixed: continuous lane-keeping measures were interface-independent, route-excursion measures favored AR-HUD within the IVR setup, and reaction times did not differ significantly across either factor, although avoidance strategy differed between setups, suggesting that the simulator setup can shape the driver behavior in addition to how AR-HUD effects are presented and measured.

Taken together, these findings should be interpreted as an exploratory evaluation case rather than as a general efficacy claim for AR-HUDs or as evidence of an isolated effect of visual immersion. Their main methodological implication is that AR-HUD outcomes should be interpreted in relation to the simulator setup in which they are measured, particularly when interfaces rely on scene-referenced cues, spatial registration, and driver-in-the-loop interaction. Because the setup also influenced the avoidance strategy adopted, this consideration extends beyond measurement sensitivity: findings obtained under one setup may reflect a different distribution of driver responses than would arise under another, which is relevant whenever simulator results are used to inform interface design decisions. In this sense, CARLA-IVR provides an open technical foundation for developing more controlled, reusable, and comparable AR-HUD evaluation protocols.

Future work should extend the CARLA-IVR foundation to richer traffic conditions, broader task sets, additional automation levels, eye-tracking-enabled protocols, and more diverse display-and-interaction configurations. In particular, future protocols can define discrete traffic-density conditions and cross them with interface type. Similarly, the position-triggered hazard-event mechanism can be extended to additional event types and multi-conflict intersections, investigating more demanding wayfinding and hazard-response tasks. Larger and more targeted studies should also examine whether AR-HUD-HDD outcome patterns remain stable across simulator setups or vary with specific components; isolating individual contributions would require dedicated factorial or ablation designs rather than comparisons between complete configurations.

VIII. Acknowledgment

The authors utilized AI-assisted tools, including Claude AI, Instatext, and Grammarly, to improve spelling, grammar, clarity, and readability during manuscript preparation. After using these tools, the authors carefully reviewed and edited the text and take full responsibility for the final content.

REFERENCES

[1] J. L. Gabbard, G. M. Fitch, and H. Kim, "Behind the Glass: Driver Challenges and Opportunities for AR Automotive Applications," *Proceedings of the IEEE*, vol. 102, no. 2, pp. 124–136, Feb. 2014.

[2] F. Hartwich and A. Koller, "A review of augmented reality in driving applications," *Journal of Intelligent Transportation Systems*, vol. 22, no. 5, pp. 373–391, 2018.

[3] M. Winkler and M. Soleimani, "A Review of Augmented Reality Heads Up Display in Vehicles: Effectiveness, Application, and Safety," *International Journal of Human–Computer Interaction*, pp. 1–16, Jan. 2025.

[4] L. Vezzani, F. Strada, and A. Bottino, "A review of automotive ar-hud interfaces across driver roles and vehicle automation levels," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–21, 2026.

[5] A. Dosovitskiy *et al.*, "Carla : An open urban driving simulator," in *Conference on Robot Learning*, 2017, pp. 1–16.

[6] P. A. Lopez, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y.-P. Flötteröd, R. Hilbrich, L. Lücken, J. Rummel, P. Wagner, and E. Wießner,

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Open Journal of Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OJVT.2026.3740837

"Microscopic traffic simulation using sumo," in *The 21st IEEE International Conference on Intelligent Transportation Systems*. IEEE, 2018.

[7] P. Czechowski, B. Kawa, M. Sakhai, and M. Wielgosz, "Deep reinforcement and il for autonomous driving: A review in the carla simulation environment," *Applied Sciences*, vol. 15, no. 16, 2025.

[8] E. Delavari, F. K. Khanzada, and J. Kwon, "A comprehensive review of reinforcement learning for autonomous driving in the carla simulator," 2025.

[9] Z. Qiao, X. Sun, H. Loeb, and R. Mangharam, "Drive right: Shaping public's trust, understanding, and preference towards autonomous vehicles using a virtual reality driving simulator," in *2023 IEEE Intelligent Vehicles Symposium (IV)*, 2023, pp. 1–8.

[10] G. Silvera, A. Biswas, and H. Admoni, "DReyeVR: Democratizing Virtual Reality Driving Simulation for Behavioural & Interaction Research," Jan. 2022.

[11] R. Currano, S. Y. Park, D. J. Moore, K. Lyons, and D. Sirkin, "Little Road Driving HUD: Heads-Up Display Complexity Influences Drivers' Perceptions of Automated Vehicles," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Yokohama Japan: ACM, May 2021, pp. 1–15.

[12] X. Li, J. Rong, Z. Li, X. Zhao, and Y. Zhang, "Modeling drivers' acceptance of augmented reality head-up display in connected environment," *Displays*, vol. 75, p. 102307, Dec. 2022.

[13] C. Merenda, H. Kim, K. Tanous, J. L. Gabbard, B. Feichtl, T. Misu, and C. Suga, "Augmented Reality Interface Design Approaches for Goal-directed and Stimulus-driven Driving Tasks," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 11, pp. 2875–2885, Nov. 2018.

[14] S. M. Russell, J. Radlbeck, J. Atwood, W. A. Schaudt, S. B. McLaughlin, and Virginia Tech Transportation Institute, "Head-Up Displays and Distraction Potential," NHTSA, Tech. Rep. DOT HS 813 293, Jan. 2023.

[15] U. Technologies, "Unity – real-time 3d game & simulation engine," https://unity.com/, 2025.

[16] AVSimulation, "Scaner – driving simulation software," https://www.avsimulation.com/scaner/, 2025.

[17] V. Charissis, "Enhancing human responses through augmented reality Head-Up Display in vehicular environment," in *2014 11th International Conference & Expo on Emerging Technologies for a Smarter World (CEWIT)*, Oct. 2014, pp. 1–6.

[18] F. Schewe and M. Vollrath, "Visualizing distances as a function of speed: Design and evaluation of a distance-speedometer," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 64, pp. 260–273, Jul. 2019.

[19] C. Maag, N. Schömig, F. Naujoks, I. Karl, A. Keinath, and A. Neukum, "Measuring workload effects of augmented reality head-up displays using detection response task," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 92, pp. 201–219, Jan. 2023.

[20] A. Calvi, F. D'Amico, C. Ferrante, and L. Bianchini Ciampoli, "Effectiveness of augmented reality warnings on driving behaviour whilst approaching pedestrian crossings: A driving simulator study," *Accident Analysis & Prevention*, vol. 147, p. 105760, Nov. 2020.

[21] R. Ma and D. B. Kaber, "The impact of augmented reality head-up displays on driver performance and attention," *Applied Ergonomics*, vol. 74, pp. 167–177, 2019.

[22] L. Morra, F. Lamberti, F. G. Pratticó, S. L. Rosa, and P. Montuschi, "Building Trust in Autonomous Vehicles: Role of Virtual Reality Driving Simulators in HMI Design," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 10, pp. 9438–9450, Oct. 2019.

[23] G. Sharma, H. Yasuda, and M. Kuehner, "Continuous Visual Feedback of Risk for Haptic Lateral Assistance," Jan. 2023.

[24] Microsoft, "Airsim," https://microsoft.github.io/AirSim/, 2017.

[25] L. S. V. Lab, "Lgsvl simulator - an autonomous vehicle simulator," https://github.com/lgsvl/simulator, 2021.

[26] L. Vezzani, F. Strada, and A. Bottino, "Towards Carla VR: Testing Head-Up Displays in Carla Driving Simulator," in *Advances in Information and Communication*, K. Arai, Ed. Cham: Springer Nature Switzerland, 2025, pp. 1–16.

[27] K. M. Stanney, R. S. Kennedy, and J. M. Drexler, "Cybersickness is Not Simulator Sickness," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 41, no. 2, pp. 1138–1142, Oct. 1997.

[28] W. Narzt, G. Pomberger, A. Ferscha, D. Kolb, R. Müller, J. Wieghardt, H. Hörtner, and C. Lindinger, "Augmented reality navigation systems," *Universal Access in the Information Society*, vol. 4, no. 3, pp. 177–187, Mar. 2006. [Online]. Available: https://doi.org/10.1007/s10209-005-0017-5

[29] S. G. Hart and L. E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," in *Human Mental Workload*, ser. Advances in Psychology, P. A. Hancock and N. Meshkati, Eds. North-Holland, 1988, vol. 52, pp. 139–183.

[30] B. Laugwitz, T. Held, and M. Schrepp, "Construction and evaluation of a user experience questionnaire," in *HCI and Usability for Education and Work*, A. Holzinger, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 63–76.

[31] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal, "Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness," *The International Journal of Aviation Psychology*, vol. 3, no. 3, pp. 203–220, 1993.

[32] T. Schubert, F. Friedmann, and H. Regenbrecht, "The experience of presence: Factor analytic insights," *Presence: Teleoperators and Virtual Environments*, vol. 10, no. 3, pp. 266–281, 06 2001.

[33] M. Melo, G. Gonçalves, J. Vasconcelos-Raposo, and M. Bessa, "How much presence is enough? qualitative scales for interpreting the igroup presence questionnaire score," vol. 11, pp. 24 675–24 685.

[34] P. Bimberg, T. Weissker, and A. Kulik, "On the Usage of the Simulator Sickness Questionnaire for Virtual Reality Research," in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. Atlanta, GA, USA: IEEE, Mar. 2020, pp. 464–467.

Leonardo Vezzani Leonardo Vezzani is currently a PhD student in the Department of Control and Computer Engineering, Politecnico di Torino, where he conducts research under the Computer Graphics and Vision Research Group. His current research interests include virtual and augmented reality for automotive head-up displays, virtual reality driving simulations and explainable AI for autonomous vehicles.

Francesco Strada Francesco Strada is currently an Assistant Professor with the Department of Control and Computer Engineering, Politecnico di Torino, where he conducts research under the Computer Graphics and Vision Research Group. His current research interests include virtual and augmented reality for learning and training, serious games, human–computer interaction, and the use of immersive technologies in collaborative scenarios.

Andrea Bottino Andrea Bottino (Member, IEEE) is currently an Associate Professor with the Department of Control and Computer Engineering, Politecnico di Torino, where he heads the Computer Graphics and Vision Research Group. His current research interests include computer vision, machine learning, human–computer interaction, serious games, and virtual and augmented reality.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/