A Practical Protocol and Governance Framework for Immersive In-vehicle HCI Usability Evaluation Using Commercial XR Analytics

Felipe Dias [1], Rafael Miguez [2], Catharina Ramos [3], Daniela Rodrigues [3], João Marcelo Teixeira [4], Sanval Santos [1], Yasmim Thasla Santos Ferreira [1,5,6,*], and Ingrid Winkler [1,5,6,*]
[1] Human-Centric Industrial Immersive Engineering Laboratory (HIIVE Lab), SENAI CIMATEC University, Salvador, Brazil
[2] Digital Innovation in Production Systems Laboratory (DIPS), NOVA University Lisbon, Lisbon, Portugal
[3] Global Automotive, Brazil
[4] Voxar Labs, Computer Center, Federal University of Pernambuco, Recife, Brazil
[5] Institute of Science, Innovation and Technology Industry 4.0 (INCITE), SENAI CIMATEC University, Salvador, Brazil
[6] Advanced Knowledge Center for Immersive Technologies (AKCIT), Federal University of Goiás, Goiânia, Brazil
Email: felipe.dias@aln.senaicimatec.edu.br (F.D.); Rafael.miguez@aln.senaicimatec.edu.br (R.M.); cramos49@ford.com (C.R.); drodr223@ford.com (D.R.); jmxnt@cin.ufpe.br (J.M.T.); sanval.santos@fieb.org.br (S.S.); yasmim.ferreira@aln.senaicimatec.edu.br (Y.T.S.F.); ingrid.winkler@doc.senaicimatec.edu.br (I.W.)
*Corresponding author

Abstract*—Virtual reality combined with embedded eye tracking offers a promising alternative to traditional in-vehicle usability assessment; however, integrating commercial extended reality analytics platforms into immersive workflows raises methodological challenges related to standardization, reproducibility, metric transparency, and version control. This study proposes a repeatable evaluation protocol and governance framework for the use of commercial extended reality analytics in immersive in-vehicle usability assessment. The framework was operationalized in an exploratory proof-of-concept study involving an immersive cockpit simulation, embedded eye tracking, a proprietary spatial analytics platform, and a physical vehicle mockup aligned with the virtual scene. Ten participants performed structured driving-related tasks, and gaze data were converted into spatial visualizations, behavioral metrics, and platform-derived presence indicators, which were descriptively compared with subjective questionnaire responses to support multimodal interpretation. The findings suggest that commercial analytics platforms can support the early identification of ergonomic inconsistencies and attention-distribution patterns in immersive cockpit evaluation. At the same time, the study highlights structural dependencies introduced by proprietary infrastructures, including metric opacity, software version instability, and limited auditability, which may affect longitudinal comparability and interpretive reliability. Practically, the study provides a reusable workflow for collision and area-of-interest validation, version documentation, multimodal benchmarking, and independent data archival to support more transparent, auditable, and repeatable immersive usability evaluation.

***Keywords*—virtual reality, eye tracking, in-vehicle usability, extended reality analytics governance**

Manuscript received March 12, 2026; revised April 2, 2026; accepted May 13, 2026; published July 23, 2026.

I. INTRODUCTION

Usability plays a central role in the design and validation of in-vehicle systems. According to the ISO-9241 standard, usability refers to the extent to which a system, product or service can be used by specified users to achieve specific goals with effectiveness, efficiency, and satisfaction in a specified context of use [1]. In-vehicle usability assessments evaluate the intuitiveness and efficiency of components during driver interactions. Conventionally, automotive usability assessment relies heavily on physical prototypes, where participant performance is captured through predominantly subjective instruments such as interviews and questionnaires. Fig. 1 illustrates this process.

Fig. 1. Conventional in-vehicle usability assessment conducted with driver and ergonomist [2].

Although widely used, this conventional approach presents several limitations. First, building and modifying physical prototypes is time-consuming and costly, particularly when design updates are frequent [3]. Second, an exclusive reliance on subjective self-report measures may overlook usability flaws [4], since participants may forget details or feel uncomfortable criticizing design elements. Beyond these operational constraints, a broader methodological limitation emerges when immersive technologies are adopted: the absence of standardized evaluation pipelines capable of structuring data collection, synchronization, and analysis across studies. Immersive usability assessments frequently rely on ad hoc procedures, complicating replication and benchmarking [5]. Furthermore, the increasing dependence on proprietary analytics platforms introduces additional challenges related to methodological transparency, metric interpretation, and version control, potentially affecting reproducibility and longitudinal comparisons. These limitations are particularly critical in complex domains such as automotive systems, where minor design flaws can compromise both safety and user satisfaction.

To overcome these challenges, automakers have increasingly adopted Virtual Reality (VR) to perform usability simulations with immersive prototypes [6]. VR enables users to interact with virtual interfaces under controlled and realistic conditions, supporting earlier identification of potential design issues before physical production [7–11]. However, these studies do not address a governance-oriented protocol for immersive Human-Computer Interaction (HCI) evaluation in sensor-rich environments, nor do they examine how proprietary Extended Reality (XR) analytics infrastructures affect traceability, reproducibility, and interpretation in usability assessment. In the automotive domain, immersive HCI evaluation also overlaps with Human-Machine Interface (HMI) evaluation, because cockpit interface decisions affect driver attention, operational safety, and the security of increasingly connected vehicle systems. Recent automotive HMI literature has therefore emphasized that usability, safety, and security must be considered jointly in interface design, while VR-based research highlights the value of immersive simulation for safety-aware HMI prototyping and training [12, 13]. This reinforces the need to position immersive cockpit evaluation not only as a generic HCI problem, but as a safety- and security-relevant HMI validation problem.

For vehicle-interior projects, VR simulators can improve cost-efficiency, repeatability, safety, and experimental control while maintaining perceptual realism through stereoscopy, depth, and spatialized audio [14]. Human-centered VR design principles further emphasize that immersion, sensory consistency, and interaction fidelity are critical factors influencing user experience and behavioral realism in virtual environments [15]. Recent work on immersive training and industrial simulation environments also demonstrates how virtual worlds can be systematically developed to support professional applications and experimentation [16].

While these characteristics enhance ecological validity and offer promising conditions for usability testing, the methodological structure for evaluating usability within immersive systems remains fragmented across studies [17]. Even in related digital transformation initiatives, such as gamified information systems and visual management platforms, structured evaluation protocols are necessary to ensure that proposed artifacts can be systematically assessed and improved [18]. Different research groups adopt heterogeneous data collection procedures, metrics, and analysis strategies, which limits comparability and benchmarking across implementations.

In addition, recent generations of Head-Mounted Displays (HMDs) integrate native eye tracking sensors, enabling the collection of fine-grained gaze data during immersive interaction. These sensors provide objective measurements of visual attention, cognitive load proxies, and interaction patterns [19]. Metrics such as fixations, gaze duration, and pupil dilation can reveal subtle usability issues that may not be captured by subjective questionnaires alone [20–23]. However, embedded eye tracking should be understood as a source of raw behavioral signals rather than a ready-to-use usability metric. Gaze data are high-frequency, time-dependent streams that require structured synchronization with system events, segmentation according to task phases, and clearly defined interpretation criteria. Without a standardized analytical pipeline, the extraction of meaningful usability indicators may vary substantially across implementations, limiting comparability, reproducibility, and benchmarking of immersive studies.

Recent studies demonstrate the potential of combining VR and eye tracking to analyze driver behavior, attention, and ergonomics [24–27]. Nevertheless, applications integrating these technologies into structured and replicable automotive usability evaluation frameworks remain limited, particularly with respect to standardized analysis workflows and governance mechanisms for automated gaze analytics.

One emerging approach to address this challenge involves using analytics platforms that transform raw eye tracking data into insights directly within immersive environments. While raw gaze data offers granular access to users' visual behavior, it typically requires substantial technical expertise and post-processing. Commercial XR analytics platforms have emerged to operationalize this process by transforming raw gaze streams into visual summaries and behavioral metrics. Among these tools, Cognitive3D [28] contextualizes gaze and interaction data into 3D visualizations such as heatmaps and fixation maps. Interpretation layers of this kind aggregate spatial and temporal data to facilitate analysis within immersive environments [29, 30].

However, these platforms represent proprietary interpretative abstractions built on top of raw behavioral signals rather than transparent measurement instruments. Like other software-mediated research infrastructures, commercial XR analytics systems do not merely record observations; they actively configure what counts as an observation by determining how data are sampled, filtered, synchronized, spatially mapped, aggregated, and rendered as usable indicators. For that reason, governance concerns arise prior to, and independently from, the empirical findings themselves. Once usability evidence depends on a vendor-controlled analytics stack, the measurement process becomes software-dependent: part of the operationalization of constructs is delegated to Software Development Kits (SDKs), dashboards, cloud services, and release cycles that remain only partially visible to the researcher. Under these conditions, the resulting indicators are contingent not only on participant behavior, but also on platform-specific data schemas, event definitions, scene-attribution rules, weighting schemes, and access policies. Despite their practical advantages, commercial XR analytics systems therefore introduce methodological and governance challenges, including version dependency, metric opacity, audit limitations, data ownership constraints, and update-driven variability. Such factors may directly affect construct validity, reproducibility, benchmarking, and longitudinal usability evaluation. Consequently, the structured evaluation and governance of these platforms remain an open research challenge.

From a theoretical standpoint, governance becomes necessary whenever usability assessment depends on layered measurement pipelines rather than direct observation alone. In immersive HCI, embedded sensors generate raw behavioral signals, analytics platforms transform these signals into derived indicators, and researchers interpret those indicators as evidence about constructs such as attention, engagement, or presence. This sequence creates a methodological vulnerability because software is not an external convenience layer but part of the measurement apparatus itself. Sampling settings determine what is captured, scene configuration and attribution rules determine how traces are assigned to objects, and proprietary analytics logic determines what is ultimately reported as evidence. Even when user behavior is stable, changes in firmware, SDK integration, game-engine versions, cloud-side processing, or metric-generation logic may alter the outputs and weaken comparability across sessions, studies, and time. Governance is therefore required as a condition for methodological accountability in software-dependent evaluation environments: it is the mechanism that restores traceability between raw signals, transformation rules, reported indicators, and the claims derived from them. Accordingly, the proposed framework is grounded in HCI evaluation validity, measurement traceability, and infrastructure-aware reproducibility, rather than only in the empirical behavior of a single platform. In this sense, the framework responds to a general property of measurements generated through analytics platforms rather than to a platform-specific anomaly revealed only by the present case.

Despite growing adoption of VR and embedded eye tracking in automotive studies, the field still lacks a practical and auditable protocol for using proprietary XR analytics platforms in ways that support repeatability, benchmarking, and methodological transparency.

We propose a repeatable protocol for using commercial XR analytics in immersive in-vehicle usability studies. The protocol specifies how to validate collision meshes and Areas of Interest (AOIs), document software versions, compare platform outputs with questionnaire data, and archive exports for later audit. We demonstrate the protocol in a cockpit study that combines a PICO 4 Enterprise headset [31], embedded eye tracking, Cognitive3D, and a physical mockup aligned with the virtual scene.

The remainder of this paper is organized as follows: Section II describes the materials and methods, Section III presents the results and discussion, and Section IV outlines the conclusions and future research directions.

II. Materials and Methods

This study is exploratory, as research on combining virtual reality and eye tracking technologies for in-vehicle usability assessment is still nascent [32].

We adhered to the Design Science Research (DSR) paradigm, which contributes to the advancement of knowledge while also addressing real-world applications related to the research problem or opportunity [33, 34]. The method followed the six steps of DSR: (1) identifying the problem; (2) defining the objectives; (3) designing and developing the artifact; (4) demonstrating the artifact; (5) evaluating the artifact; and (6) communicating the results.

This study introduces a repeatable evaluation protocol and governance structure for integrating VR, eye tracking, and XR analytics platforms. The artifact formalizes procedures for infrastructure documentation, task specification, gaze data acquisition, analytics processing, subjective benchmarking, and version control management, aiming to enhance methodological transparency, reproducibility, and long-term comparability of immersive usability studies.

This paper reports a proof-of-concept study rather than a confirmatory validation. Our aim was to test the workflow end to end and identify where proprietary analytics introduce traceability issues. For that reason, the IPQ and Cognitive3D comparison is treated as descriptive rather than inferential.

Given the sample size (n = 10) and exploratory design, the observed IPQ–Cognitive3D relationships are reported as descriptive and directional only. They do not establish generalizable associations, statistically robust correlations, or platform-validity claims, nor do they justify treating Cognitive3D outputs as validated proxies for psychological presence. Instead, the comparison serves as an initial triangulation exercise that highlights possible convergence and divergence between subjective and platform-derived indica tors within this specific instantiation. What is intended to be transferable from the study is the governance logic, the protocol structure, and the demonstration of methodological feasibility, not the numerical magnitude or inferential strength of the observed relationships.

For DSR steps 1 and 2, we identified the problem and defined our objectives by analyzing the research context and identifying limitations in current immersive automotive usability evaluations. This stage involved examining the technological landscape of XR analytics platforms, assessing the capabilities and constraints of Cognitive3D compared with raw eye tracking workflows, and reviewing internal design requirements from automotive usability studies. Consultations with specialists involved in immersive simulation projects also contributed to refining the practical needs, feasibility considerations, and expected value of the proposed solution.

In the third step, we combined several technologies. Unreal Engine was used to build the 3D environments. Initial prototype tests were conducted using Unreal Engine 4.27.2, followed by migration to Unreal Engine 5 to improve integration with the PICO 4 Enterprise device. The Unreal Engine [35] was selected due to its comprehensive documentation and ecosystem support for XR development. Additionally, we selected the PICO 4 Enterprise head-mounted display (Fig. 2) for testing because it integrates advanced virtual reality features, including eye tracking.

Fig. 2. PICO 4 Enterprise head-mounted display adopted.

We deployed the application as an Android Package Kit (APK) directly on the PICO 4 Enterprise device. Although the manufacturer specifies a maximum eye tracking sampling rate of 200 Hz, the effective application rendering frequency during the experiment was approximately 70 Hz. Detailed firmware versioning of the headset and SDK versions were not systematically logged during data collection, which may introduce constraints for strict reproducibility.

The vehicle cockpit (Fig. 3) used in the experiment simulates the interior of a standard car cabin, including the driver's seat, dashboard, steering wheel, and primary control interfaces. This physical setup was designed to replicate a realistic driving environment and facilitate participant immersion while allowing precise eye tracking and interaction data collection within the VR scenario.

Following the DSR's steps four and five, to evaluate our combination of virtual reality and eye tracking analytics, we carried out experiments with ten participants (referred to as participants P1–P10). Each participant was seated in the vehicle cockpit and asked to observe all the virtual environment to collect eye tracking data, including the head-mounted display global position, orientation, height, fixation points, and timestamps.

Fig. 3. The vehicle cockpit adopted.

The participants were recruited from Brazil. 70% reported prior experience with virtual reality. These characteristics are reported to contextualize the exploratory findings, as prior exposure to immersive systems may influence task performance and subjective presence.

Each participant performed four structured tasks. Task 1 simulated a lane-change maneuver, requiring participants to visually inspect rear-view mirrors and surrounding traffic conditions before executing the maneuver. Task 2 involved interaction with the vehicle dashboard, including adjusting air conditioning settings, modifying radio controls, and monitoring the speedometer. During the remaining time, participants verbally described their perceptions and sensations within the immersive environment. The expected relevant areas of visual attention included rear-view mirrors, side mirrors, speedometer, infotainment interface, and central dashboard controls. However, we did not conduct any formal pre-validation procedure to verify heatmap coverage or Area-of-Interest (AOI) mapping consistency prior to data collection.

Data were analyzed using the Cognitive3D platform [28], which provides an advanced analytics ecosystem for spatial data captured in immersive environments. Cognitive3D enables visualization of gaze paths, fixation heatmaps, and interaction patterns through its SceneExplorer and Dashboard tools, offering detailed insights into user attention and behavior. The platform supports integrations with major game engines, such as Unreal Engine, and allows data synchronization from multiple sensors, enabling multidimensional analysis of user performance and experience in VR. The dashboard version in use was active around December 2024. During the experimental phase, a platform update occurred without prior notification, resulting in modifications to visualization outputs and metric calculations. This episode highlighted the potential impact of version dependency and update-driven variability on data interpretation and reinforced the need for structured governance when relying on commercial XR analytics infrastructures.

Three governance controls recommended by the proposed artifact were not fully implemented during this demonstration. First, firmware and SDK versions were not logged systematically during data collection. Second, collision and AOI coverage were not formally pre-validated before the formal sessions. Third, raw exports were not independently archived outside the Cognitive3D platform. These omissions may constrain strict reproducibility, affect gaze attribution in specific scene regions, and limit later verification of platform-derived outputs.

Eye tracking and spatial interaction data were stored within the Cognitive3D cloud-based infrastructure under its Software-as-a-Service (SaaS) model. Questionnaire responses and complementary research data were stored separately using Google Sheets. We did not implement any independent raw data archival outside the Cognitive3D environment during the experimental phase.

The eye tracking module captured gaze rays, fixation points, and saccadic transitions. These streams were transmitted to Unreal Engine, where the embedded Cognitive3D SDK associated gaze samples with scene objects. The platform then generated higher-level analytical outputs such as gaze paths, heatmaps, and attention metrics based on proprietary spatial mapping algorithms.

To assess the subjective sense of presence, we applied the Igroup Presence Questionnaire (IPQ) [36], a validated instrument designed to measure the extent to which users experience the feeling of being there in a virtual environment. Presence, in this context, refers to the psychological perception of spatial immersion and involvement within mediated environments [37]. The IPQ is widely used in virtual reality and immersive systems research and is particularly suitable for studies investigating user experience, immersion, and engagement [38, 39].

The questionnaire employs a 5-point Likert scale (1 = strongly disagree to 5 = strongly agree) and comprises three validated subscales: Spatial Presence, Involvement, and Experienced Realism [37]. Additionally, the IPQ includes a general item that captures the overall feeling of being there, which is analyzed independently from the subscales. The instrument has demonstrated factorial validity and cross-cultural applicability in immersive research contexts [37].

We distinguish three levels of evidence. First are the raw session traces: gaze, timestamps, head/headset position, controller activity, and performance logs. Second are Cognitive3D's derived outputs, which depend on proprietary aggregation and scene mapping. Third are our interpretations of those outputs. This distinction matters because platform scores are useful indicators, not direct measures of psychological presence or comfort.

Finally, Step 6 of the DSR methodology involves communicating our findings, which we detail in the following sections.

The study was approved by the SENAI CIMATEC University Ethics Committee (7.012.542), and all research participants gave written and informed consent.

III. Results and Discussion

The cockpit study exposed both the value and the fragility of the workflow. We first show how the environment was configured and validated, then the usability findings from the participant tasks, and finally the implications of relying on a proprietary analytics stack.

A. Immersive Scenario Preparation and Initial Cognitive3D Exploration

First, we created a small interactive VR lobby for testing and eye tracking data collection. The immersive virtual environment was developed in the Unreal Engine 4 game engine. Within the test scene, small spheres were programmed to turn red every ten seconds (Fig. 4), directing the user's attention to different points in the virtual environment. We designed this step to validate data collection and evaluate potential interaction scenarios. Based on the captured information, heatmaps representing the areas of greatest visual interest were generated, supporting the analysis of user gaze behaviour in VR.

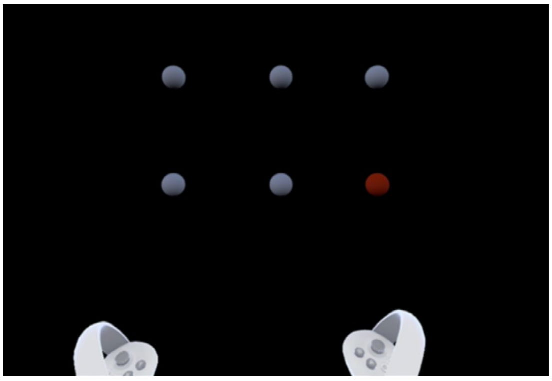

Fig. 4. Virtual-reality environment—Eye tracking test.

After defining the scenario, we selected an appropriate virtual vehicle model, prioritizing high realism and low polygon count. Two main options emerged: (1) a high-quality Audi model, widely recognized for its visual fidelity and available free on the Unreal Engine Marketplace (Fig. 5); and (2) simpler low-poly models from the free Vehicle Variety Pack (Fig. 6). A lightweight model was essential because high polygon counts strain VR hardware and degrade head-mounted display performance, especially in the case of the virtual Audi model, which contained many materials and polygons. Although the Audi offered superior detail, we ultimately adopted the Vehicle Variety Pack option due to its lower polygon count and reduced processing demand, ensuring smoother and more efficient immersion.

Fig. 5. First option of an Audi virtual vehicle model.

Fig. 6. Virtual vehicle adopted in the test environment.

Subsequently, we created a controlled and simplified immersive environment for the initial explorations with Cognitive3D. The initial design consisted exclusively of the selected virtual car placed within a low-complexity scene, ensuring an isolated test setup without interference from additional assets (Fig. 7). This minimalistic configuration enabled clearer observation of the underlying mechanisms of Cognitive3D and supported a focused evaluation of gaze and interaction data.

Fig. 7. Simplified immersive environment for the initial Cognitive3D tests.

Beyond visual outputs, Cognitive3D provides two key metrics, Presence and Comfort, ranging from 0 (poor) to 100 (excellent) [27]:

1) Comfort Level: Critical for a pleasant experience in immersive applications. Discomfort can compromise the sense of immersion and is associated with issues such as low frame rate, high frame variability, and ergonomics problems. The Application Performance parameter measures the Frame-Per-Second (FPS) rate delivered by the head-mounted display and its variability. Lower FPS, particularly below 30 Hz, can induce cybersickness. Other parameters include Neck Ergonomics, related to muscle tension from head posture, and Arm Ergonomics, which evaluates arm strain postures.
2) Presence Level: Indicates how immersive and engaging the virtual environment is. It is a composite metric summarizing the level of immersion a user experiences within the simulation. Its parameters include: (1) Engagement, which assesses the time and quality of interactions with dynamic objects; (2) Orientation, measuring head movement and how users adjust their visual perspective; (3) Spatial Coverage, the percentage of the environment explored; (4) Controller Engagement, evaluating the precision and extent of controller use; (5) Boundary, assessing proximity to the edges of the virtual tracking area. These parameters are visualized in a spider graph, and the final presence score aggregates the results across all dimensions.

In our prototype, the initial heatmap displayed missing colour markings on the dashboard and other critical areas, revealing a collision-setup issue. The vehicle model had been imported without refined collision meshes, preventing gaze data from registering on certain surfaces. After adding more comprehensive collision meshes to fully wrap the vehicle's geometry, the heatmap rendered correctly (Fig. 8).

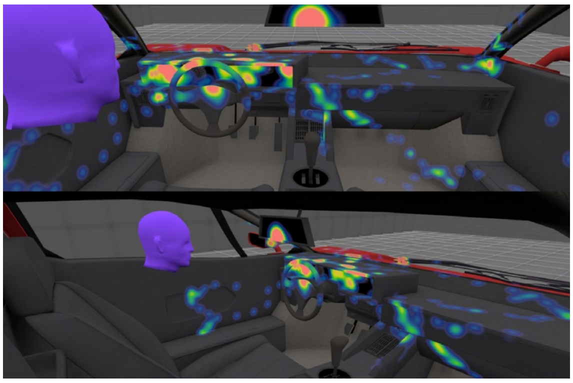

Fig. 8. Heatmap of the 1st round Cognitive3D test environment.

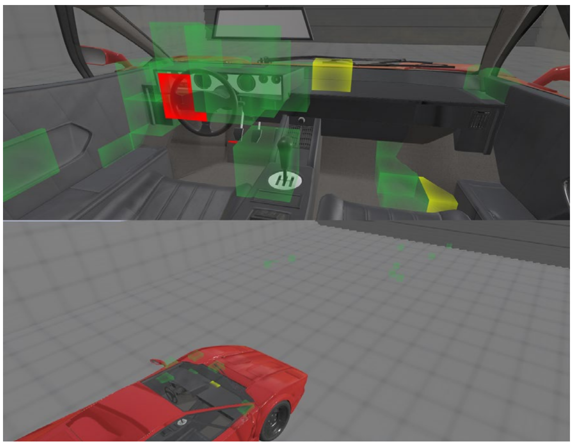

Fig. 9. Voxel-style heatmap in the Cognitive3D test environment.

The heatmap failed because the collision mesh was incomplete, showing how gaze analytics depend on scene configuration. Within the proposed artifact, collision validation becomes a mandatory protocol step, as unverified geometry can silently distort heatmap interpretation.

Additionally, Cognitive3D provided an aggregated view across sessions, dividing the space into coloured cubes in which warmer colours highlight regions of higher engagement (Fig. 9).

These preliminary configuration steps illustrate the operational layer of the proposed artifact, in which environment preparation, asset optimization, and collision validation are formalized as mandatory protocol stages rather than ad hoc technical adjustments. The demonstration phase revealed how minor configuration inconsistencies may propagate into analytical distortions, reinforcing the necessity of structured infrastructure governance.

B. *Immersive In-Vehicle Usability Assessment*

For the usability assessment, we created a simplified city environment using modular assets derived from the Modern City Downtown Megapack. Due to initial performance limitations, the scene was optimised by restricting the environment to two modular building types with varied façades. This configuration ensured stable frame rates and enabled the upload of the environment to Cognitive3D (Fig. 10).

Fig. 10. Environment for the 2nd round experiment.

We then combined the new city environment with the selected virtual vehicle model, the Cognitive3D eye tracking integration, and the physical cockpit. We manually adjusted the user's VR position to align the physical cockpit steering wheel with the virtual one (Fig. 11).

Subsequently, we conducted the in-vehicle usability assessment with participants, and their eye tracking data were collected. Each participant was immersed in the virtual cockpit for approximately one minute using the PICO 4 Enterprise head-mounted display. They were instructed to perform four tasks designed to evaluate different aspects of vehicular interaction:

(1) Simulate a lane change by observing the virtual mirrors, evaluating situational awareness;
(2) Interact with the air conditioner control panel, testing the identification and handling of secondary controls;
(3) Perform a saccadic eye movement toward the car's speedometer, analysing rapid visual search for critical information;
(4) Maintain attention on the road ahead, establishing a baseline for visual behaviour.

The next step involved generating and analysing heatmaps and extracting comfort and performance metrics from Cognitive3D.

Fig. 11. Aligning the physical and the virtual steering wheel.

The Presence metrics were interpreted according to the attributes defined by Cognitive3D. Because the study focused on gaze behavior, the Dynamic Engagement parameter scored 0, since participants did not interact with objects in the environment. The Boundary parameter scored 100, indicating that users remained within the designated physical space. The Orientation parameter reached 71, influenced by the urban setting, which introduced additional visual stimuli even though driving tasks naturally limit head movement. The Spatial Coverage parameter scored 64; despite the larger environment, we restricted exploration by the stationary cockpit setup. The Controller Engagement parameter scored 34, as arm movement was limited to simulating steering and dashboard interaction, without additional interactive elements. Fig. 12 summarises these results.

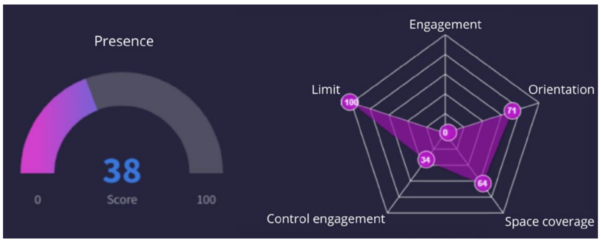

Fig. 12. Cognitive3D presence metrics.

Regarding the Comfort metrics, the Application Performance parameter maintained a 72 Hz refresh rate, ensuring smooth rendering. The Orientation parameter reproduced the same behaviour observed in the presence analysis. The Ergonomics parameter scored 20 points, as participants extended their arms to simulate steering-wheel and dashboard interaction, causing muscular fatigue and reducing comfort. Fig. 13 illustrates these values.

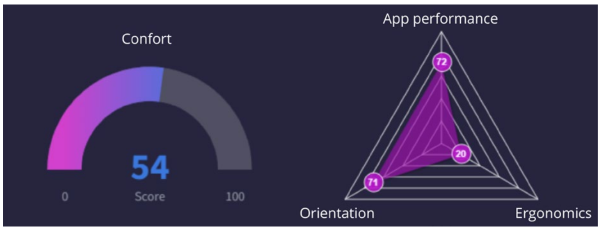

Fig. 13. Cognitive3D comfort metrics.

The Presence and Comfort results should therefore be read as platform-generated classifications of session behaviour rather than as direct observations of participants' internal states. At the empirical level, the study captured gaze, posture, controller, and performance traces. At the algorithmic level, Cognitive3D converted these heterogeneous traces into subdimensions and composite scores according to proprietary rules. At the interpretive level, we relate those outputs cautiously to usability conditions in the cockpit. Thus, a Dynamic Engagement value of 0 reflects the platform's encoding of limited object interaction under the present task structure, not an absence of immersion per se; a Boundary value of 100 indicates compliance with the tracked physical area, not a complete account of embodiment or comfort; and Comfort-related values combine performance telemetry and ergonomic heuristics rather than participants' explicit comfort judgments. These outputs are therefore used as decision-support indicators that help identify usability-relevant patterns and support multimodal comparison with questionnaire data, while avoiding stronger claims about construct equivalence, platform validity, or direct measurement of presence or comfort.

C. Comparing Cognitive3D and IPQ Presence Values

Participants answered the IPQ questionnaire to report their subjective perception of presence. Table I shows the average subscale scores. Overall, participants felt effectively situated in the immersive environment, although did not perceive the experience as highly realistic.

TABLE I. AVERAGE IPQ SUBSCALE SCORES ACROSS PARTICIPANTS

IPQ Subscale	Average Score
General Presence (G)	3.90
Spatial Presence (SP)	3.52
Involvement (INV)	3.15
Experienced Realism (REAL)	2.75

Comparisons between the IPQ scores and the Cognitive3D presence metrics revealed variations across participants (Fig. 14), showing that subjective and system-derived measures do not always align.

A consistent pattern emerged: General Presence and Spatial Presence scored highest, suggesting that participants felt adequately embedded in the virtual cockpit. In contrast, Experienced Realism received the lowest evaluations, reflecting limitations in the 3D model and indicating the need for more detailed assets when realistic driving simulation is required.

Variability among participants was also evident. For example, P6 recorded the highest presence scores across all IPQ dimensions, while P3 showed the lowest, indicating notable individual differences in perceived immersion.

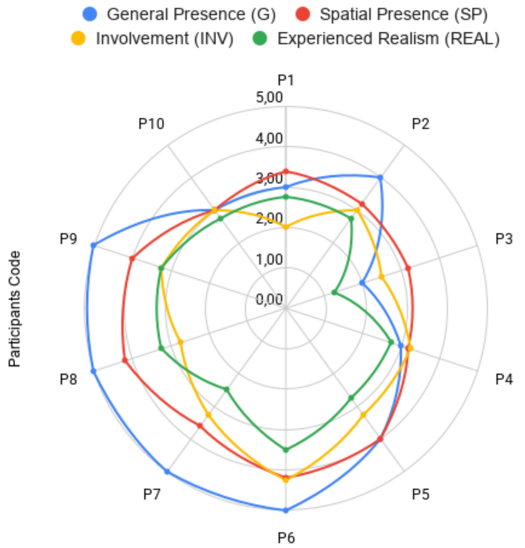

Fig. 14. Perceived presence profiles by participant, comparing the dimensions of General Presence, Spatial Presence, Involvement, and Experienced Realism.

In some cases, IPQ ratings exceeded those from Cognitive3D, whereas in others the opposite occurred. Although both instruments produced values within similar ranges (approximately 40–70), the difference per participant could reach around 19 points. Table II summarises the variables used to compare the two instruments.

TABLE II. VARIABLES USED TO COMPARE IPQ AND COGNITIVE3D PRESENCE SCORES

Variable	Description
Total Presence Score (Likert Score)	Average of the four IPQ dimensions: General Presence, Spatial Presence, Involvement, and Experienced Realism.
Total Presence Score Convert	IPQ score scaled to a 0-100 range to enable direct comparison with Cognitive3D metrics.
Presence Indicated in Cognitive3D	Platform-derived presence indicator calculated from gaze patterns, interaction behaviour, and spatial performance data.
Questionnaire vs. Cognitive3D	Absolute difference between the converted IPQ score and the Cognitive3D presence score for each participant.

Pearson's correlation analysis indicated a moderate positive relationship between IPQ scores and Cognitive3D metrics ($r = 0.60$). The p-value (0.067) was slightly above 0.05, likely influenced by the small sample size (Fig. 15). Overall, the trend suggests that the two measures tend to vary in the same direction, but participant-specific differences remain relevant. In other words, the observed pattern suggests directional convergence between the subjective and platform-derived measures, but given the reported significance level, it should not be interpreted as evidence of a statistically meaningful association.

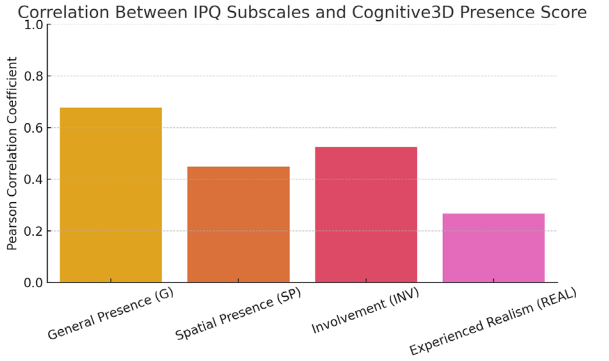

Fig. 15. Correlation between IPQ subscales with Cognitive3D presence score.

From the artifact perspective, the IPQ operates as a subjective benchmarking layer within the protocol. Rather than aiming to replace questionnaire-based presence measures, the analytics infrastructure is evaluated in terms of convergence, divergence, and interpretative complementarity. This cross-layer comparison reinforces the need for multimodal validation when relying on proprietary XR analytics. Accordingly, any convergence between IPQ and Cognitive3D should be interpreted as a relationship between a validated self-report measure and a platform-specific inferential indicator, not as equivalence between two measurements of the same construct.

Among the IPQ subscales, General Presence showed the strongest correlation with Cognitive3D (≈ 0.68), indicating that the overall sense of "being there" is more closely linked to gaze-derived indicators. Experienced Realism presented the weakest correlation, suggesting that visual fidelity is not fully captured by gaze analytics alone. Involvement and Spatial Presence showed moderate correlations.

The convergence and divergence patterns between subjective and system-derived presence indicators support the multimodal validation principle embedded in the artifact.

D. Usability Issues Detected and Design Implications

The usability tests also exposed design flaws in the vehicle model. Beyond the occluded mirror, the urban scenario introduced contextual distractions that influenced gaze distribution, producing behaviours similar to real driving. Participants frequently alternated their attention between mirrors, dashboard components, the speedometer, and surrounding traffic cues. Figs. 16 and 17 illustrate these patterns through heatmaps and aggregated gaze analyses generated by Cognitive3D.

The identified usability issues can be grouped into two categories: (i) structural interface problems, such as mirror occlusion and dashboard visibility constraints; and (ii) contextual attention competition effects introduced by environmental stimuli. This categorization supports a more systematic interpretation of gaze-derived usability evidence within immersive simulations.

Within this protocol, these findings show how structured gaze analytics, when governed by validated collision mapping and version-controlled metrics, can reveal latent ergonomic inconsistencies that might remain unnoticed in traditional static prototyping workflows. This reinforces the artifact's contribution as a replicable immersive usability assessment mechanism.

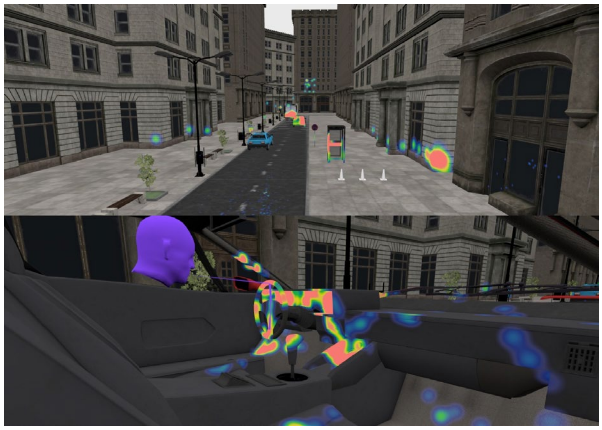

Fig. 16. Heatmap of the usability testing.

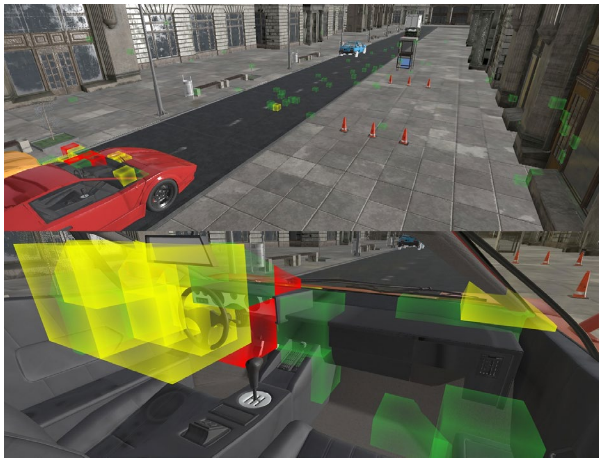

Fig. 17. Aggregated Cognitive3D analysis.

Table III summarizes the main usability issues and recommended actions. Rather than reporting the results only as case-specific observations, the table organizes them according to four practitioner-relevant dimensions: the observed issue, the corresponding evidence source, its practical interpretation, and the recommended action. This structure helps readers move from descriptive findings to operational design decisions, particularly in situations where immersive analytics are used to inform iterative improvements in interface layout, visibility, ergonomics, and attention management.

The table should be read as an application-oriented synthesis of the multimodal evidence generated in this study. In particular, it shows how gaze-based analytics, subjective measures, and technical validation procedures can jointly inform usability diagnosis in immersive HCI environments. By making these relationships explicit, the table also reinforces one of the central arguments of this paper: the value of commercial XR analytics lies not only in data capture and visualization, but in their integration into a structured interpretative workflow that supports more auditable and repeatable decision-making in other immersive and XR-based systems.

TABLE III. PRACTITIONER-ORIENTED INTERPRETATION OF IMMERSIVE USABILITY FINDINGS

Observed issue	Evidence source	Practical interpretation	Recommended action
Mirror occlusion or reduced visibility in critical viewing zones	Heatmaps, gaze trajectories, AOI dwell patterns	Safety-relevant elements may be partially blocked or poorly positioned in the immersive cockpit layout	Redesign mirror geometry, position, or surrounding visual structure; repeat validation after adjustment
Dashboard visibility constraints	Gaze concentration, re-peated search behavior, delayed task completion	Information layout may require excessive visual search effort or impose unnecessary attentional demand	Reorganize information hierarchy, improve salience of key indicators, and reassess scanning patterns
Contextual competition for attention	Distributed gaze across multiple interface elements during task execution	Interface elements may compete for attention in ways that reduce efficiency or increase cognitive load	Simplify display composition, reduce nonessential visual competition, and retest under the same task conditions
Low perceived realism	Subjective presence questionnaire results	Limited environmental fidelity may weaken ecological validity for later-stage evaluation or stake-holder demonstration	Increase visual detail, interaction fidelity, and contextual cues when the goal is decision-grade realism
Low ergonomic comfort	Self-reported comfort/usability responses and task-related behavior	Physical alignment between user posture and cockpit interaction zones may be suboptimal	Review steering wheel, dashboard, and control reach zones; adjust spatial alignment of the mockup and virtual scene
Inconsistency between questionnaire responses and platform-derived indicators	Comparison between subjective measures and proprietary analytics outputs	Single-source interpretation may be insufficient or potentially mis-leading	Use multimodal benchmarking and interpret analytics as support signals rather than standalone conclusions
Metric instability across software or platform versions	Technical logs, export differences, changed dashboard behavior	Longitudinal comparability may be threatened by platform dependency	Log all versions systematically and archive raw outputs independently whenever possible
Distorted heatmap interpretation caused by scene configuration issues	Collision validation and AOI inspection	Apparent usability problems may reflect data-capture configuration errors rather than real user behavior	Perform mandatory pre-validation of collision meshes, AOIs, and tagging before formal sessions

E. Practical Deployment Protocol for Immersive HCI Evaluation Using XR Analytics

To make the study easier to reuse in practice, the main methodological and empirical lessons were consolidated into the practical deployment protocol shown in Table IV.

These steps provide a practical guide for conducting immersive usability evaluations more consistently in future studies and applications. More specifically, the protocol offers a repeatable workflow for immersive in-vehicle HCI assessment, including mandatory validation of collision meshes and areas of interest before testing, systematic logging of hardware and software versions, multimodal comparison between platform-derived and subjective evidence, and independent archival of exported data for future verification. In this way, the protocol strengthens the auditability, repeatability, and decision value of immersive usability assessments in modern technological systems.

TABLE IV. PRACTICAL DEPLOYMENT PROTOCOL FOR IMMERSIVE HCI EVALUATION USING XR ANALYTICS

Step	Procedure	Purpose
1	Define the assessment objective and the main usability questions to be examined.	Establish a clear evaluation focus, such as visibility, reachability, attention distribution, interaction comfort, or interface search effort.
2	Prepare the virtual environment and the physical setup with sufficient fidelity and spatial alignment.	Ensure meaningful interaction and behavioral capture by aligning the mockup with the virtual representation and preserving adequate visual realism.
3	Validate collision meshes, tagged objects, and areas of interest before formal data collection.	Prevent configuration errors that may distort heatmaps, gaze attribution, and other analytics outputs, leading to incorrect usability interpretations.
4	Document all relevant technical dependencies, including headset model, firmware version, game engine version, eye tracking SDK, analytics platform version, and export settings.	Improve reproducibility and reduce the risk that future software or platform changes affect comparability across studies.
5	Conduct structured participant tasks directly linked to the evaluation goals.	Ensure that gaze patterns and interaction behaviors can be interpreted in relation to specific usability demands rather than generic exploratory activity.
6	Interpret system-generated analytics together with subjective measures, such as questionnaire responses.	Strengthen the analysis through multimodal evidence and reduce overreliance on platform-generated indicators alone.
7	Treat analytics outputs related to con-structs such as presence, engagement, or attention quality with caution.	Use these outputs as decision-support indicators rather than direct psychological measures.
8	Archive raw exports, metadata, and con-figuration records independently from the commercial platform whenever possible.	Improve traceability and enable future verification or reanalysis if needed.

F. Cognitive3D Versus Raw Eye Tracking Data: Practical Benefits and Governance Risks

Cognitive3D is described as a spatial analytics platform designed to capture 3D environments, device inputs, and user activity, providing visual and searchable insights for XR training, simulation, and research [28]. In automotive usability studies, this type of tool offers advantages over handling raw eye tracking data, particularly by simplifying access to gaze-based behavioural indicators.

Eye tracking provides essential information for in-vehicle usability evaluation, revealing attention distribution, cognitive processing demands, and potential interface issues. Metrics such as fixation duration, saccade patterns, and saccadic latency are relevant for assessing workload and visual navigation efficiency [27]. However, raw gaze data consist of low-level coordinates, demanding extensive preprocessing and specialised analytical skills.

Cognitive3D converts these data into interpretable visual outputs facilitating the identification of overlooked controls, confusing layouts, or ineffective spatial organization inside virtual cockpits. By projecting gaze vectors directly onto 3D models, the platform contextualizes user attention in relation to specific components (e.g., dashboard, mirrors, infotainment display), supporting ergonomics, safety analysis, and early-stage design decisions more efficiently than raw sensor logs.

While raw data represent gaze as 2D coordinates that are difficult to interpret spatially, Cognitive3D maps gaze and interaction data onto the actual 3D immersive environment, aligning attention with specific vehicle components, such as the dashboard or infotainment display. This enables ergonomists to assess real-world relevance, for example, how often users check mirrors or miss hazard indicators, both critical aspects for safety and ergonomic design. This type of data is especially valuable in the automotive industry, as it informs material optimization, identification of nonintuitive interface elements, and detection of deviations during task execution [27].

Additionally, Cognitive3D integrates metrics beyond gaze, offering insights into parameters such as Presence, spatial movement, and interaction behaviour (e.g., controller use, dwell time, click heatmaps). Instead of requiring manual interpretation, the platform automatically maps visual attention onto vehicle components, generates actionable heatmaps, and produces behaviour-driven usability indicators. These functionalities significantly accelerate iterative cockpit design and support the early detection of interface issues.

Table V summarizes the comparison between Cognitive3D and raw eye tracking data.

TABLE V. COMPARISON BETWEEN COGNITIVE3D AND RAW EYE TRACKING DATA

Aspect	Cognitive3D (Tool)	Raw Data (Source)
Type	Interpretable analytics platform	Unprocessed sensor output
Abstraction level	High-level, user-centric metrics (e.g., attention, presence, engagement)	Low-level logs (e.g., gaze vectors, times-tamps, coordinates)
Usability	Designed for researchers, designers, UX professionals	Requires data scientists or technical specialists
Function	Synthesizes, visualizes, and contextualizes behaviour in 3D space	Records raw sensor input without interpretation
Goal	Support decision-making and iterative design	Enable custom, flexible processing pipelines
Output	Dashboards, heatmaps, behaviour metrics, visual analytics	CSV files, coordinate logs, time-series data
Analogy	Similar to using Google Analytics to under-stand website traffic	Similar to parsing server logs line by line

Despite these advantages, reliance on proprietary analytics introduces challenges. High-level metrics may change due to platform updates, and internal calculation methods are often not transparent. This limits reproducibility and can compromise longitudinal analyses. During the study, an update modified the definitions and weightings of Cognitive3D's Presence components, preventing compatibility with previously collected data and interrupting ongoing analyses.

In automotive usability contexts, such instability can affect iterative design cycles, reduce comparability across prototypes, and diminish confidence in immersive evaluation pipelines. When metric definitions shift unexpectedly, decisions based on behavioural evidence may become inconsistent, representing a critical limitation of proprietary analytics tools.

When commercial XR analytics are used without governance controls, several failures can occur. Incomplete collision meshes or AOI tags may generate false heatmap patterns and apparent usability problems that are actually configuration artifacts. Unlogged platform updates may silently redefine metrics and compromise longitudinal comparability. Proprietary aggregations may also invite construct overreach when inferred scores are treated as direct psychological measurements, and cloud-based dependence may reduce auditability and data portability when independent exports are not maintained. In safety-relevant vehicle HMI evaluation, these failures may misdirect redesign efforts or create unwarranted confidence in cockpit adequacy [12, 13].

The findings reinforce the central premise of the proposed artifact: commercial XR analytics platforms should not be treated merely as visualization tools, but as interpretative infrastructures whose internal logic directly shapes research outcomes. Without structured governance mechanisms (including version logging, metric traceability, and raw-data archival), immersive usability evaluation risks becoming dependent on shifting computational definitions beyond the researcher's control.

From an HCI practice perspective, the value of commercial XR analytics lies in accelerating interpretation and supporting earlier design decisions; however, this value is sustainable only when accompanied by governance mechanisms that preserve traceability, comparability, and independent verification.

G. Implications for HCI in Modern Technological Systems

Although we conducted this study in an automotive cockpit context, its implications extend to Human-Computer Interaction in modern technological systems more broadly. In the automotive domain, immersive HCI evaluation overlaps with Human-Machine Interface (HMI) evaluation because interface decisions affect driver attention, operational safety, and trust in increasingly connected vehicle systems [12, 13]. Accordingly, cockpit usability problems such as mirror occlusion, visual competition, or poor control salience are not only UX concerns but also HMI risks with potential safety and security implications.

The contribution of this study can be read at three levels. Theoretically, it frames commercial XR analytics as interpretative infrastructures that condition what can be observed and claimed in immersive HCI evaluation. Rather than working only with raw behavioral signals, researchers increasingly rely on higher-level indicators generated by proprietary platforms. These abstractions accelerate analysis, but they also make HCI conclusions partially contingent on external metric definitions, weighting schemes, and update cycles.

More broadly, this distinction clarifies how proprietary analytics should be positioned in immersive HCI/HMI research. What is empirically measured is the behavioural and system-level trace; what is algorithmically inferred is the platform score produced from that trace; and what is theoretically interpreted is the researcher's claim about constructs such as presence, comfort, attention, or engagement. Treating these levels as interchangeable risks construct overreach and can make design conclusions appear more certain than the evidence warrants. The methodological value of commercial XR analytics lies instead in providing structured summaries of complex multimodal data that can guide diagnosis and redesign when embedded in a governance process. Such a process must preserve traceability across all three layers through collision and AOI validation, version logging, independent archival of exports, and benchmarking against subjective and contextual evidence. Under this view, proprietary metrics are useful not because they directly measure psychological constructs, but because they offer platform-generated scores whose decision value depends on disciplined interpretation and governance.

Methodologically, the artifact provides a governance-oriented protocol for immersive evaluation. The study showed that gaze-based spatial analytics can reveal visibility constraints, attention competition, and ergonomic inconsistencies, but also that these outputs should be interpreted together with subjective and contextual evidence rather than as self-sufficient ground truth. This reinforces the need for multimodal workflows that formalize collision and area-of-interest validation, version logging, cross-method benchmarking, and independent archival of exported data.

Practically, the framework offers a reusable workflow for early-stage cockpit assessment and decision support. For HCI/HMI practitioners, its value lies in enabling earlier diagnosis of interface-layout, visibility, ergonomics, and attention-management problems while reducing the risk that redesign decisions are based on opaque or unstable analytics alone. In safety-relevant vehicle evaluation, governance is therefore not an accessory to immersive testing; it is part of validity management.

These implications suggest that HCI in modern technological systems should increasingly be understood as an interaction not only between users and interfaces, but also between users, interfaces, sensors, and interpretative analytics layers. Accordingly, the contribution of this study goes beyond the automotive domain by showing that immersive usability evaluation with commercial XR analytics must be accompanied by transparent procedural and governance mechanisms if it is to support reliable, auditable, and reusable design decisions.

IV. Conclusion

This study proposed and demonstrated a Design Science Research artifact consisting of a structured immersive usability evaluation protocol combined with a governance framework for integrating proprietary XR analytics platforms into automotive research workflows.

This work provides a reusable protocol for integrating XR analytics into immersive usability evaluation. Its main value lies in formalizing the procedural conditions under which proprietary analytics can support decision-making without being treated as self-evident ground truth. For practitioners and researchers, the framework offers a structured path toward more auditable, repeatable, and methodologically transparent immersive evaluation in automotive applications and other modern technological systems.

The operationalization of the artifact showed that integrating virtual reality and embedded eye tracking technologies can provide actionable insights into cockpit usability, including attention distribution patterns, ergonomic inconsistencies, and contextual visual competition effects. At the same time, the study demonstrated that gaze-derived high-level metrics are not neutral psychological measures but infrastructure-dependent constructs shaped by computational abstraction layers.

A critical finding concerns the structural dependency introduced by proprietary analytics platforms. Metric definitions, weighting schemes, and internal processing models may change without researcher control, directly affecting longitudinal comparability and interpretative stability. This reinforces the necessity of governance mechanisms, including version logging, collision validation procedures, and raw data archival, as embedded components of immersive usability protocols. This reinforces the need to treat version logging, collision validation, and raw-data archival as core components of immersive usability protocols using commercial XR analytics platforms.

The limitations of this study should be considered systematically. First, construct validity is constrained because Cognitive3D's Presence and Comfort scores are proprietary aggregations; within this study they can be treated only as platform-derived indicators, not as transparent measurements of psychological constructs. Second, the empirical instantiation depends on a single commercial platform and its cloud-based infrastructure, which limits auditability, data portability, and direct comparison with alternative analytics ecosystems. Third, several governance controls recommended by the artifact were not fully implemented during this demonstration. Firmware and SDK versions were not logged systematically, collision and AOI coverage were refined during exploratory use rather than fully pre-validated before the formal sessions, and raw exports were not independently archived outside the platform. These omissions may affect gaze attribution, reduce traceability, limit independent verification, and weaken strict reproducibility. Fourth, the simplified scene and small sample size constrain ecological realism and statistical generalization. Taken together, these limitations reduce the evidential strength of the present empirical instantiation and reinforce why these controls should be treated as mandatory governance requirements in future studies rather than optional technical refinements.

Future research should therefore pursue more specific confirmatory tests, including matched multi-platform comparisons under the same cockpit tasks, larger confirmatory samples, direct benchmarking of proprietary outputs against transparent raw-data pipelines, and longitudinal studies using frozen software baselines to test whether version stability improves comparability in safety-relevant HMI evaluation. It would also be valuable to apply the protocol to cockpit scenarios with greater visual realism and to connected-vehicle HMI functions in which usability, safety, and security concerns are more tightly coupled.

Ultimately, this study reframes commercial XR analytics tools not merely as visualization utilities, but as interpretative infrastructures that require version control, validation steps, and data logging. By embedding governance principles into immersive evaluation workflows, the proposed artifact contributes to more robust, transparent, and scalable automotive usability research in virtual environments.

Supplementary Material

The following supporting information can be downloaded at https://osf.io/zvdfa/overview: (1) ExperimentMetadata_participants—participant metadata used in the experiment; and (2) ExperimentMetadata_IPQvalues—IPQ assessment results and corresponding values.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization, F.D. and R.M.; methodology, Y.T.S.F. and I.W.; software and immersive environment development, F.D. and R.M.; data collection and investigation, F.D., R.M., C.R., D.R., J.M.T., and S.S.; data analysis, F.D., R.M. and Y.T.S.F.; writing—original draft preparation, F.D. and R.M.; writing—review and editing, Y.T.S.F. and I.W.; supervision, I.W. All authors have read and agreed to the published version of the manuscript.

Acknowledgment

The authors gratefully acknowledge the support of the Human-centric Industrial Immersive Engineering Laboratory (HIIVE Lab)/ SENAI CIMATEC University, the Advanced Knowledge Center in Immersive Technologies (AKCIT), the Institute of Science, Innovation and Technology (INCITE) Industry 4.0, the Bahia State Research Support Foundation (FAPESB), and the National Council for Scientific and Technological Development (CNPq) for their support; IW is a CNPq fellow (303192/2023-2).

We are also grateful to Kevin Morey, Caetano Medeiros and Kyle Stittleburg for their valuable comments and suggestions on earlier drafts of this manuscript.

References

[1] International Organization for Standardization (ISO). (2018). ISO 9241-11:2018 ergonomics of human-system interaction—Part 11: Usability: Definitions and concepts. [Online]. Available: https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en

[2] F. V. de Freitas, "Virtual reality applied to usability testing: A method for product evaluation in the automotive industry," Master's Thesis, SENAI CIMATEC, Salvador, Brazil, 2025. (in Portuguese)

[3] C. Snyder, *Paper Prototyping: The Fast and Easy Way to Design and Refine User Interfaces*, San Francisco, CA, USA: Morgan Kaufmann, 2003.

[4] K. Hornbæk, "Current practice in measuring usability: Challenges to usability studies and research," *International Journal of Human-Computer Studies*, vol. 64, no. 2, pp. 79–102, 2006.

[5] M. Slater and M. V. Sanchez-Vives, "Enhancing our lives with immersive virtual reality," *Frontiers in Robotics and AI*, vol. 3, 74, 2016.

[6] F. Freitas, H. Oliveira, I. Winkler, and M. Gomes, "Virtual reality on product usability testing: A systematic literature review," in *Proc. 2020 22nd Symposium on Virtual and Augmented Reality (SVR)*, IEEE, 2020, pp. 67–73.

[7] D. A. Bowman, R. P. McMahan, and E. D. Ragan, "Questioning naturalism in 3D user interfaces," *Communications of the ACM*, vol. 55, no. 9, pp. 78–88, 2012.

[8] P. Caserman, A. Garcia-Agundez, and S. Göbel, "A survey of full-body motion reconstruction in immersive virtual reality applications," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 10, pp. 3089–3108, 2020.

[9] V. C. Vijay, K. R. S., G. Zhong, P. Kumar, and M. Zhao, "Realtime Artificial Intelligence (AI) and Augmented Reality (AR)-driven virtual saree dressing: Mirroring reality in fashion technology," *Journal of Advances in Information Technology*, vol. 16, no. 3, pp. 426–433, 2025.

[10] Y. Xiao, N. Funabiki, I. T. Anggraini, H. Fujiwara, J.-Y. Liu, and C.-P. Fan, "An exergame system for tenosynovitis prevention by softening thumb base," *Journal of Advances in Information Technology*, vol. 16, no. 10, pp. 1414–1422, 2025.

[11] J. Y. Wong, A. B. Azam, Q. Cao, L. Huang, Y. Xie, I. Winkler, and Y. Cai, "Evaluations of virtual and augmented reality technology-enhanced learning for higher education," *Electronics*, vol. 13, no. 8, 1549, 2024.

[12] I. Grobelna, D. Mailland, and M. Horwat, "Design of automotive HMI: New challenges in enhancing user experience, safety, and security," *Applied Sciences*, vol. 15, no. 10, 5572, 2025.

[13] Z. N. Aldoski, "Virtual reality for autonomous vehicles: A review of safety, training, and human–machine interaction," *Dasinya Journal for Engineering and Informatics*, vol. 1, no. 1, 2025.

[14] F. V. de Freitas, M. V. M. Gomes, and I. Winkler, "Benefits and challenges of virtual-reality-based industrial usability testing and design reviews: A patents landscape and literature review," *Applied Sciences*, vol. 12, no. 3, 1755, 2022.

[15] J. Jerald, *The VR Book: Human-Centered Design for Virtual Reality*, New York, NY, USA: Association for Computing Machinery and Morgan & Claypool, 2015.

[16] L. G. G. Almeida, N. V. de Vasconcelos, I. Winkler, and M. F. Catapan, "Innovating industrial training with immersive metaverses: A method for developing cross-platform virtual reality environments," *Applied Sciences*, vol. 13, no. 15, 8915, 2023.

[17] J. L. Gabbard, "Usability engineering for virtual environments," *Presence: Teleoperators and Virtual Environments*, vol. 8, no. 5, pp. 534–548, 1999.

[18] R. M. C. Leite, I. Winkler, and L. R. G. Alves, "Visual management and gamification: An innovation for disseminating information

about production to construction professionals," *Applied Sciences*, vol. 12, no. 11, 5682, 2022.
[19] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. van de Weijer, *Eye Tracking: A Comprehensive Guide to Methods and Measures*, Oxford: Oxford University Press, 2011.
[20] K. Kerstens, H. Van Kerrebroeck, K. Luyten, and K. Coninx, "The value of eye tracking and think-aloud in usability testing," *Journal of Usability Studies*, vol. 13, no. 4, pp. 207–228, 2018.
[21] J. H. Goldberg and A. M. Wichansky, "Eye tracking in usability evaluation: A practitioner's guide," in *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, J. Hyönä, Radach, and H. Deubel, Eds. Amsterdam: North-Holland, 2003, ch.13, pp. 493–516.
[22] M. Schiessl, S. Duda, A. Thölke, and R. Fischer, "Eye tracking and its application in usability and media research," *MMI Interaktiv*, vol. 6, pp. 41–50, Jan. 2003.
[23] S. Negi and R. Mitra, "Fixation duration and the learning process: An eye tracking study with subtitled videos," *Journal of Eye Movement Research*, vol. 13, no. 6, 1, 2020.
[24] J. Moreno-Arjonilla, A. López-Ruiz, J. R. Jiménez-Pérez, J. E. Callejas-Aguilera, and J. M. Jurado, "Eye-tracking on virtual reality: A survey," *Virtual Reality*, vol. 28, no. 1, 38, 2024.
[25] A. Riegler, A. Riener, and C. Holzmann, "A systematic review of virtual reality applications for automated driving: 2009–2020," *Frontiers in Human Dynamics*, vol. 3, 689856, 2021.
[26] A. Calvi, F. D'Amico, and A. Vennarucci, "Comparing eyetracking system effectiveness in field and driving simulator studies," *The Open Transportation Journal*, vol. 17, no. 1, e187444782301191, 2023.
[27] C. Ramos, R. Miguez, F. Leão, D. Rodrigues, M. Demetrescu, F. Sandrin, M. Alfonso, R. Leite, and I. Winkler, "Enhancing automotive usability testing and user experience: Insights from virtual reality and eye tracking integration," in *Proc. 2025 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR)*, IEEE, 2025, pp. 429–434.
[28] Cognitive3D. Cognitive3D Official Website. [Online]. Available: https: //cognitive3d.com/
[29] Y. Abeysinghe, K. Cauchi, V. Ashok, and S. Jayarathna, "Framework for measuring visual attention in gaze-driven VR learning environments using Meta Quest Pro," in *Proc. the 2025 Symposium on Eye Tracking Research and Applications*, ACM, 2025, pp. 1–3.
[30] A. Fotopoulos, E. Loukakis, and M. Xenos, "Evaluating pedestrian behavior in virtual reality for traffic education using eye tracking," in *Proc. International Conference on Human-Computer Interaction*, Springer, 2025, pp. 161–180.
[31] PNY Technologies. (2024). PICO 4 Enterprise. [Online]. Available: https: //www.pny.com/pico-4-enterprise
[32] J. W. Creswell and J. D. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 5th ed. Thousand Oaks, CA, USA: SAGE Publications, 2017.
[33] S. Gregor and A. R. Hevner, "Positioning and presenting design science research for maximum impact," *MIS Quarterly*, vol. 37, no. 2, pp. 337–355, 2013.
[34] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A design science research methodology for information systems research," *Journal of Management Information Systems*, vol. 24, no. 3, pp. 45–77, 2007.
[35] Epic Games. (2025). Unreal Engine: The Most Powerful Real-Time 3D Creation Tool. [Online]. Available: https://www.unrealengine.com/
[36] Igroup Project Consortium. (2025). Igroup Presence Questionnaire (IPQ) Overview. [Online]. Available: https://www.igroup.org/pq/ipq/index.php
[37] 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, 2001.
[38] J. J. Vasconcelos-Raposo, M. Bessa, M. Melo, L. Barbosa, R. Rodrigues, C. M. Teixeira, L. Cabral, and A. A. de Sousa, "Adaptation and validation of the Igroup Presence Questionnaire (IPQ) in a Portuguese sample," *Presence: Teleoperators and Virtual Environments*, vol. 25, no. 3, pp. 191–203, Dec. 2016.
[39] Y. T. S. Ferreira, A. F. d. S. Santos, M. M. Araujo, L. S. Lima, J. L. R. Neto, M. A. S. Santos, U. J. Camelier, L. M. d. A. Vale, and I. Winkler, "Assessing the usability and presence of an immersive virtual environment: Opportunities and challenges of combining SUS, IPQ and Nielsen's Heuristics," in *Proc. the 27th Symposium on Virtual and Augmented Reality (SVR)*, IEEE, 2025, pp. 272–281.