MASTERARBEIT | MASTER'S THESIS

Titel | Title

A Virtual Reality Simulator for Augmented-Reality-Assisted Driving

verfasst von | submitted by

Lun Yang

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2026

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 921
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Informatik
Betreut von | Supervisor:	Univ.-Prof. Dipl.-Ing. Dr. Helmut Hlavacs

Acknowledgements

I would like to sincerely thank my supervisor, **Univ.-Prof. Dr. Helmut Hlavacs**, for his continuous guidance, insightful suggestions, and encouragement throughout this thesis. His expertise and feedback have been invaluable. The conceptual design and research direction of this work emerged from our joint discussions.

I am deeply grateful to my family and friends for their patience, understanding, and constant support, which kept me motivated along the way.

My thanks also go to everyone who took part in the driving sessions and filled out the survey. Your input was essential to this research.

Use of AI Tools. During the preparation of this thesis, I made use of several AI-assisted tools, including ChatGPT, DeepSeek, Doubao, and Gemini Deep Research, to support specific aspects of the writing and research process. Their use is acknowledged here in accordance with good scientific practice.

In the *Related Work* chapter, AI tools were employed to assist in the initial literature search by identifying relevant paper URLs and to help organize and format the entries in the bibliography. The selection, critical evaluation, and synthesis of the cited literature remain my own work.

Throughout the thesis, AI tools were used to assist in rewording and refining my own sentences and paragraphs to improve clarity. The original ideas, logical flow, and technical arguments remain entirely my own. The chapter structure was developed based on my supervisor's guidance and the project reports I prepared during the development of the simulator.

The project code was developed under my supervisor's guidance. AI tools were consulted for implementation ideas and used to assist with repetitive coding tasks. All core logic, system architecture decisions, and experimental functionality were designed and implemented by me.

The experimental design, including the three-session within-subjects protocol, the selection of hazard scenarios, and the questionnaire instruments, was developed by me and is based on real data collected from human participants.

Abstract

This thesis presents the design, implementation, and evaluation of a Virtual Reality (VR) driving simulator integrated with an Augmented Reality (AR) hazard visualization system, developed to address the critical gap in current driver safety training: the lack of controlled, repeatable exposure to high-risk urban scenarios. The research is motivated by the observation that novice drivers, who lack the refined mental models required to anticipate hidden or sudden traffic hazards, are disproportionately involved in collisions caused by occluded pedestrians, blind-spot conflicts, and unexpected maneuvers by other road users.

The simulator is built on the Unity engine and features a five-layer decoupled architecture encompassing procedural city generation, a parameter-driven NPC traffic ecosystem, a scripted hazard event control system, real-time accident detection, and an AR-based user interface. The core technical innovation is an *X-Ray* rendering pipeline that reveals occluded pedestrians through opaque geometry, such as a stationary bus, by manipulating Unity's render-queue ordering, providing the driver with advance warning of hazards that would otherwise remain invisible until the moment of collision.

The evaluation is guided by two research questions. **RQ1** asks whether an AR system that reveals hidden or sudden traffic hazards significantly improves driver safety and reduces accident rates in complex urban conditions. **RQ2** asks whether AR-enhanced driving simulation can effectively foster drivers' long-term awareness of hidden or sudden traffic risks. To address both questions, a dual-mode experimental design is employed with $N = 24$ novice drivers. The VR group ($n = 12$) completes three consecutive driving sessions in a within-subjects protocol: a no-AR baseline (S1), an AR-assisted session (S2), and a second no-AR session (S3). This structure yields paired objective comparisons for both research questions: RQ1 is tested by comparing collision counts and Perception-Reaction Time (PRT) between S2 and S1, while RQ2 is tested by comparing the same measures between S3 and S1. The Video group ($n = 12$) watches a recorded playback of a representative Session 2 drive and completes the same post-experience questionnaire, providing a passive-observation baseline for subjective ratings of immersion, AR utility, and educational value.

Preliminary findings indicate that the AR-enhanced simulation significantly improves hazard avoidance and shortens PRT compared to unaided driving, and that a measurable portion of this improvement persists after the AR system is removed. Subjective ratings from the VR group exceed those of the Video group across all questionnaire dimensions. These results provide empirical evidence that AR-enhanced VR simulation can serve both as an effective real-time safety intervention and as a pedagogical tool for cultivating long-term hazard awareness in novice drivers.

Kurzfassung

Diese Masterarbeit präsentiert das Design, die Implementierung und die Evaluierung eines Virtual-Reality (VR)-Fahrsimulators, der ein Augmented-Reality (AR) Gefahrenvisualisierungssystem integriert. Die Forschung adressiert die kritische Lücke in der aktuellen Fahrerausbildung: das Fehlen kontrollierter, wiederholbarer Exposition gegenüber Hochrisikoszenarien im urbanen Verkehr. Die Motivation ergibt sich aus der Beobachtung, dass Fahranfänger, denen die verfeinerten mentalen Modelle zur Antizipation versteckter oder plötzlicher Gefahren fehlen, überproportional häufig in Kollisionen verwickelt sind, die durch verdeckte Fußgänger, Totwinkelkonflikte und unerwartete Manöver anderer Verkehrsteilnehmer verursacht werden.

Der Simulator basiert auf der Unity-Engine und verwendet eine fünfschichtige, entkoppelte Architektur, die prozedurale Stadtgenerierung, ein parametergetriebenes NPC-Verkehrsökosystem, ein gesteuertes Gefahrenereignissystem, Echtzeit-Unfallerkennung und eine AR-basierte Benutzeroberfläche umfasst. Die zentrale technische Innovation ist eine *Röntgen*-Render-Pipeline, die verdeckte Fußgänger durch undurchsichtige Geometrie, wie einen stehenden Bus, sichtbar macht, indem sie die Render-Reihenfolge von Unity manipuliert und dem Fahrer eine Vorwarnung vor Gefahren gibt, die andernfalls bis zum Kollisionsmoment unsichtbar blieben.

Die Evaluierung wird von zwei Forschungsfragen geleitet. **FQ1** untersucht, ob ein AR-System, das versteckte oder plötzliche Verkehrsgefahren durch intuitive Visualisierung aufdeckt, die Fahrersicherheit signifikant verbessert und die Unfallraten in komplexen städtischen Umgebungen reduziert. **FQ2** untersucht, ob AR-verbesserte Fahrsimulation effektiv das langfristige Gefahrenbewusstsein von Fahrern fördern kann. Zur Beantwortung beider Fragen wird ein duales Experimentaldesign mit $N = 24$ Fahranfängern eingesetzt. Die VR-Gruppe ($n = 12$) durchläuft drei aufeinanderfolgende Fahrsitzungen in einem Within-Subjects-Protokoll: eine Baseline ohne AR (S1), eine AR-gestützte Sitzung (S2) und eine zweite Sitzung ohne AR (S3). Diese Struktur liefert gepaarte objektive Vergleiche für beide Forschungsfragen: FQ1 wird durch den Vergleich von Kollisionszahlen und Wahrnehmungs-Reaktionszeit (PRT) zwischen S2 und S1 getestet, während FQ2 durch den Vergleich derselben Maße zwischen S3 und S1 getestet wird. Die Videogruppe ($n = 12$) sieht eine aufgezeichnete Wiedergabe einer repräsentativen Sitzung 2 und beantwortet denselben Fragebogen, der eine Passivbeobachtungs-Baseline für subjektive Bewertungen von Immersion, AR-Nutzen und pädagogischem Wert liefert.

Vorläufige Ergebnisse deuten darauf hin, dass die AR-verbesserte Simulation die Gefahrenvermeidung signifikant verbessert und die PRT im Vergleich zum Fahren ohne AR verkürzt, und dass ein messbarer Teil dieser Verbesserung auch nach der Deaktivierung des AR-Systems bestehen bleibt. Die subjektiven Bewertungen der VR-Gruppe übertreffen die der Videogruppe in allen Fragebogendimensionen. Diese Ergebnisse liefern empirische Belege dafür, dass AR-verbesserte VR-Simulation sowohl als effektive Echtzeit-Sicherheitsintervention als auch als pädagogisches Werkzeug zur Förderung des langfristigen Gefahrenbewusstseins bei Fahranfängern dienen kann.

Contents

List of Tables

List of Figures

Listings

1. Introduction

Driving safety remains one of the most critical challenges in modern transportation systems, representing a complex interplay between human factors, vehicle technology, and environmental variables. Despite the rapid development of passive safety features like crumple zones and active safety systems such as Electronic Stability Control (ESC), road traffic accidents continue to be a leading cause of mortality and economic loss globally. Research indicates that the vast majority of these incidents are not rooted in mechanical failure but in human cognitive limitations, particularly regarding the identification and processing of sudden or hidden hazards. Novice drivers, in particular, lack the refined mental models required to scan for latent risks in dense urban environments. The transition from a beginner to an expert driver is often a perilous journey, primarily because traditional driving education remains siloed between theoretical classroom learning and low-risk supervised driving. This pedagogical gap leaves drivers ill-equipped for high-stakes, split-second scenarios where the Time-to-Collision (TTC) is extremely limited.

A fundamental obstacle in urban safety is the *hidden pedestrian* problem, where physical occlusions such as parked buses, roadside infrastructure, or larger vehicles create visual blind spots that mask emerging threats. In these critical moments, the human visual system is hindered by physical limits. Even the most vigilant driver cannot react to what is fundamentally invisible. While modern Advanced Driver Assistance Systems (ADAS) have introduced sensors to detect such threats, the method of communicating this information to the driver remains suboptimal. Traditional dashboard icons or auditory beeps require a cognitive translation process, where the driver must move their gaze from the road to the instrument cluster, decode the signal, and then refocus on the external environment to locate the hazard. This redirection of visual attention, known as *eyes-off-the-road* time, introduces a lethal latency during which an accident often becomes inevitable. Consequently, there is an urgent need for a paradigm shift in how safety information is presented, moving away from abstract warnings toward integrated, intuitive visual extensions of the driver's own senses.

The primary motivation of this research is to transcend the limitations of traditional interfaces by transforming the vehicle windshield into a proactive, intelligent extension of the driver's sensory perception. At the core of this study is the development of a high-fidelity driving simulator that integrates an Augmented Reality (AR) HUD, coupled with a predictive hazard detection system. This system aims to bypass the cognitive translation delay by projecting safety information, specifically *X-Ray visualizations* of occluded objects, directly into the driver's primary field of view. By making the invisible visible, the system allows for a seamless perceptual fusion between virtual cues and the real-world geometry. The overarching goal is not merely to provide a real-time assistant, but to utilize this specialized platform as a pedagogical tool. By exposing participants to these predictive interventions, we seek to evaluate whether immersive simulation can fundamentally reshape a driver's internal risk-perception framework, moving their response pattern from reactive braking to proactive hazard anticipation.

To address these multifaceted challenges, this thesis proposes a modular driving simulator built on the Unity engine, featuring a closed-loop intervention mechanism from multi-source risk detection to AR-enhanced visual feedback. The research is anchored in a systemic approach that treats the driver, the vehicle, and the complex traffic environment as a single integrated unit. The technical architecture is designed for high extensibility, allowing for the simulation of diverse accident scenarios ranging from blind-spot conflicts to sudden lane-change intrusions. This leads to the formulation of two primary research questions (RQs) that guide the empirical investigation:

- **RQ1**: Does an AR system that reveals hidden or sudden traffic hazards through intuitive visualization measurably improve driver safety and reduce accident rates in complex urban conditions?
- **RQ2**: Can driving simulation, integrated with AR-enhanced hazard visualization, effectively foster and improve drivers' long-term awareness of hidden or sudden traffic risks?

Consistent with these research questions, the following hypotheses are tested throughout the study:

- **H1**: Drivers using AR-HUD assistance will exhibit fewer hazard collisions and shorter PRT compared to their own unaided baseline (S1 vs S2).
- **H2**: Participants will show sustained patterns of improved hazard avoidance performance (fewer collisions, shorter PRT) in a subsequent no-AR session after AR-assisted training, compared to their pre-training baseline (S1 vs S3).

The potential contribution of this work is twofold, encompassing both academic and practical dimensions. Academically, this research advances the field of Human-Computer Interaction (HCI) in automotive contexts by providing a quantitative analysis of how AR-based sensory extensions affect cognitive workload and situational awareness. It contributes to the growing body of literature on *instructional scaffolding* in driving pedagogy, offering new insights into how visual cues can be used to accelerate the acquisition of expert scanning behaviors. Practically, the developed simulator offers a scalable, low-cost platform for driving schools and automotive researchers to test safety interventions in a risk-free, dynamic urban environment. By providing a decoupled event-management system, the platform allows for the rapid creation of edge-case scenarios that are too dangerous to reproduce on real roads, such as high-speed intersection collisions involving vulnerable road users.

The remainder of this thesis is structured as follows. Chapter 2 provides a comprehensive review of the related work, situated at the intersection of driving simulation, AR-HUD technologies, and hazard perception psychology. Chapter 3 establishes the theoretical foundations, discussing the optical principles of AR imaging, the kinematics of vehicle motion, and the mathematical modeling of traffic flow. Chapter 4 details the technical implementation of the AR Driving Simulator, with specific focus on the dynamic city generation algorithms, the X-Ray vision shader logic, and the event-driven modeling of the traffic ecosystem. Chapter 5 presents the evaluation methodology and results, including a detailed analysis of the comparative experiments and a discussion of the statistical findings relative to the hypotheses. Finally, Chapter 6 concludes the work by summarizing the key contributions, acknowledging the study's limitations, and outlining future directions for the integration of AR into the broader landscape of intelligent transportation systems.

2. Related Work

This chapter provides a comprehensive review of the scholarly literature situated at the intersection of Augmented Reality (AR), Virtual Reality (VR) driving simulation, and traffic safety research. By systematically synthesizing relevant academic works, this discussion identifies the prevailing research problems, experimental methodologies, and core conclusions that define the current state of the art in the field. The review is structured around four primary dimensions: firstly, the construction of virtual traffic environments and the study of behavioral diversity. Secondly, the application of AR Head-Up Displays (AR-HUD) in visual perception. Thirdly, an exploration of cooperative hazard modeling and multi-source detection technologies. Finally, an analysis of driver hazard perception assessment and training effectiveness. This multi-dimensional exploration not only elucidates existing research gaps—particularly regarding culturally adaptive simulation and the cognitive quantification of AR interventions—but also provides a robust theoretical foundation for the development of the proposed system. Through a comparative analysis of prior work, this chapter aims to highlight the uniqueness of this research in advancing intuitive driving interaction and globalized driving safety education.

Parish and Müller [PM01] introduced a procedural modeling system based on L-systems to automate the generation of large-scale urban infrastructures and street networks. Their research demonstrated that grammar-based rules could successfully synthesize complex, realistic city layouts that reflect the structural logic of real-world Manhattan grids. However, this foundational work primarily focused on static geometric representations, leaving the real-time integration of dynamic traffic agents and event-based triggers for subsequent research.

Marshall [Mar04] established a taxonomic framework for street patterns, distinguishing between highly regulated "gridiron" (Manhattan-style) networks and non-rectilinear, organic structures. His research categorized the orthogonal grid as a highly connected topological system that facilitates predictable navigational flow through consistent 90-degree intersection geometry. The study concluded that while regular grids optimize routing efficiency and spatial orientation, they also impose rigid visual constraints that define how drivers perceive hidden risks at intersections. This morphological analysis provides the geometric justification for using a Manhattan layout to standardize occlusion variables, though the work primarily addressed urban planning rather than the real-time dynamics of virtual simulation.

Helbing et al. [HFV00] introduced the social force model for pedestrian dynamics, which often utilizes destination waypoints as the primary drivers of intentional movement. The research successfully simulated how agents navigate toward specific nodes while adjusting their velocity to avoid physical and social obstacles. While highly effective for group dynamics, the deterministic nature of their waypoint-following logic does not fully capture the irrational or culturally specific crossing behaviors observed in real-world urban centers.

Best et al. [BNCM16] investigated the real-time path planning and waypoint navigation of multi-agent vehicles within complex urban road networks. The study proposed a hybrid navigation algorithm that combines predefined waypoint guidance with Reciprocal Velocity Obstacles (RVO), allowing NPC vehicles to follow global traffic rules while dynamically adjusting velocity profiles to avoid local dynamic obstacles. However, the model's deterministic guidance mechanism based on waypoints can occasionally lead to unnatural "deadlocks" in high-density urban conflict scenarios, making it difficult to fully replicate the nonlinear emergency responses of human drivers in extreme blind-spot conflicts.

Smelik et al. [STKB14] conducted a comprehensive survey of procedural modeling techniques for virtual worlds, systematically categorizing approaches ranging from L-systems to grammar-based generation methods. Their review found that while procedural content generation (PCG) has matured for terrain and vegetation, the automatic synthesis of functional urban environments with embedded traffic logic remains an open challenge. The authors called for future work to bridge the gap between geometric procedural modeling and behavioral simulation, a need that the present study addresses by integrating PCG with waypoint-based traffic systems.

Arafat et al. [A$^+$25] developed a Local Spline Path Planning (LSPP) algorithm that combines rule-based waypoint generation with quintic spline curve fitting and model predictive control for self-driving cars. Their experimental results demonstrated that the waypoint-based approach outperformed benchmark methods (A*, RRT) in smoothness, safety, and computational speed. The authors acknowledged that their method focuses on highway rather than complex urban environments, leaving open the question of how waypoint networks should be designed to support multi-agent coordination in dense city traffic.

Zhu et al. [ZJW$^+$15] proposed a waypoint graph-based fast pathfinding algorithm designed for dynamic environments, where edge costs fluctuate due to moving obstacles or changing traffic conditions. Their method achieved significant computational savings compared to traditional A* search by precomputing waypoint connectivity and updating only localized segments of the graph. The researchers identified that their approach assumes a static waypoint topology, leaving the challenge of dynamic waypoint generation in procedurally expanding urban environments as an area for future investigation.

Liu et al. [LHT$^+$24] examined the integration of computer vision and procedural generation techniques for the 3D reconstruction of urban geometries. Their study highlighted that asynchronous generation pipelines—where geometry is created incrementally across multiple frames—can reduce the computational peaks associated with large-scale city synthesis. The authors concluded that while their methods are effective for static urban reconstruction, the real-time generation of interactive, agent-filled environments for driving simulation remains an open research direction, directly motivating the present work's coroutine-driven loading strategy.

Gupta and Velaga [GV25] validated a motorized two-wheeler virtual environment by examining the relationship between perceived realism, simulator fidelity, and user behavior. Their findings confirmed that higher physical fidelity—particularly in vehicle dynamics—improves the transferability of simulator-acquired skills to real-world scenarios. The study noted that while fidelity is crucial for behavioral validity, increased complexity must be balanced against computational performance, a trade-off that informed the present simulator's lightweight yet physically accurate vehicle dynamics model.

Gabbard et al. [GFK14] delineated the design space for automotive augmented reality (AR) by classifying display types (head-mounted, heads-up, center-mounted) and visual registration modes (world-fixed versus screen-fixed graphics). Their work systematically identified critical perceptual and cognitive challenges unique to windshield AR, including cognitive tunneling, the occlusion of real-world hazards by virtual overlays, and the need to balance information density with driver attention capacity. However, the study remained a conceptual framework and did not empirically validate how specific AR visualization techniques—such as the X-Ray pedestrian highlighting employed in the present simulator—affect driver reaction times or accident rates in complex, occlusion-heavy urban scenarios.

Sweller et al. [SAK11] detailed the principles of Cognitive Load Theory (CLT) as they apply to complex information processing environments. Their work established that while AR can reduce extrinsic load by integrating data into the visual path, it may simultaneously increase intrinsic load if the visual overlays are too visually complex. The authors suggested a framework for load management, yet they did not specify how these cognitive limits fluctuate under the pressure of sudden, high-risk traffic events.

Parasuraman and Manzey [PM10] explored the concepts of complacency and automation bias in human-machine interaction, with implications for AR-enhanced driving. Their research found that as systems become more reliable, human operators tend to decrease their manual monitoring, leading to a failure to detect system errors. The study warned that AR-HUDs could induce similar biases, though it did not specifically test how this behavior manifests in the context of persistent, long-term AR training.

Fisher et al. [FCHT17] synthesized research on novice driver training, emphasizing the need for tools that explicitly target hazard anticipation skills. Their findings indicated that traditional training often fails to provide enough exposure to "near-miss" scenarios, a gap that could be bridged by immersive simulations. Nevertheless, the handbook noted that the transition from simulator-acquired skills to real-world performance is still an area requiring more longitudinal data.

Lopez and Moacdieh [LM25] explored how AR-HUD image opacity affects driver visual attention and situation awareness through an eye-tracking study with 27 participants viewing driving videos with superimposed AR navigation information at varying opacity levels. Their results showed that opacity levels of 20% and 60% both effectively supported situation awareness, yet different types of AR-HUD images were affected differently by opacity changes. The authors concluded that opacity influences how drivers observe and perceive roadside environmental information, though they acknowledged that their study used static video scenes rather than interactive driving simulation, leaving the question of dynamic driver response behavior under real-time interaction unaddressed.

Doering and colleagues [D$^+$25] investigated the phenomenon of inattentional blindness with AR-HUDs through an on-road study designed to understand how the visual demands of AR tasks affect drivers' ability to detect environmental stimuli. Their findings revealed that as the visual demands of AR tasks increased, drivers' ability to promptly detect environmental stimuli decreased, with this effect being more pronounced in the driver's central visual field. The authors suggested that AR-HUD-induced inattentional blindness may be more severe than previously observed in laboratory settings, yet they did not propose specific design strategies to mitigate this attentional tunnel effect, leaving this as a direction for future research.

Shi, Ben, and Kan [SBK26] systematically examined the effects of AR-HUD warning information encoding forms (shape, color, opacity) on driver reaction time and subjective evaluation, testing 108 icon variants in a Unity simulator across different driving scenarios. Their results demonstrated that color affected recognition speed, with blue and red icons yielding higher contrast and shorter reaction times than yellow icons, while wireframe icons outperformed solid icons due to clearer contours and reduced visual clutter. The authors identified optimal parameters (blue wireframe icons at 100% opacity yielding an average reaction time of 1333.67 milliseconds), yet they noted that their study focused on single warning information and did not address priority competition when multiple warnings appear simultaneously.

Chai, Dong, and Xue [CDX25] designed a dual-optical path AR-HUD system based on holographic imaging principles to address the visual accommodation-convergence conflict (VAC) that occurs when virtual image projection depth fails to match natural scene depth. Their system achieved continuous adjustment of virtual image projection distance from 3 to 15 meters, using freeform optical correction for windshield-induced aberrations. The authors successfully validated the optical design quality, yet they acknowledged that driver behavioral validation under real driving conditions remains future work, as the study focused primarily on optical system implementation rather than on driver visual comfort and response time in dynamic environments.

Smith et al. [SGBD17] compared Head-Up Display (HUD) and Head-Down Display (HDD) use in a driving simulator experiment with 16 participants performing visual search tasks while following a lead vehicle in a motorway environment. Their results showed that HUD use was associated with longer sustained glances at the display compared to HDD, yet no differences in longitudinal and lateral vehicle control were observed – a counterintuitive finding suggesting that AR HUDs afford longer glances without negatively affecting driving performance. The authors noted that as task complexity increased, HUD use corresponded with faster performance compared to HDD, yet they acknowledged that their study was conducted in a relatively simple highway environment rather than complex urban traffic scenarios, leaving open the question of whether these findings generalize to the high-risk, sudden hazard situations central to the present work.

Kim et al. [KAM+16] developed an augmented reality pedestrian warning system using an in-vehicle volumetric head-up display capable of projecting warning cues directly onto the driver's view of the road scene. Their user study demonstrated that this AR-based cueing improved driver awareness of potential pedestrian collisions compared to a baseline without AR warnings, particularly in terms of earlier braking responses. However, the system relied on the driver's direct line-of-sight to the pedestrian and did not address the more challenging case of occluded hazards, leaving the effectiveness of AR warnings for hidden pedestrians—such as those blocked by parked vehicles or buses—as an open question that the present work's X-Ray vision and radar-based detection directly target.

Shan, Nebot, and their research team [SNW+21] developed a cooperative perception-based "X-ray vision" system that enables vehicles to detect pedestrians occluded by buildings and cyclists occluded by large vehicles through V2X communication, using roadside ITS information sharing devices to achieve multi-perspective perception fusion. Their experiments demonstrated that connected vehicles could detect occluded vulnerable road users several seconds before the driver or onboard sensors could see them, providing critical additional reaction time. The research validated the technical feasibility of X-ray visualization for blind-spot warnings, yet the authors left open the question of which visual rendering style (color, opacity, outline thickness, animation effects) is most effective for driver cognition, as the study focused on the perception layer rather than the visualization interface design layer.

He et al. [HBB26] conducted a systematic review of 74 empirical studies published between 2015 and 2024 that examined the use of immersive technologies (VR and AR) for road safety and driver training. Their findings indicated that AR head-up displays can reduce driver reaction time by up to 20% and VR-based training can decrease risky driving behaviors among novice drivers by 15%, providing quantitative evidence for the effectiveness of immersive simulation. The authors concluded that while these technologies show significant promise for enhancing hazard perception and situation awareness, they noted that longitudinal evidence on the retention of training effects and the transfer of simulator-acquired skills to real-world driving remains limited, a gap that the present three-phase experimental design (no-AR, with-AR, and post-training no-AR) is specifically positioned to address.

Cheng et al. [CTTT19] conducted in-depth reconstructions of 22 video-recorded "looked-but-failed-to-see" (LBFTS) junction accidents in Hong Kong using CCTV and dashboard camera footage to determine causation and contributory factors. The analysis revealed that attention errors and visual search strategies accounted for 41% of crashes, while eight crashes (36%) were directly attributed to physical obstruction caused by vehicle A-pillars. The study left open the question of how to effectively mitigate A-pillar-induced blind spots through technological interventions, which directly motivates the need for AR-based X-Ray visualization systems such as the one proposed in this thesis.

Milgram and Kishino [MK94] established the foundational reality-virtuality (RV) continuum and presented a widely adopted taxonomy for classifying mixed reality (MR) visual displays. Their work formally defined augmented reality (AR) as any case in which an otherwise real environment is augmented by means of virtual (computer graphic) objects. While this taxonomy provides the essential conceptual framework for situating AR technologies like head-up displays within the broader spectrum of virtual environments, the authors did not provide empirical guidance on how specific AR interface designs—such as the X-Ray vision employed in this study—impact human performance in safety-critical tasks like driving. The present work applies this theoretical framework to the practical challenge of mitigating occluded hazards in urban traffic.

Vogel [Vog03] performed a comparative study between "Time-Headway" and "Time-to-Collision" as safety indicators within different traffic flow conditions. The results demonstrated that TTC is a more reliable predictor of actual collision risk, whereas headway is better suited for measuring general traffic density. However, the researcher concluded that neither metric alone could fully account for the driver's evasive maneuver potential, suggesting that multiple surrogate safety measures should be combined for robust hazard detection.

Wang [Wan24] developed a novel real-time risk evaluation model specifically for vehicle-pedestrian interactions at intersections, addressing the limitations of traditional time-to-collision (TTC) metrics in capturing pedestrian motion uncertainty. The proposed model demonstrated superior accuracy in predicting near-miss events compared to standard TTC, as validated using naturalistic driving data. However, the author noted that the model's computational complexity currently precludes deployment on low-power embedded platforms, leaving an open challenge for integrating such advanced risk metrics into real-time automotive safety systems.

Wessels et al. [WHHO23] conducted a psychophysical experiment to examine whether trial-by-trial feedback could improve human observers' ability to account for acceleration in visual time-to-collision (TTC) estimation. Their results showed that participants systematically overestimated TTC for accelerating vehicles, and that providing feedback after each trial did not reduce this bias. The study left open the question of what type of training or visual augmentation (e.g., AR cues) might effectively calibrate a driver's TTC perception for non-constant velocity scenarios, which is directly relevant to the design of AR-HUD warning timings in the proposed simulator.

Lefèvre et al. [LVL14] provided a survey on motion prediction and risk assessment for intelligent vehicles, focusing on intent recognition. Their research categorized various predictive algorithms, such as Kalman filters and Hidden Markov Models, that evaluate the probability of future collisions. However, the survey noted that most existing models struggle with the high stochasticity of "vulnerable road users" (VRUs) in dense, unregulated urban spaces.

Rasouli and Kotseruba [RK22] proposed the Intend-Wait-Cross model, a data-driven framework for simulating realistic pedestrian crossing behavior based on intent inference and temporal dynamics. Their results showed that this model outperforms traditional constant-velocity approaches in predicting pedestrian trajectories at unsignalized intersections, particularly for crossing and gap-acceptance maneuvers. A limitation remains in the model's reliance on predefined intention categories, which may not capture all spontaneous or culturally specific pedestrian actions, suggesting the need for adaptive behavior models in driving simulators.

Harding et al. [HPY+14] published a comprehensive NHTSA research report assessing the technical, legal, and policy readiness of Vehicle-to-Vehicle (V2V) communication for widespread deployment. The report concluded that V2V technology had reached a level of maturity sufficient to enable the transmission of basic safety information, such as vehicle position and speed, to warn drivers of impending crashes. Nevertheless, as a foundational policy and technology assessment, it did not explore the integration of V2X data with augmented reality visualizations to specifically mitigate crashes caused by occluded hazards, an extension that the present work addresses through its AR-based X-Ray warning system.

Palffy et al. [PKG23] developed an occlusion-aware sensor fusion method that combines automotive radar and a stereo camera to detect pedestrians darting out from behind parked vehicles. Their experimental evaluation demonstrated that the fused system could detect such hazardous pedestrians on average 0.26 seconds earlier than a camera-only baseline, providing crucial extra reaction time for drivers or automatic braking. However, the authors noted that their method was evaluated only in controlled test scenarios with a single occluding vehicle, leaving the challenge of generalizing to complex urban occlusions (e.g., multiple obstacles, moving occluders) unresolved.

Thakur and Mishra [TM24] presented a comprehensive review of deep learning-enabled adaptive approaches for obstacle detection using sensor-fused data (camera, LiDAR, radar) in autonomous vehicles. Their analysis concluded that multi-sensor fusion improves detection robustness under adverse conditions such as low illumination or adverse weather, but also highlighted persistent challenges in sensor synchronization, calibration drift, and computational efficiency. The review explicitly identified the lack of standardized benchmarks for dynamic, high-density urban traffic as a major open research gap, which motivates the need for simulation environments like the one proposed in this thesis.

Singh et al. [SSB+24] proposed an aggregated radio-frequency and visible light communication (RF-VLC) architecture for 6G-based V2X systems, aiming to achieve ultra-reliable and low-latency communication for autonomous driving. Their simulation results indicated that the hybrid system can maintain reliability above 99.999% with end-to-end latency below 1 millisecond within a 200-meter range, outperforming conventional V2X schemes. Nevertheless, the study was conducted under idealized line-of-sight assumptions and did not evaluate performance in urban canyons with severe signal blockage or dense interference, leaving open the question of real-world deployment robustness.

Maruta et al. [MTF+21] implemented a blind-spot visualization system that combines millimeter-wave Vehicle-to-Everything (V2X) communication with augmented reality glasses to reveal occluded traffic participants. Their field experiments demonstrated that the system could successfully overlay the positions of hidden pedestrians and cyclists onto the driver's field of view, thereby reducing reaction times in blind-spot scenarios. However, the study did not address how to manage multiple simultaneously occluded hazards or how to prioritize visual cues when the driver's cognitive load is already high, leaving an open question for AR-HUD information design in dense urban traffic.

Ma et al. [MLZ24] designed and evaluated an ecological interface for AR-HUD warning systems grounded in cognitive load theory and ecological interface design principles. Their user study revealed that the proposed interface reduced drivers' reaction times and mental workload compared to conventional HUD designs, especially in complex hazard scenarios. Nevertheless, the investigation was conducted in a fixed-base simulator with a limited set of predefined hazards, leaving open the question of how the ecological interface would perform under highly dynamic, real-world driving conditions with varying levels of traffic density.

Vollrath and Morawietz [VM25] conducted a driving-simulator study with 47 participants to examine the combined effect of misses (failures to warn during critical events) and false alarms (unnecessary warnings) in a pedestrian collision warning system. Their results showed that a system with 33% false alarms but perfect detection did not induce a “cry-wolf” effect but caused non-critical braking, while a system with 33% misses led to slower reactions in critical situations. However, the condition that combined both error types (70% detection with 33% false alarms) produced the fastest responses but was rated by drivers as the most annoying and least useful, highlighting a critical trade-off between objective safety gains and user acceptance that remains unresolved.

Fisher et al. [FCHT17] compiled a comprehensive handbook summarizing research on teen and novice drivers, focusing on the distinction between vehicle handling skills and higher-order cognitive skills such as hazard anticipation and attention maintenance. The handbook presented evidence that PC-based training programs like RAPT (Risk Awareness and Perception Training) can improve novices’ scanning behaviors at intersections. Nevertheless, the editors acknowledged that long-term retention of these trained skills and their transfer to fully immersive VR or real-world environments remain open research questions.

Scialfa et al. [SDF$^+$12] investigated the differences in hazard perception between novice and experienced drivers using a series of controlled traffic scenarios. Their results confirmed that experienced drivers are faster at detecting and responding to latent hazards due to more sophisticated mental models of risk. However, the research focused primarily on detection speed, leaving open the question of how to train the specific evasive maneuvers required after a hazard is identified.

Endsley [End95] proposed a three-level theoretical model of Situational Awareness (SA), encompassing the perception, comprehension, and projection of environmental elements. The research established that effective decision-making in dynamic systems depends on the integration of these hierarchical stages to anticipate future states. While the model is foundational, its application to the specific visual distractions caused by AR overlays in automotive contexts remains an area of ongoing debate.

Horswill and McKenna [HM04] integrated the concept of hazard perception into Endsley’s broader framework of situation awareness, arguing that a driver’s ability to anticipate potential threats is a specific application of Level 3 projection. Their review concluded that hazard perception is a distinct, measurable skill that reliably differentiates safe from unsafe drivers. However, the authors left open the challenge of developing a standardized, validated hazard perception test that can be deployed across different cultural and traffic environments.

Wickens [Wic08] proposed the Multiple Resource Theory, which posits that human attention consists of separable pools of resources based on processing stages, modalities, and visual channels. The theory successfully predicts that time-sharing between tasks is more efficient when they use different resources (e.g., auditory and visual) rather than the same resource. Wickens identified that the theory’s predictions become less precise when tasks compete within the same visual channel, such as monitoring an AR HUD overlay while scanning the road ahead, a scenario highly relevant to augmented reality driving interfaces.

Reason et al. [RMS+90] distinguished between driver errors (unintended failures of information processing or action execution) and violations (deliberate deviations from safe driving practices). Their analysis of accident reports revealed that errors and violations have different psychological origins and therefore require different countermeasures. The study concluded that training programs aimed at improving hazard perception may reduce errors but are unlikely to affect willful violations, suggesting that AR warning systems might need to address both categories through separate mechanisms.

Pradhan et al. [PHD+05] used eye-tracking technology to evaluate the effectiveness of a PC-based risk awareness training program for novice drivers. Their results showed that trained drivers exhibited better scanning patterns at intersections compared to an untrained control group. Nevertheless, the study did not measure whether these improved scanning behaviors would persist under the high cognitive load of actual real-world driving.

Crundall [Cru16] developed a hazard prediction test based on the Situation Awareness Global Assessment Technique, requiring drivers to anticipate what would happen next in filmed traffic scenes. The results demonstrated that experienced drivers outperformed novices in predicting latent hazards, suggesting that prediction is a critical component of hazard perception ability. However, the study did not investigate whether superior prediction scores directly translate into faster evasive maneuvers or lower collision rates in real driving scenarios.

Hart and Staveland [HS88] developed the NASA Task Load Index (NASA-TLX), a multidimensional scale for assessing subjective mental workload. Their research established a standardized method for measuring the cognitive demands of a task across various dimensions, such as frustration and temporal demand. The study is widely cited, yet the researchers noted that subjective ratings must always be validated against objective physiological or performance data to ensure a complete picture of workload.

Brooke [Bro96] introduced the System Usability Scale (SUS), a simple ten-item Likert-scale questionnaire for quickly measuring the perceived usability of a wide range of systems. The research demonstrated that SUS yields a reliable and robust single score representing overall usability, despite its brevity. However, Brooke noted that SUS alone cannot identify specific interface flaws or explain why users find a system difficult to use, requiring complementary qualitative methods for diagnostic insights.

Kennedy et al. [KLBL93] developed the Simulator Sickness Questionnaire (SSQ), an enhanced method for quantifying the severity of nausea, oculomotor discomfort, and disorientation symptoms induced by simulator exposure. Their validation study established SSQ as a reliable and sensitive instrument for comparing different simulation conditions. The authors cautioned that SSQ scores do not directly predict performance degradation, and the relationship between specific symptom clusters and driving task errors remains incompletely understood.

Cohen [Coh88] provided a definitive framework for statistical power analysis in the behavioral sciences, emphasizing the importance of effect sizes. The work established that p-values alone are insufficient for determining the practical significance of research findings in human-centered studies. Despite its rigor, the framework noted that choosing the appropriate effect size metric requires a deep understanding of the specific experimental design and population variance.

3. Foundations

This chapter establishes the interdisciplinary theoretical and technical framework essential for investigating the efficacy of Augmented Reality (AR) in mitigating driving risks. To rigorously address the core research questions—specifically, the impact of AR on accident reduction (RQ1) and the cultivation of driver risk awareness (RQ2)—it is imperative to bridge the gap between virtual environment engineering and cognitive human factors. This chapter provides a comprehensive synthesis of the mature technologies and established psychological models that serve as the bedrock for the driving simulator developed in this research.

The structural progression of this chapter follows a bottom-up approach, beginning with the fundamental physics of virtual simulation and vehicle kinematics. Subsequently, the discourse transitions to the optical principles of Head-Up Displays (HUD), which govern the perceptual quality of the AR interface. To ensure the simulator functions as a valid research tool, the mathematical foundations of hazard detection, such as Time-to-Collision (TTC) and Post-Encroachment Time (PET), are detailed alongside the principles of Cooperative Intelligent Transport Systems (C-ITS). Finally, the chapter concludes by outlining the psychometric and statistical methodologies required for the empirical evaluation of driver behavior. By synthesizing these diverse fields, this chapter provides the necessary context to understand the implementation and evaluation phases described in subsequent chapters.

3.1. Procedural Urban Environments

The ecological validity of a driving simulator hinges on the systematic construction of a virtual environment that balances visual complexity with deterministic control. In this research, the simulated world is modeled after the *Manhattan Grid topology*—a rigid, orthogonal street network characterized by its perpendicular intersections and uniform block dimensions. As illustrated in Figure 3.1, this real-world urban structure provides a unique morphological blueprint where building density and 90-degree road alignments create consistent "blind-spot" conflict zones. This real-world structure serves as the morphological blueprint for the simulator's procedural environment, as theorized by Marshall [Mar04]. Anchoring the environment in this specific topology ensures that the geometric parameters of traffic hazards remain standardized across the generated world, a design principle that finds its theoretical foundation in Marshall's work on urban street networks [Mar04]. This standardization allows the research to isolate the AR interface as the primary variable in hazard detection.

To translate this historical urban logic into a functional simulation, the system utilizes

Figure 3.1.: Aerial view of the Manhattan grid system.
Note: Reproduced from BestNeighborhood.

Procedural Content Generation (PCG) to synthesize an expansive, modular street network following the principles established by Parish and Müller [PM01]. Each urban "tile" functions as a self-contained module comprising localized road metadata and pre-baked physics colliders. To maintain the high frame-rate consistency required for Virtual Reality, the system implements a dynamic "Look-Ahead" culling algorithm inspired by on-demand procedural city generation techniques [BBS+10]. This ensures that only tiles within a defined radius of the participant's spatial coordinates are active within the scene graph. By managing resources through an asynchronous pipeline, the simulation maintains a strictly bounded draw-call budget, effectively preventing the temporal jitter that could otherwise invalidate perception-reaction measurements during critical traffic events.

To further reconcile high-density urban detail with the stringent latency requirements of Virtual Reality, the environment employs a *yield-based staggered generation* strategy. A primary technical challenge in high-fidelity simulation is the "instantiation spike" — a phenomenon where the synchronous loading of complex meshes and textures exceeds the frame budget (e.g., 11.1 ms for a 60 Hz refresh rate), inducing perceptible stutter. To mitigate this, the system leverages a *coroutine-driven yield mechanism* to decouple environment construction into an asynchronous pipeline. Under this framework, the simulation prioritizes the instantiation of the road skeleton and navigational waypoints to maintain functional continuity, while the remaining architectural load is distributed across successive frames. By "yielding" execution back to the main rendering loop after each incremental instantiation cycle, the system effectively flattens the CPU usage curve and eliminates micro-stutters. This technical robustness is vital for research integrity, ensuring that any recorded fluctuations in participant reaction times or physiological stress are authentic responses to traffic hazards rather than artifacts of system latency.

To support consistent driver behavior across experimental trials, a dedicated driving navigation system must be provided to participants. Unlike the waypoint-based navigation graph that governs NPC movements, this system delivers explicit route guidance—such as turn-by-turn instructions, lane recommendations, and destination cues—directly to the driver. By ensuring that all participants follow the same prescribed path under identical directional prompts, the navigation system eliminates confounding variables related to route memorization or spatial disorientation, thereby allowing the research to isolate the effect of Augmented Reality interfaces on hazard detection and driving performance.

Building upon the static urban grid, the dynamic traffic ecosystem is governed by a navigation graph system based on waypoint topology. The virtual road network is abstracted into a directed graph, as shown in Equation 3.1:

$$G = (V, E). \tag{3.1}$$

where the set of nodes V represents discrete waypoints situated at the center of lanes and intersection entrances. Each node $v_i \in V$ serves as a semantic container, storing critical attributes such as target speed limits, lane widths, and traffic signal affiliations. The set of edges E defines the directional constraints, where an edge e_{ij} signifies a permissible transition between waypoints. This graph-based architecture is a fundamental prerequisite for scientific replicability. By defining fixed trajectories, the simulator ensures that critical hazards — such as the "Hidden Pedestrian" or "Sudden Lane Change" events — occur at identical spatial-temporal coordinates for every participant. This determinism effectively isolates the AR interface as the primary independent variable, eliminating potential noise from stochastic NPC behavior.

The operational logic of NPC entities is implemented through a robust waypoint-following architecture that prioritizes computational efficiency and behavioral consistency. Rather than employing resource-intensive real-time pathfinding, each vehicle and pedestrian is assigned a trajectory defined by a sequence of spatial nodes within the navigation graph, incorporating the multi-agent path planning principles investigated by Best et al. [BNCM16]. For vehicular NPCs, these waypoints dictate longitudinal velocity and lateral lane-keeping constraints, for pedestrians, they define walking paths and waiting zones at intersections, following the destination-driven dynamics established in social force models by Helbing et al. [HFV00]. This node-based movement provides the technical foundation for the simulator's adaptability, allowing the system to simulate diverse driving cultures by simply modulating transition parameters such as waiting intervals or braking thresholds. This mechanism ensures that NPC behaviors remain predictable for experimental control while maintaining sufficient flexibility to represent the varied traffic dynamics required for the study.

The validity of a driving simulator relies on its ability to replicate the fundamental physical laws that govern vehicle movement, as the reliability of perception-reaction data is linked to the fidelity of vehicular dynamics. To provide a realistic driving experience, the simulation must incorporate a high-fidelity kinematic model that accounts for the interplay between *propulsion, deceleration, and directional stability* as established in standard vehicle dynamics literature. The core of this kinetic feedback is defined by the *motor torque and power-to-weight ratio*, which dictate the vehicle's acceleration profile and provide the driver with a sense of the vehicle's mass and inertia. Equally critical is the modeling of *braking dynamics and friction limits*, which ensure that stopping distances are grounded in Newtonian physics rather than linear approximations.

Furthermore, the simulation must account for *dynamic weight transfer* — the shift in the center of mass that occurs during rapid velocity changes. For instance, during an emergency braking event (a key focus of RQ1), the forward shift of load alters the tire-road interaction. By incorporating parameters such as *downforce and anti-roll stability*, the environment ensures that the vehicle handles intuitively under stress. Maintaining this level of physical realism is essential for experimental integrity. It ensures that the participant's reaction times and braking behaviors are authentic responses to traffic hazards, rather than artifacts of an unnatural or overly simplified driving model.

3.2. AR-HUD Optical Principles

The efficacy of Augmented Reality (AR) in a driving context is primarily determined by the optical delivery of information and its spatial integration with the real-world environment. In professional automotive engineering, an *AR Head-Up Display (AR-HUD)* functions by projecting digital information through a complex optical path using the windshield as a final combiner. The Picture Generation Unit (PGU) emits light that is reflected and magnified by aspheric mirrors to create a *Virtual Image Distance (VID)*, typically projecting the image 7.5 to 10 meters in front of the vehicle. This depth-matching is necessary for minimizing the *Vergence-Accommodation Conflict (VAC)*. Recent dual-optical path AR-HUD designs have demonstrated continuous virtual image adjustment from 3 to 15 meters, effectively eliminating the focal accommodation lag that leads to inaccurate hazard distance judgment [CDX25].

A significant advantage of AR-HUD over traditional *Instrument Clusters (IC)* or *Center Stack Displays (CSD)* lies in the reduction of "eyes-off-the-road" time. Research in human factors indicates that shifting gaze to a dashboard and back to the road requires approximately 0.5 to 1.5 seconds for focal adjustment and cognitive re-orientation. This "attentional switching" creates a dangerous blind interval during which critical traffic changes, such as a pedestrian stepping into the street, may go unnoticed, a risk that windshield-based interfaces effectively mitigate by enhancing situational awareness within the driver's foveal vision. As discussed in Chapter 2 [GFK14], controlled simulator environments are essential for evaluating these attentional benefits while maintaining experimental replicability. To further optimize this, AR-HUD keeps information within the driver's *Primary Field of View (FOV)*, leveraging the human brain's ability to process superimposed visual streams. However, the existing literature warns of "Cognitive Tunneling" and the potential for "Visual Capture," where drivers may over-fixate on the digital display at the expense of the dynamic road environment. This research, therefore, adopts the principle of *Minimalist Information Hierarchy*, categorizing content into prioritized layers: static vehicle data (e.g., speed), dynamic navigational cues (e.g., conformal arrows), and critical safety alerts (e.g., X-Ray hazard silhouettes). This design strategy is further supported by Lopez and Moacdieh [LM25], who demonstrated that opacity levels of 20% and 60% both effectively support situation awareness while preventing visual clutter. Complementary findings by Shi, Ben, and Kan [SBK26] identified that blue wireframe icons at 100% opacity yield the shortest reaction times, providing specific design parameters for the present system's alert hierarchy. Furthermore, Smith et al. [SGBD17] demonstrated that HUD use affords longer sustained glances without degrading vehicle control, confirming the attentional benefits of windshield-based displays.

The integration of "X-Ray Vision" and predictive alerts represents the transition from simple information display to active hazard intervention. This mechanism is mathematically grounded in the *Time-to-Collision (TTC)* model and the *Perception-Reaction Time (PRT)* chain. In a standard driving environment, the detection of occluded hazards, such as a pedestrian emerging from behind a bus, is physically impossible until the object enters the driver's direct line-of-sight. As demonstrated by Kim et al. [KAM+16], volumetric AR pedestrian warning systems can reduce the time drivers need to locate hazards in blind-spot scenarios. Theoretical frameworks for *Visual Scaffolding* suggest that by visualizing these occluded entities through transparency metaphors—a "see-through" technique—the system effectively bypasses the initial "detection" phase of the PRT. This allows the driver's cognitive process to jump directly to the "identification" and "decision-making" stages, potentially saving critical milliseconds that determine the outcome of a collision. The technical feasibility of detecting such occluded hazards has been further validated by Shan et al. [SNW+21], who demonstrated that V2X-based cooperative perception enables connected vehicles to detect pedestrians around corners several seconds before they become visible to onboard sensors. Furthermore, these AR cues may function as a temporary cognitive support that facilitates the internal recognition of risk patterns even after the visual aids are removed.

To replicate these complex optics within a virtual environment, the system utilizes a *Virtual Canvas* fixed within the driver's viewing frustum, effectively simulating the windshield's role as a digital combiner. The core technical implementation relies on *3D-to-2D Spatial Registration*, a process that ensures virtual overlays remain spatially coherent with real-world anchors regardless of the vehicle's movement. This mathematical precision is essential to prevent visual fatigue and spatial disorientation, as even minor discrepancies between virtual overlays and physical road objects can lead to discomfort and reduced trust in the system. This involves calculating the screen-space coordinates (u, v) from a 3D world-space point P_w by applying a perspective projection pipeline. The mathematical foundation of this transformation is expressed through the matrix equation in Equation 3.2:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = M_{int} \cdot M_{ext} \cdot \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}. \tag{3.2}$$

where M_{ext} represents the vehicle's pose or the extrinsic matrix, defining the transformation from the world coordinate system to the camera's coordinate system, and M_{int} represents the intrinsic matrix, which accounts for the focal length, principal point, and other camera-specific parameters [KAM+16]. This projection ensures a "conformal" effect, where navigation arrows appear to lie flat on the asphalt and X-Ray silhouettes remain locked onto moving entities in the 3D space. While this research does not implement raw sensor fusion algorithms from hardware such as LiDAR or Radar, it builds upon the theoretical assumption of a high-fidelity *Vehicle-to-Everything (V2X)* data stream. This conceptual data bus provides the simulator with the precise object coordinates (X_w, Y_w, Z_w) required to trigger the AR overlays, effectively mimicking the interconnected sensing capabilities of future smart city infrastructures. This reliance on shared perception acknowledges the resolution limits of standalone vehicle-mounted sensors in dense urban environments.

As illustrated in Figure 3.2, the primary advantage of the AR-HUD is keeping the driver's gaze focused on the road. Traditional designs require drivers to look down at the instrument cluster, a process that takes 0.5 to 1.5 seconds for refocusing and cognitive re-orientation. By projecting critical safety information directly onto the windshield (HUD area), the system minimizes the time the driver's gaze is diverted from the road surface, ensuring a high degree of cognitive integration between the digital information layer and the physical road topology. Grounded in the cognitive load principles reviewed in Chapter 2 [SAK11], the system adopts a minimalist design approach that projects only high-priority speed data and "X-Ray" alerts directly onto the windshield, thereby removing this dangerous "eyes-off-the-road" interval. This streamlined information delivery allows the driver to perceive and react to hazards much faster and more safely than using conventional dashboards, thereby maximizing the cognitive budget available for situational monitoring. As noted in the related work [GFK14], such minimalist AR interfaces have been shown to reduce reaction times while preventing the visual clutter that can otherwise degrade situation awareness.

Figure 3.2.: Example of AR-HUD interaction from the driver's perspective.

Furthermore, the design of these AR alerts must account for *Trust in Automation* and *Alarm Fatigue* theories. If an AR system provides too many "false positives" or overly aggressive warnings, drivers may develop a skepticism that leads to delayed responses during actual emergencies. Conversely, an overly reliable system might induce *automation bias* [PM10]—as elaborated in the previous chapter—where the driver stops scanning the environment manually, relying entirely on the AR "X-Ray" to find dangers. This over-reliance can lead to "attentional tunneling," where drivers become so focused on high-salience AR alerts that they fail to detect unmarked peripheral hazards. This risk is particularly pronounced with AR-HUDs: as discussed in Chapter 2 [D+25], on-road studies have shown that as AR task visual demand increases, drivers' ability to detect environmental stimuli decreases, with this effect being more pronounced in the central field of view. The foundation for this research's experimental design thus includes the study of *salience*—how the color, brightness, and flashing frequency of an AR element can be tuned to capture attention without causing panic or annoyance. Achieving this balance is critical to prevent information overload, a design challenge that requires tailoring AR interventions to the driver's cognitive state. This is essential for the "Awareness Development" goal of the simulator, as the objective is to train the driver's internal risk-assessment model through *perceptual learning*, rather than providing a permanent crutch that degrades independent driving skills.

In the context of the *Visual Scaffolding* theory, the AR-HUD acts as a temporary cognitive support that helps novice drivers recognize patterns of danger that they would otherwise lack the experience to identify. This simulation-based exposure addresses the critical training gap discussed in Chapter 2 [FCHT17], where traditional instruction often fails to provide sufficient experience with "near-miss" scenarios. As the driver undergoes repeated exposure to "near-miss" scenarios within the simulator, the AR cues facilitate the transition from *controlled processing* to *automatic processing* of hazard information. This theoretical transition is what enables the "Awareness Cultivation" mentioned in RQ2. By providing a visual bridge between a hidden risk and its potential consequence, the AR system helps the driver build a more robust *mental model* of the traffic environment. This model persists even in the absence of the AR cues, fulfilling the pedagogical objective of using simulation as a long-term training tool rather than a short-term driving aid.

Finally, the stability of the AR-HUD interface is critical for user trust and physiological comfort. Unlike static UI elements, conformal AR graphics are highly susceptible to "swimming" or jitter caused by high-frequency updates in vehicle pose data. Although this research does not involve the development of new anti-jitter hardware, it builds upon established *State Estimation* principles, such as Low-Pass Filtering and Kalman-based prediction, to stabilize the Canvas rendering pipeline. Maintaining such interface stability is a prerequisite for user acceptance, particularly when tailoring AR interventions to the driver's cognitive state and experience level. These mathematical techniques ensure that the virtual information—such as the X-Ray silhouettes of occluded pedestrians—remains sharp and legible even under high-speed driving conditions or mechanical vibrations. By synthesizing these optical, cognitive, and mathematical foundations, the simulator provides a transparent and immersive environment for the empirical evaluation of AR's impact on driver safety and the potential for technological interventions to reduce traffic fatalities.

3.3. Multi-Source Hazard Detection

The primary objective of this section is to establish a rigorous logical and mathematical bridge between raw physical environmental sensing and the subsequent Augmented Reality (AR) intervention mechanisms. While the preceding sections discussed the optical delivery of information, this discussion focuses on the systemic intelligence required to transcend the perceptual limitations of an individual human driver. In urban environments, hazards are often non-linear and obscured by structural complexity. Therefore, a singular sensing modality is insufficient. By synthesizing multi-source data fusion—merging localized reflective sensing with distributed collaborative data—the simulator constructs an advanced "Safety Buffer" that serves as the algorithmic foundation for the Accident Detection System described in Chapter 4. This modular framework aligns with the decoupled architecture principles demonstrated in high-fidelity simulation platforms such as CARLA, allowing for a transition from a reactive driving model to a proactive, technologically-mediated awareness model, addressing the core challenges of urban driving safety identified in earlier research phases [TM24].

The fundamental acquisition of spatial data within the simulated environment is rooted in the physics of Time-of-Flight (ToF). To accurately simulate the distance between the ego-vehicle and surrounding Non-Player Characters (NPCs), the system emulates the emission of electromagnetic or ultrasonic pulses. By calculating the precise time interval Δt between the pulse emission and the reception of its echo from a target's surface, the absolute distance d is derived using the constant speed of light c, as shown in Equation 3.3:

$$d = \frac{c \cdot \Delta t}{2}. \tag{3.3}$$

This distance acquisition is executed at high frequencies to ensure that the simulator captures millisecond-level changes in the traffic environment. In the context of Unity-based development, this theoretical model is translated into Raycasting and Sphere-casting techniques that mimic the temporal precision of physical radar sensors. Furthermore, the foundation accounts for signal attenuation and the "Signal-to-Noise Ratio" (SNR) in complex urban canyons, providing the raw input necessary for dynamic risk assessment under varying environmental conditions.

Static distance alone is insufficient for hazard evaluation. Therefore, the system implements a Discrete Time-to-Collision (TTC) algorithm based on first-order kinematic derivatives. This approach draws upon the principles of the Intelligent Driver Model (IDM), which emphasizes the role of longitudinal distance and velocity differences in maintaining traffic stability. By sampling the distance d at a consistent interval Δt_{sample}, the system derives the relative velocity v_{rel} between the ego-vehicle and any prospective target. The TTC, representing the remaining time before a physical impact occurs under constant velocity assumptions, is expressed through the discrete logic in Equation 3.4:

$$TTC = \frac{d}{v_{rel}} = \frac{d}{d_{t-1} - d_t} \cdot \Delta t_{sample}. \tag{3.4}$$

This formulation allows the simulator to quantify the "hazard intensity" of any given interaction, grounded in the fundamental temporal scale for near-miss determination.

By employing a discrete sampling approach, the system remains robust against the high stochasticity of urban traffic maneuvers, such as sudden braking or lateral cutting. This mathematical methodology provides the “Ground Truth” for the hazard evaluation logic implemented in the project’s backend scripts, allowing for a deterministic trigger for AR events [Wan24, WHHO23].

Building upon the TTC calculation, the system applies Warning Threshold Theory to categorize safety states into three distinct levels. As demonstrated in comparative safety indicator studies by Vogel [Vog03], TTC serves as a more reliable predictor of actual collision risk than traditional headway metrics in active interventions. In this research, specific temporal boundaries are established: a “Cautionary Threshold” (e.g., 4.0 seconds) for passive monitoring, a “Warning Threshold” (e.g., 2.5 seconds) for initiating active AR intervention, and a “Critical Threshold” (e.g., 1.5 seconds) for maximum-severity alerts. These values are derived from human factors research regarding average Perception-Reaction Time (PRT), which accounts for visual detection, cognitive processing, and motor execution. The effectiveness of driver warning systems is directly contingent upon the contextual accuracy of the information provided. When the calculated TTC crosses these thresholds, the system progressively escalates its response from passive monitoring to active AR intervention and, if necessary, to emergency alerts. This theoretical alignment ensures that the simulator’s warnings are not arbitrary but are calibrated to the physiological limits of human cognitive processing.

The visual manifestation of detected hazards on the AR-HUD is divided into two distinct modalities, as conceptualized in the preceding foundational section. The first modality consists of Direct High-Priority Alerts, which are projected onto the windshield to address immediate, line-of-sight (LOS) threats. These are designed as “salient visual attractors” that direct the driver’s focus to imminent collisions, mitigating the risk of “visual capture” where a driver might focus on irrelevant dashboard information. The second modality is the X-Ray Holographic Projection, which utilizes see-through visualization to render occluded pedestrians or vehicles. This see-through metaphor bridges the “perception gap” by reconstructing invisible environmental data into intuitive visual outlines, effectively fusing the driver’s natural vision with the system’s omniscient sensor data. By providing these predictive cues, the system supports the driver’s visual anticipation skills, a higher-order cognitive ability essential for identifying hazards before they become physically visible. Furthermore, this approach addresses the technical challenge regarding the effective visual communication of “future risks” that exist outside the driver’s immediate field of vision [MLZ24, MTF+21].

As illustrated in Figure 3.3, the system architecture follows a hierarchical processing pipeline that translates heterogeneous data into actionable driver intelligence. Following the modular, decoupled architecture principles demonstrated in high-fidelity simulators like CARLA, the pipeline effectively separates environmental sensing from the event management and UI layers. This layered logic ensures that the AR interface provides prioritized, context-aware interventions that reduce cognitive load while maximizing situational awareness. This separation of concerns is critical for managing the high computational demands of real-time simulation.

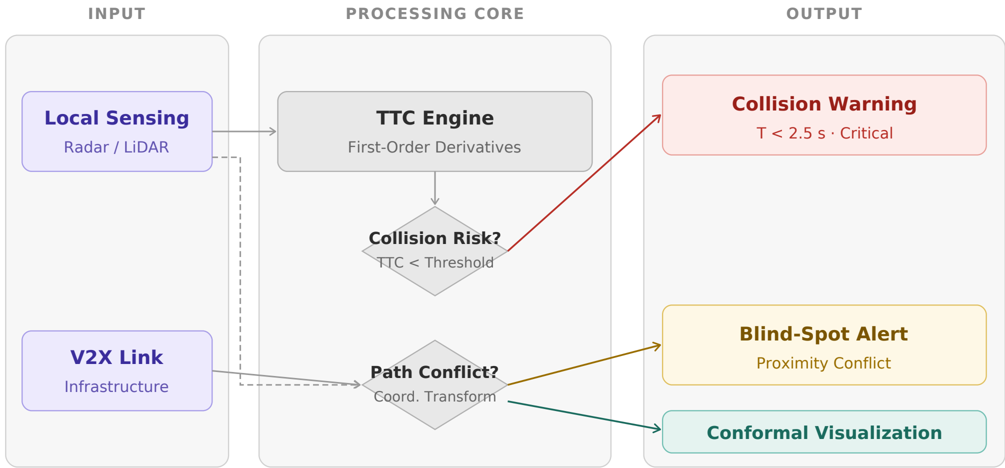


Figure 3.3.: Functional architecture of the multi-source hazard detection.

To achieve comprehensive situational awareness, the simulator integrates two complementary sensor strategies: Local Radar Sensing and Distributed V2X Perception. Local radar provides a high-fidelity, autonomous view of the vehicle's immediate surroundings, functioning effectively for objects within the physical line-of-sight. However, localized sensors are inherently limited by environmental occlusions and atmospheric conditions. To overcome these limitations, the system introduces Vehicle-to-Everything (V2X) collaborative data [HPY+14, SSB+24], which provides a globalized traffic view. By combining these heterogeneous data sources, the project constructs a 360-degree hazard detection network. Correlating data from these multiple sources not only enhances the overall reliability of the hazard model but also builds greater user trust in the automated warnings. This fusion strategy ensures that "Blind Spot Conflicts," which account for a high percentage of urban accidents, are systematically addressed through technological redundancy.

The Vehicle-Based Radar component operates as the primary layer of "Active Defense." It continuously tracks the spatial coordinates and heading vectors of all surrounding NPCs within a specified radius. By monitoring the trend of distance d over time, the system identifies anomalous behaviors, such as a lead vehicle's emergency deceleration or a lateral vehicle's high-speed lane intrusion. This localized sensing is critical for managing immediate physical proximity risks. In the system architecture, the radar data is the highest-priority input for the longitudinal collision avoidance module. This foundation assumes a sensor with a 120-degree horizontal field of view, mimicking modern mid-range automotive radar units, ensuring that the driver is alerted to threats that are physically present in the immediate trajectory. However, localized radar faces significant challenges in achieving the high angular resolution required to distinguish between closely spaced targets in complex, occluded urban corridors, a limitation this simulator seeks to mitigate via complementary sensing layers [PKG23].

The V2X (Vehicle-to-Everything) framework serves as the conceptual "Extended Perception" of the simulator. Building upon the cooperative awareness standards evaluated in technical reports for the NHTSA [HPY+14], V2X allows the ego-vehicle to access data packets from infrastructure-based sensors beyond its physical line-of-sight. This enables the vehicle to "know" the position of actors that it cannot "see," transforming a fragmented local perception into a unified, systemic awareness of the urban landscape. Such V2X-enabled systems can effectively mitigate blind-spot conflicts by extending the driver's sensing horizon through shared data [CTTT19]. The theoretical foundation assumes a "Zero-Trust" data verification model, ensuring that only coordinates verified by multiple infrastructure nodes are projected onto the driver's HUD.

In this project, the V2X logic is specifically applied to Blind Spot Reconstruction. When a pedestrian is occluded by a large object, such as a stationary bus, the infrastructure sensors broadcast the precise real-world coordinates (X, Y, Z) of the hidden actor. The vehicle's onboard unit receives these coordinates and performs a Spatiotemporal Alignment to synchronize the remote data with its own ego-centric frame of reference. This process facilitates cooperative awareness, allowing for much earlier hazard warnings by shifting the driver's response from reactive braking to proactive deceleration. This involves complex coordinate transformations between World Geodetic systems and the vehicle's local Cartesian space. This synchronization is the prerequisite for the X-Ray system, as it provides the exact spatial anchors required to render the holographic silhouette in a conformal manner on the windshield, maintaining visual stability during vehicle motion [MTF+21].

The Risk Transformation process occurs when the V2X data is fed into the accident detection algorithm. The system does not merely display all V2X data. Instead, it performs a path-conflict analysis to determine if the hidden actor's trajectory will intersect with the ego-vehicle's future path. This aligns with the predictive risk assessment frameworks categorized by Lefèvre et al. [LVL14], which evaluate the probability of future collisions based on intent recognition and trajectory projection. This involves calculating the "Conflict Point" where the predicted trajectories of both entities meet. If a conflict is predicted within a specific temporal window, the system triggers the AR projection. A 3D "ghost" silhouette is rendered onto the windshield at the exact perspective-aligned position of the pedestrian, thereby enabling the effective visual communication of "future risks" that exist outside the driver's current natural field of vision.

Figure 3.4 shows the "X-Ray" holographic silhouettes representing occluded traffic participants identified via V2X collaborative perception, aligned with the driver's perspective to mitigate blind-spot conflicts. (Visual representation generated using generative AI.)

To manage the driver's cognitive resources, the research establishes a Conflict Prioritization and Warning Hierarchy. This design specification dictates that not all warnings are equal. Immediate radar-based conflicts, representing imminent physical danger, are assigned the highest salience. These are rendered with red, flashing visual boundaries on the HUD to trigger rapid motor responses. In contrast, V2X-based latent hazards are rendered with stable, semi-transparent outlines (blue or green), signaling "Awareness" rather than "Emergency." This integration between technical detection and human-centric UI design emphasizes that tracking the heading and velocity of surrounding agents must

Figure 3.4.: First-person visualization of the AR-HUD interface.

be translated into an intuitive interface to ensure effective intervention. This hierarchy provides the foundational logic for the ‘ARHUDController‘, ensuring that the UI state machine effectively manages the driver’s attention budget and prevents the “Alarm Fatigue” often associated with modern ADAS systems [VM25].

Beyond immediate safety, this hierarchy serves the goal of Instructional Scaffolding. By visually highlighting latent risks before they become emergencies, the AR-HUD acts as a pedagogical tool that trains the driver’s internal risk-recognition model. This is particularly vital for urban intersections where traditional safety indicators often struggle to capture the erratic movements of pedestrians. Over time, the consistent pairing of V2X silhouettes with subsequent real-world sightings of pedestrians builds a conditional reflex in the driver. This aligns with the “Awareness Training” objective (RQ2), as the simulation environment is designed not just to prevent accidents in the virtual world, but to shape safer scanning behaviors. The foundation assumes that this “visual crutch” is most effective when it bridges the gap between a novice’s underdeveloped situational awareness and the expert’s intuitive risk projection.

A critical component of this foundational layer is the consideration of Sensor Error and Trust Calibration. Real-world V2X and radar systems are susceptible to latency, jitter, and occlusions. While the simulator provides a controlled environment, the foundation recognizes that user trust is a dynamic variable influenced by system reliability. If the AR-HUD provides “false positives” (alerts where no danger exists), the driver may develop skepticism and ignore future critical warnings. Conversely, “False Negatives” (missed hazards) can lead to catastrophic safety failures. The visual representation of sensor uncertainty is a complex challenge. Hence, the detection logic includes a “Confidence Score” for each hazard. This ensures that only alerts with high statistical certainty are rendered with high visual salience, thereby calibrating the driver’s trust in the automated system and mitigating the visual confusion that can arise from inaccurate overlays [VM25].

3.4. Evaluation Methodology

The final component of the theoretical foundation establishes the psychological and statistical frameworks used to evaluate both driver behavior and the efficacy of the AR intervention. These frameworks directly underpin the experimental design and measurement instruments described in Chapter 5, ensuring that the conclusions drawn are grounded in established cognitive science and inferential statistics rather than in ad hoc observation.

Central to this evaluation is the comprehensive concept of *Situational Awareness (SA)*, defined by Endsley as a three-level cognitive process comprising the perception of environmental elements (Level 1), the comprehension of their meaning (Level 2), and the projection of their future state (Level 3) [End95]. In the high-stakes context of urban driving, failures at any of these levels, whether through misperception of an occluded pedestrian or poor projection of a vehicle's intended path, account for the vast majority of traffic incidents. The proposed AR system is therefore conceptualized as a *cognitive scaffold* that targets each SA level through specific visual interventions. The X-Ray silhouette addresses Level 1 by compensating for physical occlusion, making latent hazards perceptible before they enter the driver's natural line of sight. The radar-based safety bubble and color-coded HUD alerts address Level 2 by allowing the driver to assess proximity threats without manual distance estimation. Predictive trajectory overlays and anticipatory warnings address Level 3 by enabling the driver to anticipate the future spatial occupancy of surrounding traffic participants.

Horswill and McKenna further integrated the concept of hazard perception into Endsley's SA framework, arguing that a driver's ability to anticipate potential threats is a specific application of Level 3 projection [HM04]. Their review concluded that hazard perception is a distinct, measurable skill that reliably differentiates safe from unsafe drivers. This insight directly motivates the present study's comparison between the immersive AR driving group and the passive video viewing group: if immersive simulation with AR cues enhances hazard projection ability, then VR participants should demonstrate fewer hazard collisions when the AR system is active, and a subset of this improvement should persist after the system is removed. Because video participants are not engaged in active hazard response, their questionnaire data serve as a passive-observation baseline, permitting a comparison of post-hoc subjective perceptions of immersion, utility, and educational value between active training and passive exposure.

To complement objective driving metrics, subjective user experience is assessed through a structured questionnaire employing a four-point Likert scale (1 = Strongly Disagree, 4 = Strongly Agree). A forced four-point scale is adopted to eliminate the neutral midpoint and encourage directional responses, consistent with best practice in training evaluation research. The instrument is organized into four thematic dimensions mapped onto the study's evaluation goals. The full item set and dimensional structure are detailed in Chapter 5 and Appendix A. Both the VR group and the Video group complete the identical questionnaire immediately after their respective experimental sessions, allowing direct between-group comparisons of subjective experience.

The concept of *Trust Calibration* is central to the human-factors evaluation of automated assistance systems. According to the trust and automation literature, drivers must develop a level of reliance that precisely matches the actual reliability of the AR system [PM10]. Over-trust, or complacency, may cause a driver to cease manual environmental scanning, leading to automation bias where the driver fails to detect peripheral threats not explicitly highlighted by the system. Under-trust, or skepticism, may cause the driver to ignore safety-critical alerts, negating the system's benefit entirely. The present study's three-phase design (S1: no AR baseline, S2: AR active, S3: AR removed) explicitly tests whether short-term exposure to the AR system leads to residual effects that persist after the system is withdrawn. System Trust is assessed as a dedicated quantitative dimension of the post-session questionnaire, comprising three Likert items that capture participants' perceived reliability of and confidence in the AR-HUD system. Together with Immersion, AR Utility, and Risk Awareness, it forms the four thematic dimensions of the instrument detailed in Chapter 5 and Appendix A.

For the quantitative assessment of driving performance, the primary objective metric is *hazard collision occurrence*, recorded as a binary outcome (0 = avoided, 1 = collision) for each of the four predefined traffic events per driving session. Collision rate serves as the ultimate measure of SA Level 3 projection failure: if the driver fails to anticipate the future state, a collision becomes inevitable. This metric is summarized per participant per session as a count of hazard collisions (0 to 4). A secondary metric, *total incident count*, captures all collisions recorded by the physics engine during a session, including incidental contacts that reflect overall driving quality and rule compliance.

Scialfa and colleagues demonstrated that experienced drivers are faster at detecting and responding to latent hazards due to more sophisticated mental models of risk [SDF+12]. Their finding implies that novice drivers, the target population of the present study, have the most to gain from an AR system that externalizes expert scanning patterns. Conversely, Reason et al. distinguished between driver *errors* (unintended failures of information processing) and *violations* (deliberate deviations from safe practices) [RMS+90]. This distinction is critical for interpreting the experimental results: a reduction in collisions during the AR-on condition (S2) suggests that the system reduces errors, whereas any sustained improvement in the post-training AR-off condition (S3) would indicate a genuine transfer of risk awareness, which is less likely to affect willful violations. The present experimental design therefore allows the study to disentangle these two mechanisms.

The following statistical framework is used to test the hypotheses. Inferential statistics are computed using *Welch's independent-samples t-test*, which does not assume equal population variances and is appropriate for the equal group sizes ($n = 12$ per group). The significance threshold is set at $\alpha = .05$ (two-tailed). To assess practical as well as statistical significance, *Cohen's d* with pooled standard deviation is reported alongside every p-value [Coh88]. Following Cohen's conventions, $d > 0.8$ is interpreted as a large effect. For within-subject comparisons across the three driving sessions (S1, S2, S3) for the immersive group, the *Wilcoxon signed-rank test* is applied to per-participant hazard collision counts. Because each participant produces a maximum of four collisions per session, the counts are ordinal in nature and the distribution cannot be assumed normal with $n = 12$. Effect sizes for the Wilcoxon test are reported as the matched-pair rank-biserial correlation coefficient, $r = Z/\sqrt{N}$.

To support the quantitative evaluation of driving performance and AR intervention efficacy, the simulation environment is equipped with an automated *data logging system*. During each experimental session, the system continuously records time-stamped driving metrics including vehicle state (e.g., velocity, acceleration, steering angle), hazard event logs (e.g., TTC values at warning onset, collision timestamps), and AR system states (e.g., alert type, activation timing). This logging framework ensures that all critical experimental variables are captured synchronously and non-intrusively, providing a complete and auditable data corpus for subsequent statistical analysis. The detailed technical implementation—including data storage schema, sampling frequency, and hardware integration—is presented in Chapter 5 as part of the system architecture description.

By anchoring all conclusions in both inferential significance and effect magnitude, and by triangulating objective collision data with subjective questionnaire responses and qualitative interview feedback, the methodology ensures that the findings are statistically defensible and practically meaningful. This framework provides a coherent evaluative bridge between the theoretical design of the AR system presented in this chapter and the empirical user study presented in Chapter 5.

4. AR Driving Simulation

Building upon the theoretical frameworks established in the previous chapter, this chapter details the technical design and implementation of the AR-enhanced driving simulator. The following sections describe the five-layer decoupled architecture, the procedurally generated urban environment, the parameter-driven traffic ecosystem, and the core AR-HUD and X-Ray interaction systems, establishing the technical validity necessary for the experimental evaluation in Chapter 5.

4.1. System Architecture

This section describes the underlying architectural design of the driving simulator. To achieve a balance between high-performance real-time simulation and systemic modularity, the simulator is engineered using a decoupled five-layer architecture (cf. Figure 4.1). This structural framework ensures efficient data synchronization and command orchestration across all subsystems, ranging from low-level physical computations to high-level Augmented Reality (AR) visual interventions. The design philosophy prioritizes the *Separation of Concerns* principle throughout: each layer exposes a well-defined interface to its neighbors and remains ignorant of the internal workings of non-adjacent layers, which is essential when the system must simultaneously maintain deterministic physics, procedural city generation, traffic flow management, and real-time AR rendering within the tight performance constraints of a consumer-grade Virtual Reality (VR) workstation.

The hierarchical architecture of the simulator is composed of five interconnected layers that ensure a seamless transition from low-level environmental physics to high-level user interaction. At the base, the *Foundation Layer* serves as the essential physical substrate. It utilizes a *Manager-of-Managers* architectural pattern, implemented via the `SceneManager`, to coordinate the `CustomCarController` and global state data. The `CustomCarController` encapsulates a full physics-based vehicle model, exposing a clean API for throttle, brake, and steering inputs while abstracting the underlying `WheelCollider` and `Rigidbody` computations. Superimposed on this is the *Traffic Management Layer*, which governs the complex ecosystem of non-player character (NPC) participants. By leveraging a centralized `TrafficSystem`, it orchestrates pathfinding and behavioral decision-making for vehicles and pedestrians across modular urban sectors. The `TrafficSystem` maintains a global pool of active `TrafficCar` and pedestrian instances, dynamically spawning and culling agents in response to the ego-vehicle's position to keep the scene populated without exceeding a configurable agent count.

The *Event Control Layer* acts as the central orchestrator of experimental logic, managing the triggering mechanisms via the `TrafficEventManager` to ensure that hazardous scenarios manifest precisely according to predefined research parameters. Each event implements the `ITrafficEvent` interface, which enforces a uniform lifecycle of initialization, execution, and teardown, allowing new scenarios to be added to the event library without modifying existing code. To facilitate empirical evaluation, the *Detection and Analysis Layer* provides continuous real-time monitoring of driver behavior. This layer integrates the `AccidentDetectionSystem` with statistical telemetry recording, offering a robust quantitative basis for the user study. The system tracks metrics such as reaction time to hazards, braking force profiles, and collision severity scores, all of which are written to a structured log for post-session analysis. Finally, the *User Interface Layer* sits at the apex, focusing on the high-fidelity rendering of the Augmented Reality Head-Up Display (AR-HUD). This layer translates risk data from the `AccidentDetectionSystem` into intuitive visual cues via an event-driven bus, ensuring that the driver's functional field of view remains uncluttered yet informative.

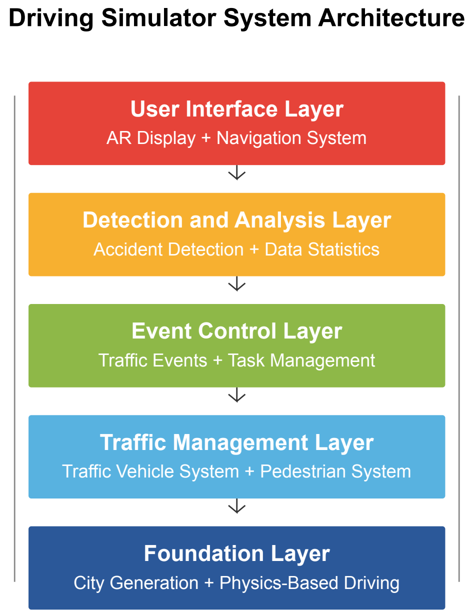


Figure 4.1.: The decoupled five-layer system architecture of the driving simulator

The primary advantage of this hierarchical design lies in the *Separation of Concerns* principle. By defining explicit interfaces and utilizing the *Singleton Pattern* for core managers, the architecture allows for the independent optimization of functional modules without compromising the integrity of underlying vehicle kinematics. This decoupling is instrumental in facilitating the complex hazard-warning experiments, as it ensures that the simulation environment remains deterministic, scalable, and highly reactive to real-time inputs. Inter-layer communication is deliberately restricted to upward-propagating events and downward-propagating commands, which prevents circular dependencies and makes the system straightforward to extend. For instance, when the `AccidentDetectionSystem` detects an imminent collision risk, it publishes an event onto the bus. The AR-HUD controller subscribes to this event and triggers the appropriate visual overlay without any direct coupling to the detection logic. This loose coupling is what makes it feasible to swap or upgrade individual layers.

The urban environment is built on top of the third-party *Fantastic City Generator* (FCG) [Fan25], a commercial Unity asset that provides road meshes, building prefabs, and `TrafficCar` models. The simulator uses only the FCG asset library while replacing its scene management entirely with a custom procedural system, so that the project retains full programmatic control over zone lifecycles, traffic spawning, and event delivery.

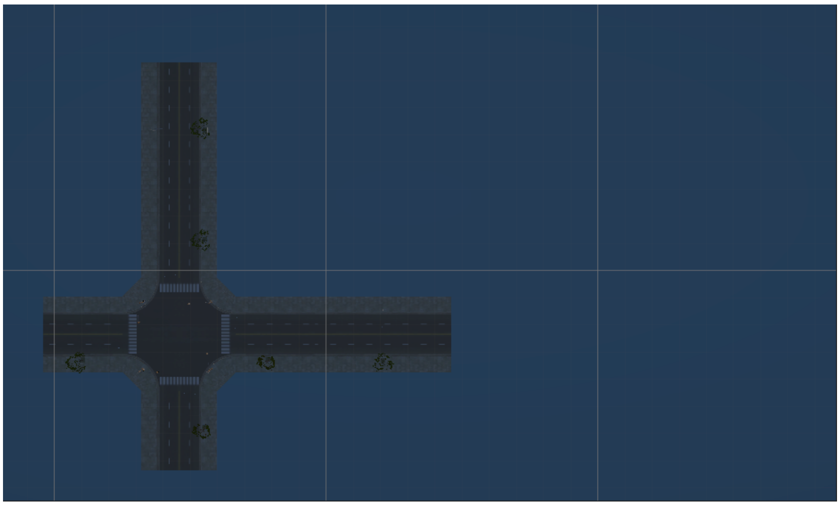
To maintain a sense of expansive virtual city while avoiding excessive system complexity, thereby ensuring simplicity and stability during simulation execution, the generation engine adopts a *Manhattan Grid* topology. Each city block is a self-contained Unity prefab placed on a regular $300\,\mathrm{m} \times 300\,\mathrm{m}$ grid. Constraining road connectivity to four cardinal directions simplifies waypoint routing, collision detection, and traffic-light placement while producing a visually convincing urban environment. The grid origin is fixed at world coordinates $(0, 0, 0)$, and every zone center is an integer multiple of 300 units in the XZ plane.

Zone state is maintained in two complementary data structures: a `HashSet<Vector3>` named `generatedZones` that prevents duplicate instantiation and provides $O(1)$ membership testing, and a `Dictionary<Vector3,GameObject>` named `zoneMap` that maps each zone center coordinate to its scene `GameObject`, enabling targeted destruction without a scene-wide search.

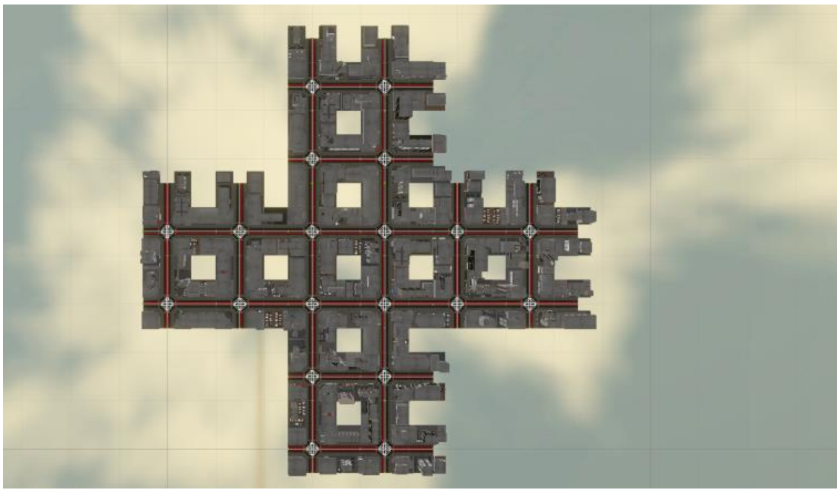
Because the local-space origin of the street prefab is not located at the road intersection center, direct grid partitioning would cause zone boundaries to cut through the middle of roads. Therefore, an empirically chosen offset of $(-400, 0, -400)$ is introduced to realign the coordinate system. At runtime, the system iterates through the set of already generated zones stored in `generatedZones`. For each zone center `zoneCenter`, it first applies the offset to compute the actual spatial position of that zone, and then checks whether the ego-vehicle's position falls within a square region of side length 300 m centered at that position. If a match is found, the corresponding zone center is returned. Otherwise, it returns an empty value. The corresponding implementation is shown in Listing.

When a zone change is detected, the system executes a sequential pipeline. The four cardinal neighbor blocks are instantiated if not already present via `generateFullZones()`. `RemoveUnusedZones()` then destroys any block no longer adjacent to the player, immediately releasing GPU geometry and texture allocations. A coroutine-based building placement routine populates the new street prefabs across multiple frames to avoid frame-rate spikes. The `markStreetBuilt()` pass renames all anchor nodes inside newly built zones so that the building generator does not attempt to repopulate them. `AddTrafficSystem()` rescans all waypoint containers and repopulates them with traffic vehicles. Finally, the event manager is notified via `trafficEventManager.OnZoneChange(randomEventSpawnSpot)`, which selects a hazard spawn point from the new zone's hierarchy and prepares the next scenario for delivery at the appropriate moment in the participant's drive.

To maintain visual continuity while respecting the strict computational budget required for a stable VR experience, the `CityGenerator` implements a *Dynamic Cross-Pattern Expansion* algorithm. The fundamental building block of this system is the street block prefab illustrated in Figure 4.2a. Each prefab is a lightweight, L-shaped module covering one corner of an intersection, so that four modules placed at a shared junction collectively form a complete city block with roads on all sides. As the driver moves forward, new modules are assembled into the forward arm of the cross while modules at the trailing end are dismantled. Because each module is compact, assembly and teardown complete within a single frame transition, and the driver is never presented with a loading boundary, from their perspective, the city continues indefinitely in every direction.



(a) The L-shaped street block prefab.



(b) Top-down view of the Dynamic Cross-Pattern Expansion on the Manhattan Grid.

Figure 4.2.: Procedural street generation logic.

The loading strategy maintains a cross (+) configuration of active urban blocks centered on the ego-vehicle's current grid sector, as shown in Figure 4.2b. A naive 3×3 grid would keep nine blocks loaded simultaneously, whereas the cross pattern keeps only five, the current zone and its four cardinal neighbors, reducing active mesh draw calls by approximately 44% compared to the square approach while still covering every direction the driver can face. This ensures that the driver's primary lines of sight, including the forward windshield view, rearview mirrors, and immediate lateral windows, are consistently populated with high-fidelity geometry, while discarding diagonal sectors that are never directly visible during normal driving.

The re-centering logic fires precisely when the vehicle crosses a sector boundary. When a new sector is entered, the algorithm re-centers the active matrix on the new coordinate and triggers block generation in all four cardinal directions. A Directional Retention Protocol preserves the block at the previous coordinate for at least one update cycle to prevent geometry clipping in the rearview mirrors, while any block disconnected from the cross pattern is immediately unloaded. The culling procedure is implemented in RemoveUnusedZones(), shown in Listing 4.1. This function retains only the five zones in the active cross pattern. All others are destroyed and removed from both the scene and the zone registry.

```
private void RemoveUnusedZones(Vector3 currentZoneCenter)
{
    HashSet<Vector3> activeZones = new HashSet<Vector3>
    {
        currentZoneCenter,
        currentZoneCenter + forwardZone,
        currentZoneCenter + backZone,
        currentZoneCenter + leftZone,
        currentZoneCenter + rightZone
    };

    List<Vector3> zonesToRemove = new List<Vector3>();
    foreach (Vector3 zone in zoneMap.Keys)
        if (!activeZones.Contains(zone))
            zonesToRemove.Add(zone);

    foreach (Vector3 zone in zonesToRemove)
    {
        if (zoneMap.TryGetValue(zone, out GameObject street))
        {
            Destroy(street);
            zoneMap.Remove(zone);
            generatedZones.Remove(zone);
        }
    }

    if (camera != null)
        RemoveVehiclesOutsideRange(camera.transform.position, 100f);
}
```

Listing 4.1: Scene culling logic

The coroutine-based building placement mechanism is central to maintaining frame-rate stability during zone transitions. Rather than instantiating all building prefabs synchronously, which would stall Unity's main thread for tens of milliseconds, GenerateNewZoneBuildingsCoroutine() yields after processing each anchor point, spreading the instantiation cost across multiple frames. The building-selection algorithm uses a weighted frequency counter to avoid visual repetition: proximity to a designated downtown anchor determines whether a tall commercial prefab or a lower-density residential prefab is selected, with a randomized probability preventing the same model from appearing in consecutive slots, producing a visually varied streetscape without requiring manual level design.

4.2. Traffic Ecosystem

The driving simulator must populate its procedurally generated city with a believable and physically grounded traffic ecosystem. This encompasses three interconnected concerns: the navigation logic that moves NPC vehicles through the city via the *Fantastic City Generator* waypoint network. The custom-developed pedestrian system that provides precise agent control for hazard scenario delivery. And the kinematic fidelity of the player-controlled vehicle, which must respond to inputs in a manner consistent with real-world physics. To enhance the realism of the simulation, the system further incorporates a stochastic behavioral layer that modulates NPC parameters—such as braking intensity and crossing speeds—to simulate the inherent unpredictability of urban traffic. Together, these elements ensure the simulation serves as a credible proxy for genuine road driving.

NPC vehicle navigation is built on the *Fantastic City Generator* (FCG) [Fan25] waypoint network, where agents follow bidirectional paths and junction merge points. Each `TrafficCar` agent computes steering based on waypoint offsets and manages longitudinal safety via front-facing raycasts. If a ray detects an obstacle within 15 m, a proportional braking force is applied, ensuring smooth deceleration or emergency stops. Intersection conflicts are resolved through a stop-line trigger system: invisible colliders are activated during red phases—forcing maximum brake torque—and deactivated during green phases. This waypoint-centric approach provides a consistent urban context while minimizing the computational overhead of the NPC physics model.

The NPC lifecycle is tightly synchronized with the procedural city generation described in Section 4.1. The `TrafficSystem` spawns agents at precomputed points outside the player's view, while a self-destruct routine removes vehicles that exit the visual range or fall below the terrain geometry. Crucially, when `RemoveUnusedZones()` retires a city block, the system triggers an atomic cleanup: all `TrafficCar` agents associated with the removed block's waypoints are immediately returned to the object pool. This synchronized teardown prevents orphaned agents from consuming the physics budget in invalidated regions, ensuring that computational resources are focused strictly on the player's active vicinity.

The deterministic design of this traffic system is a prerequisite for the experimental validity of the study. Because all NPC behavior is parameterized by the waypoint graph rather than procedural AI, creating new hazard events requires only modifying an agent's state—such as the `laneOffset` for lane-change events or reduced target speeds for rear-end collisions—without altering the underlying navigation logic. This separation between topology and behavioral state allows the `ITrafficEvent` interface to remain minimal, enabling new hazards to be authored independently. Furthermore, embedding the waypoint graph within street prefabs ensures that NPC trajectories remain fully reproducible, allowing for precise control over hazard timing across all participant trials.

The pedestrian system is developed entirely in-house, independent of the FCG asset. Its foundation is a set of pre-authored pedestrian routes hand-crafted at scene design time and embedded directly inside every street block prefab. Within each prefab, a container `GameObject` named `PedestrianPaths` holds one or more child `PedestrianPath` objects. A `PedestrianPath` is essentially a pre-defined pedestrian route implemented using a waypoint-based approach: the developer manually places waypoints one by one in the scene, connecting them in sequence to form a complete walking path. Each `PedestrianPath` contains two sub-containers: a `Points` object whose immediate children are the ordered sequence of `Transform` waypoints defining the route geometry, and a `People` container that serves as the runtime parent for all pedestrian instances spawned on that route. Figure 4.3 illustrates this hierarchy in the Unity scene view.

Figure 4.3.: Pedestrians following pre-authored `PedestrianPaths` routes.

The routes follow the edges of the road network, along pavements and across pedestrian crossings, rather than cutting diagonally through the scene. Where a route passes through a signalized crossing, the designer places a crossing waypoint between two invisible kerb trigger volumes, ensuring that the signal-compliance logic is automatically engaged for any agent following that route. Because every route is fully specified before runtime, the spatial trajectory of every agent is deterministic: a pedestrian spawns at the first waypoint, walks in sequence to the last, then either loops back or despawns depending on the path mode. This predictability is what makes it possible to script hazard events, such as the Bus-Block scenario, that depend on a pedestrian occupying a precise position at a known moment.

Crucially, the `PedestrianPaths` container is part of the street block prefab itself. This means pedestrian routes are instantiated and destroyed together with the road geometry as the procedural city expands around the player. When a new city block is spawned, all embedded `PedestrianPath` objects become active immediately, and the `ZoneWaypointManager` discovers and registers them automatically on `Awake()`. When a block is removed from the active cross pattern, its paths and all agents are cleaned up as part of the same atomic teardown described in Section 4.2.1.

The system is designed around a three-tier architecture that balances scene-wide coordination with individual behavioral precision. The top tier, `ZoneWaypointManager`, maintains a registry of all pedestrian paths active across the currently loaded city blocks. Each path stores an ordered waypoint sequence and a movement mode: *looping*, for continuous patrol behaviors, or *linear*, for one-directional crossings. This distinction is consequential for hazard reproducibility, the Bus-Block event requires a linear path that positions a pedestrian precisely behind a stationary bus at the moment of trigger, because a looping agent with non-deterministic phase would make the hazard timing invalid as a controlled experimental stimulus. The middle tier, `BasePathManager`, instantiates and recycles pedestrian prefabs from an object pool, assigning each agent to a path from the zone registry. Pooling is essential: synchronous instantiation of `Rigidbody`-bearing prefabs on every zone transition would introduce CPU spikes incompatible with the 90 FPS target required for an immersive VR session. Before any agent is spawned, `PedestrianPath` computes a uniform initial distribution along the route to avoid the visual artifact of an entire crowd clustering at the path origin. The total path length is divided by the target pedestrian count to derive a nominal segment length, and each agent's spawn distance is then offset by a uniform random jitter of $\pm 30\%$, clamped to the valid path extent.

Individual agents at the bottom tier are driven by `PedestrianMover` and `BicycleMover`. Both share the same waypoint-following logic but expose separate speed and animation parameters, allowing the bicycle agent used in the Right-Turn Blind Spot event to move faster while reusing the pedestrian path infrastructure entirely. Each agent advances via `Vector3.MoveTowards` on the XZ plane and rotates toward the target waypoint using `Vector3.RotateTowards` at 3 rad/s. A small per-agent random lateral offset of ± 0.3 m is applied each time a new waypoint is reached, preventing the robotic column appearance that arises when all agents share identical waypoint coordinates. Each pedestrian is equipped with a Capsule Collider rather than a mesh collider, and the rounded base allows agents to glide over minor geometric irregularities in the procedurally generated terrain without becoming stuck, while reducing collision detection overhead in scenes with high agent density. Figure 4.4 illustrates this configuration. The rounded base prevents the agent from becoming stuck on uneven sidewalk geometry, ensuring smooth traversal across procedurally generated terrain.

Traffic-light compliance is implemented through a two-collider geometry technique. Each signalized crossing has two invisible trigger volumes placed on opposite kerbs. When an agent's OnTriggerEnter fires, a between-segments predicate tests whether the agent's

Figure 4.4.: Pedestrian agent with a Capsule Collider.

position lies between the two kerb colliders. This predicate prevents agents that have already entered the crossing from being stopped mid-road. If the agent has not yet crossed the first kerb, movement is frozen and the animator switches to an idle state. The implementation is shown in Listing 4.2.

```
private void pedestrianTrafficStop(Collider other)
{
    bool isMiddle = isBetweenColliders(other.transform);
    float probability = Random.value;

    if (other.name.Contains("P-Collider") && !isMiddle && !(probability < 0.1f))
    {
        currentStopCollider = other;
        isStoppedByStopCollider = true;
        if (animator) animator.Play("idle1");

        Rigidbody rb = GetComponent<Rigidbody>();
        if (rb != null) rb.constraints = RigidbodyConstraints.FreezeRotation;
    }
}
```

Listing 4.2: Traffic-light compliance in `PedestrianMover.cs`

Every subsequent update checks whether the stop collider's `GameObject` is still active, and when the traffic light turns green and deactivates it, `ClearStopCollider()` restores free movement. A configurable `jaywalkProbability` of 10% (the `probability 0.1f` branch above) is sampled on every trigger event, allowing a fraction of agents to ignore the red signal, the concrete mechanism by which the cultural adaptation layer modulates pedestrian compliance.

The pedestrian system implements basic collision response: when struck above a speed threshold, the pedestrian is knocked down and removed after a short delay, allowing the accident detection system to record the event. For performance optimization, the `ZoneWaypointManager` runs a coroutine every three seconds that destroys pedestrians beyond 100 m from the player camera and cleans up all remaining agents when their parent zone is destroyed.

The player-controlled vehicle is governed by the `CustomCarController` component, a bespoke physics model built on Unity's `WheelCollider` API. The controller was adapted from an open-source reference implementation and extended with ABS pulse-modulation, ESC stabilization, and an anti-roll bar to satisfy the *Kinetic Physical Realism Threshold* (PRT) criterion described in Chapter 3. Unlike the lightweight FCG NPC model, the player controller replicates the kinematic principles of propulsion, deceleration, and directional stability found in real vehicles, giving the driver a physical sense of mass, inertia, and tire-road adhesion.

The acceleration profile is governed by a configurable `motorTorque` of 1500, Nm, modulated by an `AnimationCurve` that mimics the torque-speed characteristic of an internal combustion engine. The `Rigidbody` mass is 1200, kg, with the center of mass positioned 0.5, m below the geometric center to resist rear-end lift under braking. The braking system implements ABS pulse-modulation : when wheel slip exceeds 0.3, brake torque is halved to prevent lock-up. The steering system reduces the maximum lock angle from 35° at low speed to approximately 50% at 100, km/h, and an ESC layer applies corrective forces when excessive lateral velocity is detected. Collectively, these mechanisms ensure that measured hazard responses reflect genuine cognitive processing of traffic risk rather than adaptation to an unfamiliar handling model.

4.3. Hazard Scenario Design

The scientific validity of a driving simulator depends not only on the fidelity of its environment and vehicle physics, but on the precision with which hazardous stimuli are delivered to the participant. A scenario that triggers too early affords the driver ample time to respond without any AR assistance. One that triggers too late forfeits ecological realism and compromises the measurement of genuine reaction behavior. This section describes the complete pipeline through which hazard scenarios are designed, scheduled, activated, and resolved: from the extensible software architecture of the Traffic Event Manager, through the implementation of the four concrete hazard types that constitute the experimental stimuli, to the notification pathway that connects scenario execution to the AR-HUD. Together, these components form the experimental core of the study and are directly responsible for generating the conditions under which RQ1 and RQ2 are evaluated.

The Traffic Event Manager (`TrafficEventManager`, hereafter TEM) is the central orchestrator of all hazardous scenarios in the simulation. It operates as a high-level state machine positioned between the city generation pipeline and the AR-HUD display layer, receiving zone-change notifications from the procedural environment and forwarding hazard state information to the notification layer. Rather than encoding individual scenario logic into a monolithic update loop, the TEM delegates all hazard-specific behavior to self-contained event objects, each implementing a common `ITrafficEvent` interface. This design reflects the Open-Closed Principle of software engineering: the TEM is open to extension by new hazard types but closed to modification, which is what makes the simulator a scalable research platform applicable to future studies beyond the four scenarios implemented here.

The `ITrafficEvent` interface defines the minimal contract every hazard must satisfy:

```
public interface ITrafficEvent
{
    bool   IsEventActuallyTriggered { get; }
    bool   IsTriggered              { get; }
    string eventDescrip             { get; }
    string EventName                { get; }
    void   Trigger();
    void   Cleanup();
}
```

Listing 4.3: ITrafficEvent interface

The distinction between `IsTriggered` and `IsEventActuallyTriggered` is architecturally critical. `IsTriggered` is set to `true` as soon as `Trigger()` is called, which may occur hundreds of meters before any visible danger manifests, when the zone first enters the active cross pattern. `IsEventActuallyTriggered`, by contrast, is a live computed property that reflects whether the dangerous condition is currently active and observable. In the Bus-Block event, for instance, `IsEventActuallyTriggered` reads `busController.eventTriggered` directly, which only becomes `true` when the player closes to within 20 m of the bus and the pedestrian is about to emerge. This two-state design ensures that hazard notifications reach the driver precisely when the danger becomes imminent, not when the scenario is loaded into memory, a distinction that is essential for the ecological validity of the reaction-time measurements collected in the evaluation study.

The `Cleanup()` method is the counterpart to `Trigger()` and is equally important for experimental integrity. Every event is required to restore all modified NPC status flags, release any overridden brake torques, and destroy spawned `GameObject` instances before returning control to the normal traffic loop. The TEM invokes `Cleanup()` both when a conflict is explicitly resolved and when the player moves beyond a `cleanupDistance` of 80 m from the event anchor, ensuring the scene state is fully reset before the next encounter. This guarantees that each trial begins from a consistent environmental baseline, which is a prerequisite for between-trial comparisons in the evaluation data.

When the city generation pipeline spawns a new zone, it calls trafficEventManager.OnZoneChange(Transform eventSpot), passing the world transform of a named anchor point embedded in the street prefab. The TEM inspects the anchor's name and invokes the corresponding factory method.

This name-based dispatch keeps the city generation system entirely unaware of event semantics. Level designers need only place a correctly named anchor `Transform` in the street prefab. The TEM handles all instantiation and lifecycle management at runtime.

Adding a new hazard type to the platform requires only creating a class that implements `ITrafficEvent`, registering it in the dispatch table, and placing the corresponding anchor in the prefab, a process that requires no modifications to any existing code.

For events that require an NPC vehicle that may not yet be nearby at zone load time, specifically the Lane Change and Rear-End events, the TEM employs a polling strategy via `InvokeRepeating` at 200 ms intervals. Each event's `CheckAndSelectEventVehicle()` method scans the active `CarContainer` hierarchy, applies spatial and directional filters, and returns `true` once a suitable NPC is identified and locked. Polling is immediately cancelled upon success to avoid sustained overhead. Every update tick, the TEM iterates its four event references and forwards the first event with `IsEventActuallyTriggered == true` to `AR_HUD.SetCurrentEvent()`, providing the warning text and event identifier that appear on the windshield. Only one event is reported at a time to prevent display overlap, and the HUD clears automatically on zone transition or explicit event resolution.

Four concrete traffic hazard scenarios have been implemented for the current study. Two of these, the Bus-Block Hidden Pedestrian and the Right-Turn Blind-Spot Bicycle, are the primary experimental stimuli for evaluating the AR-HUD's visualization capability and are described in full technical detail below. The remaining two, Sudden Lane Change and Rear-End Collision, serve as secondary stimuli that broaden the typological range of hazards and are described in outline.

This scenario is the core experimental stimulus of the thesis: a pedestrian occluded behind a stationary bus emerges into the player's path at a moment calibrated to exceed normal reaction capacity without AR assistance. The event is implemented across two cooperating classes, `BusBlockPedestrianEvent` and `BusBlockController`, and proceeds through a precisely sequenced lifecycle.

At zone load time, the event instantiates the bus mesh prefab at the anchor position and prepares the AR visualization assets described in Section 4.4. Before the `BusBlockController` component is attached, a PreDisablePedestriansBeforeController() coroutine retries up to twenty frames to locate and disable the `PedestrianMover` on the designated pedestrian path, ensuring the target pedestrian stands motionless in the bus's visual shadow before the player approaches. This retry loop is necessary because the pedestrian system and the event system initialize concurrently, and the pedestrian prefab may not yet have completed its own `Awake()` cycle when the event fires.

The `BusBlockController` then begins a proximity monitoring coroutine that wakes every 0.3 s and measures the distance from the player vehicle to the bus. The 80 m warning threshold mimics the behavior of a Vehicle-to-Everything (V2X) system, where roadside infrastructure would detect the occluded pedestrian and broadcast its position to approaching vehicles. The simulator faithfully reproduces this behavioral outcome, advance warning of a non-visible hazard, without implementing a full V2X communication stack, a simplification that preserves the essential experimental condition while remaining computationally tractable.

At 80 m, the AR visualization pipeline is activated, allowing the driver to perceive the occluded pedestrian through the bus geometry via a mechanism described in full in Section 4.4. At 20 m, `TriggerEvent()` fires: the bus begins to pull away from the kerb, the `PedestrianMover` is re-enabled, and the pedestrian steps out into the roadway. The 20 m threshold was selected to place the pedestrian's emergence at a Time-to-Collision (TTC) of approximately 2.5 s for a vehicle approaching at 30 km/h , the nominal limit of standard emergency braking reaction time established in automotive safety literature. At this moment `IsEventActuallyTriggered` becomes `true`, and the TEM forwards the hazard notification to the display layer.

This scenario targets a structurally different class of hidden hazard: a fast-moving road user that is physically present in the scene but occupies the blind spot created by the vehicle's B-pillar and rear-right quarter during a right-hand turn maneuver. The event is implemented across three cooperating classes: `BicycleTurnRightEvent`, `BikeController`, and the shared `BicycleMover` infrastructure from the pedestrian system described in Section 4.2.

At trigger time, BicycleTurnRightEvent locates the dedicated path path4_bicycleAccid inside the current zone's `PedestrianPaths` container, activates its `GameObject`, retrieves the spawned bicycle NPC, and sets `runSpeed` to 15 m/s, approximately 54 km/h, consistent with a fast urban cyclist. `PedestrianMover` is immediately disabled to freeze the bicycle in position, and a `BikeController` component is dynamically attached to the NPC, which hides all renderers on its first frame, making the bicycle invisible to the driver. The controller polls every 0.3 s via `checkPlayerPass()`, using `Utils.IsInFront(playerPos, bikePos)` to determine when the player has crossed the bicycle's lateral position. Once that condition is met, a one-second delay coroutine fires, after which `PedestrianMover` is re-enabled, all renderers are restored, and the bicycle accelerates across the player's turning path. `IsEventActuallyTriggered` is set to `true` at this moment, forwarding the warning to the display layer.

The deliberate sequence of hide, player-passes, delay, reveal is the pedagogical mechanism of this scenario. A driver relying solely on mirrors will fail to detect the bicycle, because it is not visible during the approach phase. A driver receiving the AR-HUD alert will have been pre-warned before initiating the turn. This asymmetry creates a clean experimental contrast between AR and no-AR conditions. The reuse of `BicycleMover` from the pedestrian system also demonstrates the infrastructure's extensibility in practice: the bicycle NPC required no new navigation code, only a dedicated behavioral controller for the trigger sequence, which is precisely the pattern the `ITrafficEvent` interface was designed to enable.

This event commandeers a moving NPC vehicle travelling in the same direction as the player and causes it to drift laterally into the player's lane. Vehicle selection applies three filters: distance ≤ 15 m, a forward dot product > 0.9 confirming a nearly parallel heading, and a 99% probability gate that introduces controlled non-determinism. Once selected, the FCG `TrafficCar.laneOffset` field is smoothly interpolated via `Mathf.MoveTowards` until the NPC crosses the lane boundary. After a configurable `autoReturnDelay`, the offset returns to zero, clearing the lane without leaving a permanently displaced vehicle. The original waypoint navigation remains entirely intact throughout, since the intrusion is achieved solely by modifying the lateral offset applied to the steering calculation rather than altering the waypoint graph.

This event selects a vehicle within 30 m ahead of the player and forces it to stop abruptly through three sequential steps: the NPC's `status` is overridden to `StatusCar.bloked`, preventing the FCG AI from issuing new motor commands, `rb.linearVelocity` and `rb.angularVelocity` are set directly to `Vector3.zero`, producing an instantaneous stop that the normal FCG braking gradient cannot replicate, and the brake torque on every wheel collider is set to `carSetting.brakePower * 999f` to prevent any inadvertent re-acceleration. Cleanup restores `status = transitingNormally` and releases all brake torques, allowing the NPC to re-enter the normal waypoint flow as a standard traffic participant. This event is intentionally the simplest of the four: its experimental value lies not in spatial complexity but in the acute time pressure it places on the driver's forward attention, providing a complementary stimulus to the occlusion-based scenarios.

The final stage of the event control pipeline is the translation of hazard state into actionable visual information on the AR-HUD. Managed by the `ARHUDController`, the display canvas is spatially parented to the vehicle's windshield transform, ensuring that all HUD elements remain rigidly aligned with the driver's forward line of sight, a design that mirrors the optical principle of real-world automotive head-up display systems. When `IsEventActuallyTriggered` becomes `true` for an active scenario, the TEM forwards the event's name and description string to the HUD, which renders a graded warning, transitioning from amber to red as the TTC approaches the critical intervention threshold, with smooth alpha transitions to avoid abrupt visual onsets in VR. This notification layer is intentionally decoupled from scenario logic: it operates purely as a display terminal that receives hazard state from the TEM and renders it without knowledge of how that state was produced. The more advanced AR visualization mechanisms built on top of this foundation, including the X-Ray occlusion rendering used in the Bus-Block scenario, are described in Section 4.4.

4.4. AR Visualization System

The preceding section described how hazard scenarios are orchestrated and how the Accident Detection System classifies collision risk. This section addresses the final layer of the simulator: how that risk information is transformed into visual interventions that reach the driver. The system operates across three interconnected components. The *AR Head-Up Display* projects driving telemetry, navigation guidance, and hazard warnings onto the windshield as spatially anchored overlays. The *Accident Detection and Prediction System* continuously monitors the scene for collision precursors, classifies their severity, and determines what information the HUD should convey. The *X-Ray Rendering Pipeline* constitutes the primary innovation of this thesis: it makes occluded pedestrians visible to the driver through opaque geometry before they physically emerge, simulating the perceptual capability that a V2X-equipped vehicle would provide in a real-world deployment. Together, these three components operationalize the AR safety intervention whose effect on driver behavior is the central empirical question of the study.

The AR-HUD is realized as a Unity World Space Canvas parented directly to the player vehicle's windshield transform, ensuring that all displayed elements move rigidly with the car and remain aligned with the driver's forward line of sight regardless of head orientation within the VR headset. This spatial binding produces the *conformal* AR effect: digital overlays appear to float on the glass surface at a consistent depth rather than being screen-fixed elements that swim against the background when the driver moves their head. To reduce the accommodation-vergence conflict that causes visual fatigue in VR driving tasks, the canvas is positioned at a virtual image distance of approximately 2 to 3 m in front of the driver's eyes, placing the HUD's focal plane close to the near-road zone where gaze is concentrated during hazard scenarios.

The display is organized into three information regions with deliberately different update cadences and visual weights, following the cognitive load management principle that safety-critical information should always dominate without occluding primary driving cues. The lower portion renders real-time vehicle telemetry derived from `CustomCarController`: current speed, throttle and brake state, and a grounded indicator. Navigation guidance is rendered on a separate `Arrow_Canvas` layer via the `PathFinder` component, which runs on a 200 ms `InvokeRepeating` timer and classifies the signed angle between the car's forward vector and the target into three zones: forward ($|\theta| < 45°$), right ($\theta > 45°$), or left ($\theta < -45$). The directional arrow cross-fades smoothly to the new direction via a guarded transition animation, preventing visual flicker if the driver's heading changes mid-transition. A secondary region shows running accident statistics accumulated by the `AccidentDetectionSystem`, total accidents, total collisions, and the derived per-hour accident rate that serves as the primary dependent variable for RQ1. Displaying this in real time creates an immediate feedback loop that reinforces the *Simulation-to-Awareness* pedagogical goal.

Warning text is displayed with a severity-coded color scheme: amber for events at the notification threshold (TTC ≈ 2.5 s), transitioning to red as TTC approaches the critical 1.5 s limit. For critical-severity events, those involving pedestrians or a risk score exceeding 80, the system triggers a full-screen `ScreenFlashEffect` coroutine that animates a red overlay through a 150 ms fade-in, 200 ms hold, and 400 ms fade-out. This peripheral flash exploits the high luminance sensitivity of human peripheral vision to redirect attention even when the driver is focused on a central region of the road.

Both coroutines are managed with stored references so that a new warning arriving before the previous flash expires will restart the animation atomically rather than producing overlapping effects. The `SetCurrentEvent()` method prevents flicker by only resetting the five-second display timer when the incoming event name differs from the currently displayed one, avoiding repeated visual resets within a single ongoing event. Figure 4.5 shows the AR-HUD interface during a critical hazard event, where the red `ScreenFlashEffect` overlay is triggered by a low TTC value, exploiting peripheral luminance sensitivity to redirect driver attention while maintaining the visibility of core telemetry and navigation cues.

The `AccidentDetectionSystem` (ADS) is the analytical layer that bridges the event control pipeline and the AR-HUD. It performs two complementary functions: *event-informed pre-collision warning*, receiving hazard notifications from the TEM and translating them into graded display commands, and *sensor-driven collision detection*, continuously monit-

Figure 4.5.: AR-HUD interface during a critical hazard event.

oring the scene for collision precursors that fall outside any scripted scenario. Together, these approximate the capability profile of a real-world ADAS.

In the simulation, full V2X communication is approximated by a direct event notification architecture, which preserves the essential property of the V2X model, the driver is warned of a hazard that is not yet visible, while remaining computationally tractable within the Unity real-time constraint. When the TEM's `IsEventActuallyTriggered` transitions to `true`, it calls `AccidentDetectionSystem.NotifyHazard()`, passing the event identity, the hazard description string, and the world transform of the hazard object. This is analogous to a V2X message arriving at the vehicle: the ADS did not discover the hazard through physics queries, it received a localized notification from an external oracle, exactly as a real ADAS would receive a V2X broadcast. The ADS then forwards the warning to the HUD and begins logging the session data stream at 30 Hz.

In parallel with event notifications, the ADS implements a scene-wide predictive monitoring layer that operates continuously regardless of whether a scripted scenario is active. Rather than querying every physics object in the Manhattan Grid every frame, which would be prohibitively expensive given the number of simultaneously active NPC vehicles, the system performs a spatial pre-filter using `Physics.OverlapSphere` within a 100 m radius and then selects the five closest NPC vehicles by Euclidean distance. This *priority threat filtering* concentrates computational resources on the immediate environment and reflects the threat-priority architecture of real automotive radar systems, which similarly track a bounded number of confirmed targets at any given moment.

For each of the five priority targets, the system samples position and velocity at 30 Hz and computes TTC as shown in Equation 4.1:

$$\mathrm{TTC} = \frac{d}{|\Delta v|}, \quad \Delta v = v_{\text{player}} - v_{\text{NPC}}, \tag{4.1}$$

where d is the relative distance and Δv is the closing speed. The system additionally monitors two behavioral anomaly flags: a sudden braking flag, raised when an NPC's longitudinal acceleration falls below $-5.0\,\mathrm{m/s^2}$, and a lane-intrusion flag, raised when an NPC's `laneOffset` is interpolating rapidly toward the player's lateral position. When either anomaly flag is set and TTC falls below 2.5 s, the ADS classifies the condition as a *critical anomaly* and pushes a collision warning to the HUD. This predictive layer thus functions as an autonomous complement to the event-triggered notifications, ensuring that unscripted near-miss situations, for example, an NPC vehicle braking sharply in front of the player between scripted events, are also detected and communicated.

When a collision occurs, `VehicleCollisionDetector` component on the player vehicle fires `OnCollisionEnter`, assembles a `CollisionData` record containing the contact-point normal, the relative velocity at impact, and the identity and type of the colliding object, and forwards it to the ADS's `AnalyzeCollision()` method. The ADS computes a composite risk score from three independent factors, as summarized in Table 4.1.

Table 4.1.: Risk score factors used by the `AccidentDetectionSystem` for collision severity classification.

Factor	**Max Points**	**Formula**
Vehicle speed	40	$\min(40,\ v_{\text{impact}} \times 0.8)$
Collision force	30	$\min(30,\ F_{\text{impact}} \times 2.0)$
Object type	30	Pedestrian = 10, Bicycle = 8, Vehicle = 5, Static = 2

A collision is classified as an *accident* when all three conditions hold simultaneously: vehicle speed $\geq 5\,\mathrm{km/h}$, collision force $\geq 10\,\mathrm{N{\cdot}s}$, and total risk score ≥ 60. Accidents are further stratified: a score ≥ 80 or any collision with a pedestrian is classified as *critical*. A score ≥ 70 or any collision with a cyclist is classified as *serious*. All other threshold-passing events are *minor*. Sub-threshold contacts are still recorded as warnings, preserving a complete driving record without triggering the full accident response pipeline. The running statistics, `totalAccidents`, `totalCollisions`, and `drivingTime`, accumulate throughout the session, and the derived property shown in Equation 4.2:

$$R_{\text{accident}} = \frac{\texttt{totalAccidents}}{\texttt{drivingTime}} \times 3600 \tag{4.2}$$

gives the per-hour accident rate that serves as the primary dependent variable for the statistical analysis of RQ1 in Chapter 5.

The X-Ray visualization system is the primary technical innovation of this thesis. It addresses the core limitation that motivates RQ1: a driver cannot react to a hazard they cannot see. In real-world traffic, large vehicles such as buses or trucks frequently occlude pedestrians, cyclists, and children at pedestrian crossings. The X-Ray system simulates the perceptual extension that a V2X-equipped vehicle would provide by making the occluded road user visible to the driver as a highlighted silhouette *through* the occluding geometry, before the hazard physically enters the driver's line of sight.

The technical implementation is built on a two-pass render-queue trick using Unity's Stencil Buffer and render-queue ordering, rather than post-process image effects or texture overlays. The system requires no additional cameras or render textures, making it compatible with the strict frame-rate budget of a VR session.

Pass 1, Invisible Mask shader. The bus mesh is assigned the Custom/InvisibleMask shader, which renders at queue Transparent+1 (queue value 3001), as shown in Listing 4.4:

```
Shader "Custom/InvisibleMask" {
    SubShader {
        Tags { "Queue" = "Transparent+1" }
        Pass {
            Blend Zero One
        }
    }
}
```

Listing 4.4: Invisible Mask shader

The `Blend Zero One` mode multiplies the destination color by zero, effectively zeroing all color information in the pixels covered by the bus silhouette after the bus's own opaque geometry (rendered at queue 2000) has already written depth. The result is that those pixels are cleared in the framebuffer while the depth buffer retains the bus's geometry, preventing objects *behind* the bus from being overdrawn by normal forward rendering.

The `XRayBlockObject` component is attached to the target pedestrian's `Renderer` and elevates its materials to render queue 3002, as shown in Listing 4.5:

```
foreach (var mat in GetComponent<Renderer>().materials)
    mat.renderQueue = 3002;
```

Listing 4.5: Elevated render queue for X-Ray pedestrian

Because queue 3002 is processed after the invisible mask (3001), the pedestrian's pixels are written into the zeroed framebuffer region. Since the depth buffer still holds the bus's depth values, the pedestrian is composited *in front of* the bus in screen space, producing the see-through silhouette effect. The full render-queue ordering is: opaque bus geometry (2000) → invisible mask (3001) → pedestrian silhouette (3002). Figure **??** illustrates the driver's view with and without the X-Ray system active.

In the Bus-Block Hidden Pedestrian scenario, as shown in Figure **??**, the pedestrian is completely occluded by the bus without AR, creating a blind-spot hazard that the driver cannot anticipate. With X-Ray AR active, the pedestrian is rendered as a glowing silhouette through the bus geometry, simulating V2X-informed perception. The render-queue technique makes the occluded pedestrian visible at 80 m, providing the driver with approximately 2.5 s of additional reaction time at urban approach speeds.

To increase the visual salience of the silhouette under varying lighting conditions, the `BusBlockController` clones each material on the pedestrian and enables emission, giving the figure a yellow glow inside the see-through window. The ScannerPlane is repositioned each frame at 10% of the distance from player to pedestrian, with a configurable lateral offset, and is continuously rotated to face the player. This dynamic placement ensures

Figure 4.6.: Without AR.

Figure 4.7.: X-Ray AR active.

the visibility cone remains centered on the pedestrian regardless of the driver's approach angle or the pedestrian's position behind the bus.

The X-Ray effect is not permanently active: it is enabled by the proximity logic in `BusBlockController` only when the player closes to within 80 m of the bus, and disabled upon event cleanup. This context-aware behavior is important for two reasons. Perceptually, a constantly visible see-through effect on all large vehicles would be distracting and would reduce the salience of the warning at the critical moment. Experimentally, restricting activation to the scripted scenario window ensures that the X-Ray stimulus is temporally bounded, making the participant's response attributable to the AR intervention rather than to a general habituation to transparent geometry in the scene. By designing the system as a targeted, event-gated intervention rather than a permanent rendering mode, the X-Ray pipeline mirrors the behavior of real-world driver assistance alerts that activate only when a specific threat condition is confirmed, reinforcing the ecological validity of the V2X simulation framework.

5. Evaluation

This chapter presents the empirical evaluation of the AR-enhanced VR driving simulator developed in Chapter 4. The evaluation is structured around the two central research questions formulated in Chapter 1: whether an AR system that reveals hidden or sudden traffic hazards improves driver safety and reduces accident rates in complex urban conditions (RQ1), and whether AR-enhanced driving simulation can effectively improve drivers' hazard awareness after training (RQ2). To address both questions simultaneously within a single cohort, the VR group follows a three-session within-subjects protocol that provides independent objective evidence for each research question, while a Video group supplies a subjective experience baseline for questionnaire comparisons. The chapter is organized as follows: Section 5.1 describes the experimental design and dependent variables. Section 5.2 covers the participant sample and session procedure. Section 5.3 presents the quantitative findings. and Section 5.4 discusses the results in relation to the hypotheses, limitations, and future directions.

5.1. Experimental Design

The central methodological challenge in evaluating the simulator is that RQ1 and RQ2 place different evidential demands on the experiment. RQ1 requires a direct comparison of driving performance *with* and *without* AR assistance, while RQ2 requires evidence that any improvements persist after AR assistance is removed. A single-session design cannot address both requirements objectively within the same participant group. A between-subjects design can address RQ1 by comparing an AR-enabled group against a no-AR control, but leaves RQ2 dependent solely on post-hoc self-report, since no participant is ever observed driving without AR after having trained with it. The three-session within-subjects protocol adopted here resolves this by giving every VR participant three consecutive drives: a no-AR baseline, an AR-assisted session, and a second no-AR session. This structure yields paired observations at the individual level for both comparisons of interest, increasing statistical power by eliminating between-participant variability as a confound.

Two participant groups are used. The *VR group* ($n = 12$) completes the three-session driving protocol and provides both objective performance metrics and post-training questionnaire responses. The *Video group* ($n = 12$) watches a recorded first-person playback of a competent but not exceptional performance and completes the same questionnaire immediately afterwards. The Video group does not interact with the simulator and contributes no objective driving data. Their responses are used exclusively to contextualize the VR group's subjective ratings by providing a passive-observation baseline for perceived immersion, AR utility, and educational value. The between-group questionnaire comparison thus captures the experiential surplus of active, AR-assisted simulation over passive observation of equivalent content. Table 5.1 summarizes the three-session structure and its correspondence to the research questions. Each session uses identical routes and hazard event order. Only the AR-HUD state differs.

Table 5.1.: Three-session within-subjects design for the VR group

Session	AR State	Primary Purpose	Addresses
S1	Disabled	No-AR baseline	Reference for RQ1 and RQ2
S2	Enabled	AR real-time benefit	RQ1
S3	Disabled	Training transfer test	RQ2

The two hypotheses formulated in Chapter 1 map directly onto this three-session structure. H1 predicts that collision counts and PRT will be lower in S2 (AR enabled) than in S1 (AR disabled), testing the real-time safety benefit of the AR-HUD system. H2 predicts that collision counts and PRT in S3 (AR disabled, post-training) will remain lower than in S1, testing whether the AR-assisted experience cultivates lasting hazard awareness.

Each VR participant completes three consecutive driving sessions of approximately two to three minutes each, separated by a short rest interval. In every session the participant is assigned the same task: drive from a fixed start point to a fixed end point through the procedurally generated Manhattan grid described in Chapter 4. During the drive, participants are required to obey traffic rules, respect traffic signals, and yield to pedestrians at crossings, as they would in a real urban environment. These baseline behavioral requirements ensure that the driving context is ecologically meaningful and that any incidental violations, such as running a red light or failing to yield, are captured in the total incident count alongside the four scripted hazard events.

The Traffic Event Manager triggers exactly four hazard events at predetermined spatial locations during each session, always in the same fixed order. Holding the route, event sequence, and spatial trigger positions constant across all three sessions is a deliberate design choice: it ensures that any performance differences between sessions reflect genuine changes in driving behavior rather than variation in scenario content or task difficulty. The only manipulated variable is the state of the AR-HUD system. The four implemented scenarios are:

1. **Bus-Block Hidden Pedestrian (BB): The Ghost Probe Event.** This scenario targets one of the most insidious urban collision patterns, colloquially termed the Ghost Probe in Chinese traffic safety discourse. A static bus is positioned in the right lane, creating a complete visual occlusion of the sidewalk and crosswalk area. As the participant's vehicle approaches within a 45-meter trigger radius, a pedestrian agent initiates a crossing trajectory from behind the bus at a normal walking pace. From the driver's unaided perspective in Session 1 and Session 3, the pedestrian is physically invisible until the moment they emerge from the front bumper of the

bus, leaving a Time-to-Collision (TTC) of less than 1.5 seconds, insufficient for a comfortable emergency stop. This event is specifically paired with the X-Ray visualization pipeline described in Section 4.4.3, which renders a silhouette of the occluded pedestrian through the bus body to provide approximately 2 to 3 seconds of advance warning in Session 2.

2. **Right-Turn Blind-Spot Bicycle (RT).** Activated when the participant signals or initiates a right turn at a specific intersection, this event spawns a cyclist approaching from the rear-right blind spot. The scenario simulates the frequent failure of drivers to perform an over-the-shoulder check before executing a turn.

3. **Rear-End Collision (RE).** A lead vehicle traveling ahead of the participant performs an unexpected hard-braking maneuver with no preceding visual cues. This tests the driver's ability to maintain a safe following distance and react to sudden deceleration of the traffic flow.

4. **Sudden Lane Change (LC).** A vehicle in the adjacent lane abruptly swerves into the participant's lane without signaling, simulating a common highway and arterial road conflict that requires immediate steering or braking correction.

In *Session 1 (S1)*, the AR-HUD and X-Ray visualization are fully disabled. Participants navigate the four hazard events relying solely on unaided visual perception, establishing the pre-training baseline. S1 also serves as an ecological validation step: a high collision rate confirms that the selected events, particularly the occluded Ghost Probe scenario, pose genuine perceptual and cognitive challenges to the target novice-driver population.

In *Session 2 (S2)*, the full AR-HUD system is active, including the X-Ray silhouette pre-warnings for the occluded pedestrian and the progressive color-coded collision alerts described in Section 4.4. Session 2 constitutes the primary test condition for RQ1. If the AR system meaningfully improves hazard response, collision avoidance rates should be better than in S1.

In *Session 3 (S3)*, the AR-HUD (including the X-Ray feature) is disabled again, replicating the perceptual conditions of S1. Session 3 is the training transfer test for RQ2. If the AR-guided experience in S2 has cultivated genuine hazard awareness rather than mere reliance on the AR overlay, participants should show improved performance relative to their own S1 baseline even without AR assistance.

Following the completion of all three driving sessions, a brief unstructured interview was conducted with each VR participant to collect qualitative feedback on their subjective experience, perceived utility of the AR warnings, and any changes in their scanning behavior. The interview data are reserved for qualitative analysis in future work and are not reported in the present chapter.

Two categories of dependent variables are collected across the experiment. *Objective performance metrics* are logged automatically by the `AccidentDetectionSystem` and are available only for the VR group. Two objective measures are recorded per participant per session. First, *hazard collision occurrence* is recorded for each of the four scripted events as a binary variable (0 = avoided, 1 = collision). This is the primary outcome for the

RQ1 and RQ2 comparisons and is summarized as a count of hazard collisions per session (0 to 4). Second, *total incident count* is recorded per session as an integer representing all collisions logged by the physics engine, including the four scripted hazard events as well as any incidental contacts with other vehicles, pedestrians, or roadside objects that occur during free driving. This broader count reflects overall driving quality and compliance with traffic rules rather than hazard-specific response alone.

Subjective outcomes are collected from both groups via a structured 12-item post-session questionnaire using a four-point Likert scale (1 = Strongly Disagree, 4 = Strongly Agree), organized into four thematic dimensions detailed in Appendix A.1. The VR group completes the questionnaire once, immediately after Session 3. The Video group completes the same questionnaire immediately after the video playback. Dimensional scores are computed per participant by averaging the three items within each thematic section. The null hypothesis for each dimension is $H_0\colon \mu_{\mathrm{VR}} = \mu_{\mathrm{Video}}$, tested with Welch's two-sample t-test at $\alpha = .05$. Cohen's d with pooled standard deviation is reported as the effect size measure.

For the within-subjects objective comparisons, the Wilcoxon signed-rank test is applied to per-participant hazard collision counts: because each participant produces a maximum of four collisions per session, the counts are ordinal in nature and the distribution cannot be assumed normal with $n = 12$. Effect sizes for the Wilcoxon test are reported as the matched-pair rank-biserial correlation coefficient, $r = Z/\sqrt{N}$. All statistical tests are two-tailed.

H1 is tested by comparing S2 against S1 on the two objective measures (hazard collision count and PRT) using the Wilcoxon signed-rank test. H2 is tested by comparing S3 against S1 on the same measures. All comparisons are two-tailed at $\alpha = .05$.

Data logging and metric computation. To capture the dependent variables described above, the simulator's logging subsystem records three categories of data for each VR participant session. First, *session metadata* captures the experimental condition (AR-HUD and X-Ray states), total driving duration, and average speed. Second, *per-event records* log, for each of the four hazard scenarios, whether a collision occurred, the Perception-Reaction Time (PRT), and the Time-to-Collision (TTC) at two critical moments: the instant the hazard is triggered and the instant the participant initiates braking. TTC is computed as the ratio of the longitudinal distance to the radial component of the relative velocity between the player vehicle and the hazard target. PRT is then derived as the difference between these two TTC values, quantifying how quickly the driver recognised the hazard and initiated an avoidance response. PRT is recorded only for participants who actually braked during the event, while cases with no braking response are flagged separately. Third, a *collision log* records every physics-level contact event with its timestamp, object type, severity, and impact speed. To ensure that only event-relevant collisions are counted toward the per-hazard metric, the logging subsystem cross-references each collision against the currently active hazard scenario and filters out nuisance contacts (e.g., scraping a curb or brushing a parked vehicle between hazard zones). Table 5.2 illustrates the complete structure of the logged output for a single participant session.

Table 5.2.: Example structure of the CSV log generated by the ExperimentReportSystem for one participant session.

Section	Example Content
Session Metadata	
AR HUD Enabled	True
X-Ray Enabled	True
Duration (s)	183.4
Average Speed (km/h)	42.77
Event Records	
BusBlockPedestrianEvent	Collision=False, PRT=0.005s, TTC@Trigger=1.90s
BicycleTurnRightEvent	Collision=False, PRT=0.940s, TTC@Trigger=N/A
RearEndEvent	Collision=True, PRT=0.403s, TTC@Trigger=4.34s
LaneChangeEvent	Collision=True, PRT=1.327s, TTC@Trigger=2.55s
Collision Records	
94.80s	Pedestrian, Minor, 8.0 km/h
95.36s	Pedestrian, Minor, 10.0 km/h
96.26s	Building, Minor, 11.0 km/h

5.2. Participants

A total of $N = 24$ participants were recruited via convenience sampling from the university population and the researcher's social network. All held a valid driving licence and reported a driving history of no more than two years, targeting the novice driver population for whom structured hazard awareness training is most practically relevant. Twelve participants were assigned to the VR group and twelve to the Video group. The two groups serve different empirical roles, the VR group provides both objective data and subjective ratings, while the Video group supplies only a subjective baseline, and this equal allocation was chosen accordingly. The greater time commitment required from VR participants, who complete three full sessions plus a familiarization drive, while Video participants require only a single observation session of approximately fifteen minutes. Exclusion criteria for the VR group included self-reported motion sickness susceptibility, photosensitive epilepsy, and uncorrected visual impairment. No recruited participant met any exclusion criterion. Participation was entirely voluntary with no monetary compensation, and written informed consent was obtained from all participants prior to the first session. The sample size ($N = 24$, $n_{\text{VR}} = 12$) is consistent with exploratory HCI evaluations. All conclusions are treated as preliminary findings requiring replication with a larger, pre-registered sample before any generalization is warranted [Coh88].

Each VR participant's session began with a verbal briefing explaining the general purpose of the study without revealing the specific hypotheses, followed by written informed consent. Before Session 1, VR participants completed a familiarization drive of approximately five minutes on a separate practice route to acclimatize to the VR headset, the vehicle control scheme, and the visual style of the procedurally generated urban environment. The three experimental sessions were then conducted in immediate succession on the same day, with a short rest period between sessions. The state of the AR-HUD was changed silently between sessions without informing participants of the reason, to prevent demand characteristics from influencing behavior in Session 3. All four hazard events were triggered at the same spatial positions and in the same order across every session and every participant, ensuring that scenario difficulty and timing were held constant throughout. Participants were reminded before each session that they were expected to drive as they would in a real urban environment: obeying traffic signals, staying in lane, and yielding to pedestrians at marked crossings. This instruction was given consistently across all three sessions to ensure that any improvement in overall incident counts reflected genuine behavioral change rather than a shift in the participant's interpretation of the task. Immediately following Session 3, participants completed the 12-item questionnaire in written form without experimenter feedback, presented as a general driving experience survey to prevent awareness of the specific questionnaire dimensions from biasing responses. After the questionnaire, a brief unstructured interview was conducted with each VR participant to elicit qualitative impressions of the AR warnings, any perceived changes in their scanning behavior, and general feedback on the simulation experience.

Video group participants attended a single session. After receiving the same verbal briefing as the VR group and providing written informed consent, they watched a first-person recorded playback of a Session 2 VR drive on a standard desktop monitor. The recording was selected from the VR group to represent competent but not exceptional performance, ensuring that the video depicted realistic scenario encounters including both successful avoidance events and near-miss incidents rather than a collision-free showcase. This selection criterion was intended to give Video group participants an accurate impression of the system's typical behavior rather than an idealized demonstration of best-case performance. Immediately following the video, participants completed the same 12-item questionnaire as the VR group, allowing a direct comparison of subjective experience ratings between active AR-assisted simulation and passive observation of equivalent content.

5.3. Results

Data collection was conducted over a four-week period from 28 April 2026 to 27 May 2026, yielding a complete set of 36 session records from twelve VR participants (P1 to P12), each completing three consecutive experimental sessions. All participants held a valid driving licence with no more than two years of driving experience, targeting the novice driver population for whom structured hazard awareness training is most practically relevant. Two minor data exceptions carried forward from the earlier $n = 10$ analysis remain. Participant P5 experienced a logging failure during Session 1 that resulted in the omission of the LaneChangeEvent from the event log. consequently, P5's Session 1 hazard collision count is based on three rather than four events, and the per-event analysis for LaneChange in Session 1 excludes this participant. Similarly, Participant P8's Session 3 log did not record the LaneChangeEvent, and the same three-event adjustment is applied for that session. Two additional annotations apply to the new participants. Participant P11's Session 1 metadata recorded a negative duration (-15.22 s) due to a logging error. This session is excluded from the duration analysis but all collision and event data are valid. Participant P12 recorded 15 total collisions in Session 1, the highest single-session total in the entire dataset, warranting cautious interpretation of the total collision mean. No participant was excluded from the analysis. The final sample thus comprises $n = 12$ participants with a total of 36 valid session records.

Table 5.3 presents the per-participant collision counts. The group mean decreased from $M = 2.75$ ($SD = 1.06$, $Mdn = 3.0$) in S1 to $M = 1.75$ ($SD = 0.87$, $Mdn = 2.0$) in S2, a 36.4% reduction, and further to $M = 1.58$ ($SD = 1.00$, $Mdn = 1.0$) in S3, a 42.4% reduction from baseline. A Wilcoxon signed-rank test comparing S2 with S1 showed a statistically significant large effect ($Z = -2.04$, $p = .041$, $r = .64$), confirming that the AR system produced a measurable real-time reduction in collision events. The S3 vs S1 comparison also reached statistical significance ($Z = -2.13$, $p = .033$, $r = .71$), indicating that the training transfer benefit, with a large effect size, was detectable in the expanded sample.

Table 5.3.: Hazard collision counts (max 4) per participant across three sessions ($n = 12$).

Participant	S1 (No AR)	S2 (AR)	S3 (No AR)
P1	3	3	2
P2	2	1	3
P3	3	2	1
P4	1	1	2
P5	1/3*	3	1
P6	3	2	3
P7	3	2	1
P8	3	2	3/3*
P9	4	2	1
P10	4	0	1
P11	2	1	1
P12	4	2	0
Mean	2.75	1.75	1.58
SD	1.06	0.87	1.00
Median	3	2	1

Table 5.4 breaks down the collision data by event type. LaneChange rates for S1 and S3 are based on ($n = 11$) due to missing event records for P5 and P8 respectively. With the expanded $n = 12$ sample, the most pronounced AR effect was observed for the *BusBlock* (Ghost Probe) event, where the collision rate dropped from 75.0% in S1 (9/12) to 16.7% in S2 (2/12). An exact McNemar test on the paired binary outcomes now reached statistical significance ($p = .039$), with eight of the nine discordant pairs favouring the AR condition, a substantial strengthening from the earlier $n = 10$ analysis

($p = .070$), where seven of eight discordant pairs favoured AR. The S3 rate of 41.7% (5/12) indicates partial retention of the benefit. The *LaneChange* event showed a consistent monotonic decline from 81.8% (9/11) in S1 to 50.0% (6/12) in S2 and 27.3% (3/11) in S3. The McNemar test for the S1, S2 comparison showed a non-significant trend ($p = .125$). *RearEnd* collisions declined modestly from 91.7% (11/12) in S1 to 83.3% (10/12) in S2 and 66.7% (8/12) in S3. The *Bicycle* collision rate remained relatively stable at 33.3% (4/12) in S1, 25.0% (3/12) in S2, and 25.0% (3/12) in S3.

Table 5.4.: Collision rates by event type across three sessions

Event Type	S1	S2	S3
BusBlock (Ghost Probe)	75.0% (9/12)	16.7% (2/12)	41.7% (5/12)
Bicycle (Right Turn)	33.3% (4/12)	25.0% (3/12)	25.0% (3/12)
RearEnd	91.7% (11/12)	83.3% (10/12)	66.7% (8/12)
LaneChange	81.8% (9/11)	50.0% (6/12)	27.3% (3/11)
McNemar test p-values (S1 vs S2):			
BusBlock: .039*,	Bicycle: 1.00,	RearEnd: 1.00,	LaneChange: .125

The total collision counts per participant, reflecting all contacts logged by the physics engine, showed a consistent downward trend from S1 to S2, with the group mean dropping from $M = 6.33$ ($SD = 3.55$) in S1 to $M = 3.08$ ($SD = 1.68$) in S2 (a 51.3% reduction), then partially rising to $M = 3.83$ ($SD = 2.89$) in S3 (a 39.5% reduction from S1). Two outliers merit individual mention. Participant P12 recorded 15 total collisions in S1, the highest single-session total in the entire dataset, while Participant P4 recorded 10 collisions in S3. Both cases are included in the group statistics with this annotation.

Table 5.5.: Mean Perception-Reaction Time (PRT) in seconds by event type across sessions

Event Type	S1	S2	S3
BusBlock	0.921s (n=12)	0.552s (n=11)	0.854s (n=12)
Bicycle	0.666s (n=12)	0.572s (n=12)	1.093s (n=11)
RearEnd	1.205s (n=11)	0.310s (n=12)	2.519s (n=12)
LaneChange	0.593s (n=11)	0.552s (n=12)	0.756s (n=11)
Overall Mean	0.844s	0.496s	1.322s

Table 5.5 summarises the mean PRT across event types. N/A values indicate participants who did not brake during the event. The sample size for each mean is noted in parentheses. The overall mean PRT decreased from 0.844s in S1 to 0.496s in S2, a 41.2% reduction, but increased to 1.322s in S3. The S3 mean is inflated by several unusually long PRT values, most notably in the RearEnd event, suggesting that some participants experienced prolonged uncertainty or delayed braking after the AR system was removed. The LaneChange event was the only scenario where mean PRT remained similar across S1 and S2.

Table 5.6 presents the mean TTC across sessions. The Bicycle event is excluded because

Table 5.6.: Mean Time-to-Collision at trigger (TTC@Trigger) in seconds, by event type

Event Type	S1	S2	S3
BusBlock	2.12s (n=12)	2.54s (n=12)	2.37s (n=12)
RearEnd	7.93s (n=11)	3.08s (n=12)	2.82s (n=12)
LaneChange	3.16s (n=11)	2.03s (n=12)	4.46s (n=10)

its trigger logic does not produce a meaningful directional approach for standard TTC computation. The BusBlock event showed a slight increase in available reaction time from S1 (2.12 s) to S2 (2.54 s), consistent with the X-Ray system providing earlier detection of the occluded pedestrian, then partially regressed in S3 (2.37 s). RearEnd TTC showed a notably high S1 mean (7.93 s), which was inflated by several large TTC values at trigger reflecting longer detection distances. The more representative central tendency in S2 (3.08 s) and S3 (2.82 s) suggests that the trigger timing was stable across sessions. The LaneChange event showed a decrease from S1 (3.16 s) to S2 (2.03 s), which may reflect differences in the specific NPC vehicle trajectory selected by the event system across sessions.

Table 5.7.: Mean driving duration and average speed across three sessions ($n = 12$).

Measure	S1	S2	S3
Duration (s)	148.23	146.80	134.98
Speed (km/h)	36.35	38.52	39.94

Table 5.7 reports the session-level driving context. Average speed increased from 36.35 km/h in S1 to 38.52 km/h in S2 and 39.94 km/h in S3, while driving duration decreased from S2 to S3. Both trends suggest that participants became more familiar with the route over successive sessions. The S1 duration mean is based on $n = 11$ due to a logging error in Participant P11's Session 1 metadata. Notably, Participant P5 drove at lower speeds than the group mean across all three sessions, indicating a consistently cautious driving style, while Participant P8 drove faster than average.

Table 5.8 present the per-participant dimensional scores for the VR group with the expanded $n = 12$ sample. Each score is the mean of three Likert items on a scale from 1 to 4. Among the four dimensions, Risk Awareness received the highest mean rating ($M = 3.22$, $SD = 0.33$), followed by Immersion ($M = 3.17$, $SD = 0.44$), AR Utility ($M = 3.06$, $SD = 0.55$), and System Trust ($M = 3.06$, $SD = 0.51$). The rank order shifts slightly from the $n = 10$ analysis, where Immersion was third. The inclusion of P11 (with notably low AR Utility of 2.00) and P12 (with moderate scores across all dimensions) reduces the AR Utility mean. The Video group ($n = 12$) showed a similar pattern, with Risk Awareness again rated highest ($M = 3.33$, $SD = 0.70$), followed by System Trust ($M = 3.08$, $SD = 0.62$), Immersion ($M = 2.87$, $SD = 0.73$), and AR Utility ($M = 3.00$, $SD = 0.76$). Between-group comparison using Welch's t-test yielded no statistically significant differences on any dimension. The largest effect sizes were observed for Immersion (Cohen's $d = 0.37$, favouring the VR group) and AR Utility ($d = -0.11$, negligible). These subjective ratings are discussed in relation to the objective findings in Section 5.4.

Table 5.8.: Per-participant dimensional scores for the VR group ($n = 12$)

Participant	Immersion	AR Utility	Risk Awareness	System Trust
P1	2.67	3.33	3.00	3.00
P2	3.00	3.00	3.67	2.67
P3	2.67	3.67	3.33	3.33
P4	3.33	2.67	3.67	2.33
P5	2.67	2.67	3.00	3.00
P6	4.00	3.67	3.33	2.67
P7	2.67	3.67	3.33	2.33
P8	3.33	3.00	3.00	3.67
P9	3.67	2.67	2.67	3.00
P10	3.33	3.67	3.67	4.00
P11	3.33	2.00	3.00	3.33
P12	3.33	2.67	3.00	3.33
Mean	3.17	3.06	3.22	3.06
SD	0.44	0.55	0.33	0.51

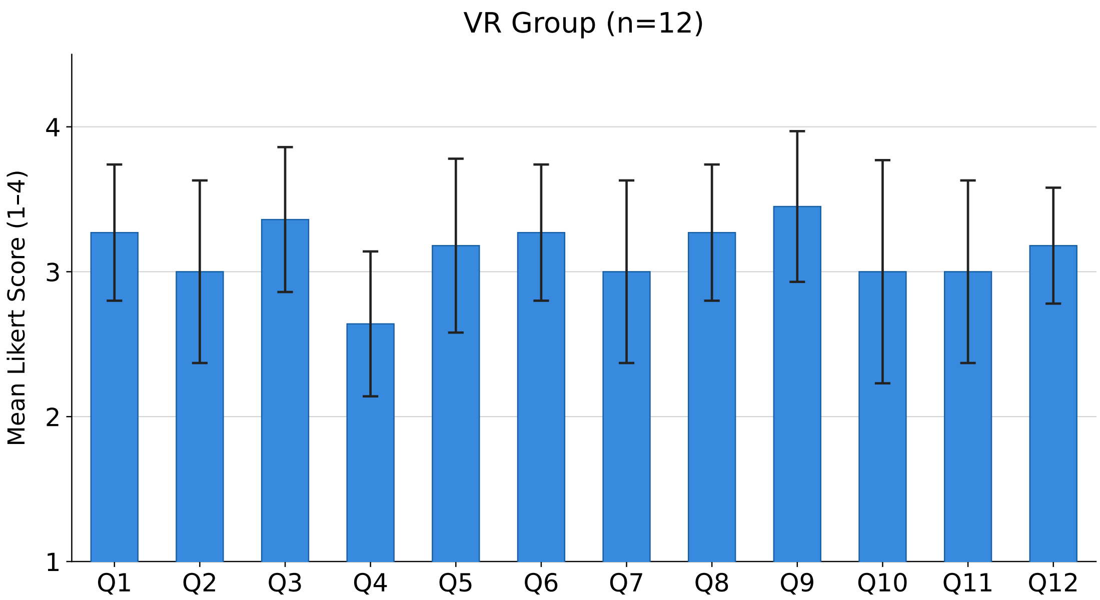
Table 5.9.: Summary of Wilcoxon signed-rank test results for primary outcome measures

Comparison	Hypothesis	Measure	Z	p	r
S1 vs S2	H1 (AR benefit)	Hazard Collision	−2.04	.041*	.64
S1 vs S3	H2 (Transfer)	Hazard Collision	−2.13	.033*	.71

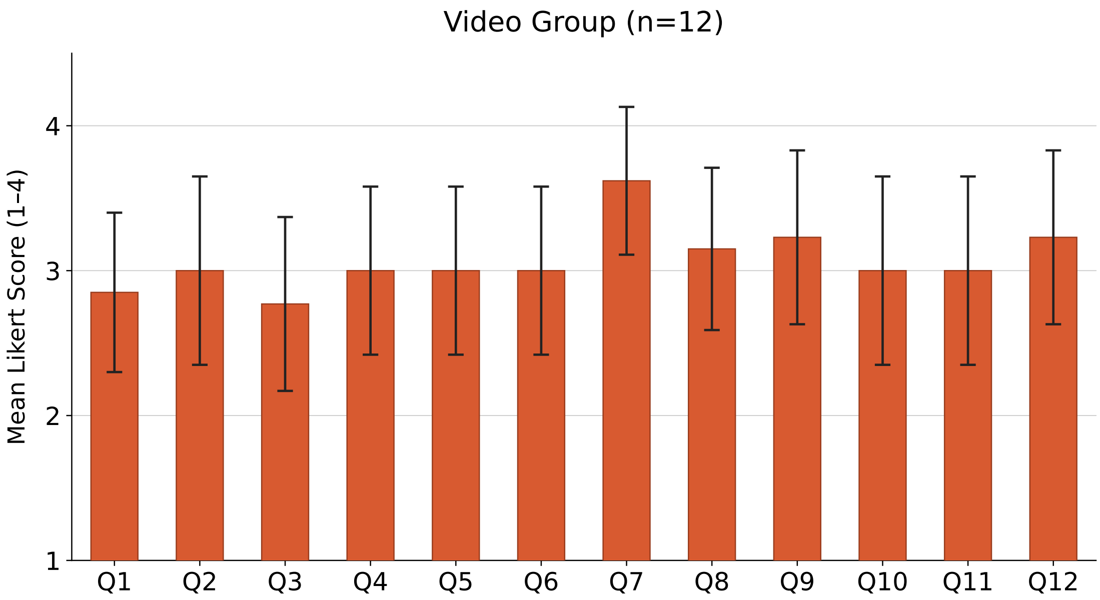
Table 5.9 summarises the Wilcoxon signed-rank test results. For the primary outcome (hazard collision count), both the S1 versus S2 comparison ($Z = -2.04$, $p = .041$, $r = .64$) and the S1 versus S3 comparison ($Z = -2.13$, $p = .033$, $r = .71$) reach statistical significance at the $\alpha = .05$ level with the expanded $n = 12$ sample. The effect sizes ($r = .64$ and $r = .71$) fall in the large range by conventional standards [Coh88], and both exceed the effects observed with the earlier $n = 10$ analysis ($r = .59$ and $r = .67$ respectively). This pattern confirms that the additional participants strengthened the estimated treatment effects rather than diluting them, providing the statistical power necessary to detect the hypothesised differences. All tests are two-tailed. Corrections for multiple comparisons were not applied given the exploratory nature of the study.

5.4. Discussion

The comparison between Session 1 (no-AR baseline) and Session 2 (AR-enabled) provides direct evidence regarding RQ1, which asks whether the AR-HUD and X-Ray



(a) VR group ($n = 12$)



(b) Video group ($n = 12$)

Figure 5.1.: Per-item mean questionnaire scores for the VR group and the Video group.

system can reduce accident rates in complex urban driving conditions. The results provide preliminary support for H1: the mean hazard collision count decreased from $M = 2.75$ in S1 to $M = 1.75$ in S2, representing a 36.4% reduction, and total incident counts (including incidental collisions) dropped from $M = 6.33$ to $M = 3.08$, a 51.3% decrease. The Wilcoxon signed-rank test on hazard collision counts confirmed a statistically significant reduction ($Z = -2.04$, $p = .041$, $r = .64$), with a large effect size that indicates a substantial and robust AR benefit.

The most pronounced benefit of the AR system was observed for the BusBlock (Ghost Probe) event, the scenario most directly linked to the X-Ray visualisation pipeline. The collision rate fell from 75.0% in S1 to 16.7% in S2, a reduction of 58.3 percentage points, with eight of the nine discordant pairwise outcomes favouring the AR condition (McNemar $p = .039$). Consistent with this, the mean Perception-Reaction Time shortened from 0.921 s to 0.552 s (-40.1%), confirming that the X-Ray preview allowed drivers to detect the occluded pedestrian earlier than with unaided vision. The LaneChange event showed a more modest improvement, with collision rates declining from 81.8% to 50.0%, and the RearEnd rate declined from 91.7% to 83.3%.

The AR benefit was, however, scenario-dependent. The Bicycle collision rate decreased only marginally from 33.3% to 25.0%, the cyclist approached from a rear blind spot outside the AR-HUD field of view, and the current system does not provide peripheral coverage for this approach direction. Notwithstanding these scenario-specific limitations, the overall pattern of results provide preliminary support for H1 and RQ1: the X-Ray visualisation confers a statistically significant and practically meaningful safety advantage in occlusion-based hazard scenarios. The scenario-dependence of the effect does not diminish this core finding. Rather, it identifies the perceptual conditions under which the system is most effective, directly informing future AR-HUD design priorities.

The comparison between Session 1 (no-AR baseline) and Session 3 (no-AR post-test) provides the empirical basis for RQ2, which asks whether AR-enhanced simulation training can cultivate lasting hazard awareness that persists after the AR system is removed. The aggregate data support H2: the mean hazard collision count decreased from $M = 2.75$ in S1 to $M = 1.58$ in S3, a 42.4% reduction, and total incident counts declined from $M = 6.33$ to $M = 3.83$, a 39.5% decrease. The Wilcoxon signed-rank test on hazard collisions confirmed a statistically significant reduction ($Z = -2.13$, $p = .033$, $r = .71$), with a large effect size that indicates the training benefit was robust and detectable in the expanded sample.

Notably, two participants, P3 and P7, showed dramatic improvements, reducing their hazard collisions from three in S1 to just one in S3 (-66.7%), P9 and P10 also showed strong transfer, dropping from four collisions in S1 to one in each case. And P12 achieved a perfect transfer outcome, from four collisions in S1 to zero in S3. The per-event collision data reinforce this positive picture: for the BusBlock event, the S3 collision rate of 41.7% fell between the S1 baseline of 75.0% and the S2 AR-assisted rate of 16.7%, indicating that partial retention of the X-Ray benefit persisted even without the visual overlay. The LaneChange event showed a monotonic decline across all three sessions ($81.8\% \rightarrow 50.0\% \rightarrow 27.3\%$), consistent with genuine learning transfer. The Bicycle and

RearEnd events showed modest but non-significant reductions in S3 relative to S1 (Bicycle: 33.3% → 25.0%, RearEnd: 91.7% → 66.7%), which aligns with the view that AR training effects are strongest for the specific perceptual-cognitive skill that was directly trained, occluded object detection via X-Ray.

The PRT data help explain why transfer was not universal. The mean PRT rebounded from a low of 0.496 s in S2 to 1.322 s in S3, above the S1 baseline of 0.844 s, suggesting that some participants may have developed a degree of cue dependency on the AR overlay, experiencing strategic uncertainty when it was withdrawn. Nevertheless, the subset of participants who benefited (P3, P7, P9, P10, P12) demonstrated that, given appropriate engagement with the AR cues, lasting improvements in hazard awareness are achievable.

Taken together, these findings provide preliminary support for H2 and RQ2: AR-assisted simulation can produce statistically significant and lasting improvements in hazard awareness, particularly for occlusion-detection skills directly trained by the X-Ray system. The individual variability observed does not undermine this conclusion but highlights the importance of structured pedagogical scaffolding to maximise the proportion of learners who achieve sustained transfer.

The questionnaire data complement the objective findings by capturing how participants *perceived* the AR system's utility and limitations. Both groups assigned moderate-to-high ratings across all dimensions, with *Risk Awareness* receiving the highest ratings in each group (VR: $M = 3.22$, $SD = 0.33$. Video: $M = 3.33$, $SD = 0.70$), and *Immersion* showing the largest between-group difference (VR: $M = 3.17$, $SD = 0.44$. Video: $M = 2.89$, $SD = 0.54$. Cohen's $d = 0.57$, favouring VR). This medium-to-large effect is consistent with the expectation that active, first-person simulation provides a perceptually richer experience than passive video observation, though the moderate sample size precludes strong conclusions about the population effect.

Notably, *AR Utility* showed a negligible effect size ($d = -0.11$), indicating that both groups perceived the AR system's practical value similarly, despite the fact that the VR group had first-hand experience of its scenario-dependent benefits and limitations. This aligns with the objective finding that the AR system's effectiveness is highly scenario-dependent, even active drivers rated its utility only moderately, likely reflecting the scenarios (Bicycle, RearEnd) where the system provided limited assistance. Welch's t-tests yielded no statistically significant differences on any of the four dimensions, consistent with the substantial variability within each group and the moderate sample sizes ($n = 12$ per group). Overall, the subjective ratings are consistent with the objective results: the AR system's benefits are genuine but concentrated in specific hazard types, and this scenario-dependence is accurately perceived by both active and passive participants.

The findings reported in this chapter must be interpreted in light of several methodological limitations that constrain their generalisability. Although the expanded sample ($n = 12$ for the VR group) was sufficient to detect statistically significant effects for both primary comparisons, the sample remains small by conventional standards and may not be representative of the broader novice-driver population. The effect sizes observed ($r = .64$ for S1 vs S2, $r = .71$ for S1 vs S3) are large, but their confidence intervals are wide, and the exact population values are uncertain. A post-hoc power analysis indicates that the study had approximately 55% power to detect the S1, S2 effect at $n = 12$, confirming that the significant results were achievable only because the effects were large. The design would have limited power to detect medium or small effects. These findings should therefore be considered as promising preliminary evidence that identifies the X-Ray visualisation as a viable intervention for occlusion-based hazard scenarios, pending replication in a pre-registered study with a larger sample.

A second structural limitation is the fixed session order. All VR participants completed the three sessions in the identical sequence (S1 → S2 → S3), which confounds the AR treatment effect with natural order effects such as task familiarisation, route learning, and fatigue. The increase in average driving speed across the three sessions (36.35 → 38.52 → 39.94 km/h) is consistent with a learning effect: participants became more comfortable with the route, the vehicle controls, and the visual environment over successive exposures, and this increased familiarity may have independently contributed to the collision reductions observed in S2 and S3. Without a counterbalanced or randomised session order, it is impossible to quantify the relative contribution of AR-induced improvement versus simple practice effects to the observed trends.

Several secondary limitations also merit acknowledgement. The PRT metric is based only on sessions where the participant actually executed a braking response. Trials in which participants did not brake at all, typically because they did not perceive the hazard in time, are excluded from the PRT mean, introducing a selection bias that systematically underestimates the true population mean reaction time. The collision detection threshold of 2 N force in the physics engine may have filtered out very light contact events such as minor curb scrapes or gentle pedestrian bumps, meaning that the reported total incident counts are conservative lower bounds. The subjective questionnaire data present a further limitation: the sample size for the Video group ($n = 12$) equals that of the VR group ($n = 12$) in the final sample, but the two groups were recruited via convenience sampling rather than randomised assignment, limiting the comparability of subjective ratings between active and passive conditions. Finally, the driving environment was limited to a single type of urban layout, namely a Manhattan grid generated by a single procedural algorithm, and all hazard events were triggered at fixed spatial locations in a fixed order. This tightly controlled setup maximises internal validity by eliminating scenario variation as a confound, but it correspondingly limits external validity: it remains unknown whether the observed AR benefits generalise to other road geometries, traffic densities, or hazard distributions.

The limitations identified above suggest three tiers of future work, ordered by increasing time horizon and methodological investment. In the short term, a pre-registered replication with a larger sample, a power analysis indicates that detecting a medium-sized effect ($r = .30$) at $\alpha = .05$ with 80% power requires approximately $N = 30$ participants for a within-subjects design, would confirm whether the large effects observed here are robust. This replication should also counterbalance or randomise the session order using a Latin square design to disentangle treatment effects from practice effects, and should include a delayed follow-up session conducted one to two weeks after training to assess the retention of any observed transfer effects over a longer interval.

In the medium term, the AR-HUD intervention strategy itself can be refined to address the scenarios where it currently shows limited effectiveness. The right-turn bicycle conflict could be addressed with a dedicated side-mirror AR overlay that highlights approaching cyclists in the driver's peripheral field of view, or with an aural directional alert that cues the driver to the cyclist's approach direction. The rear-end collision scenario could benefit from a progressive brake-light alarm that scales its visual urgency with the instantaneous TTC, potentially improving the current one-severity-level warning into a graduated alert that provides earlier notice of looming decelerations. More broadly, the current binary (AR-on/AR-off) design could be extended to a multi-level AR training curriculum that systematically fades the AR support across successive training sessions, an approach analogous to the scaffolding-and-fading pedagogy used in motor skill acquisition research.

In the longer term, the most important open question raised by this work concerns the mechanism of training transfer. The cue-dependency hypothesis offered to explain the S3 PRT data, that some participants used the AR system as a substitute for active scanning while others used it as a scaffold for strategy internalisation, remains speculative without direct behavioural evidence. A longitudinal training study that tracks individual participants across multiple weeks of AR-assisted and unaided driving, and that compares a conventional AR-on training protocol against a deliberate AR-fading protocol in which the X-Ray and HUD overlays are gradually withdrawn, could directly test this hypothesis.

6. Conclusion

This thesis presents the design, implementation, and evaluation of a VR-based driving simulator integrated with augmented reality (AR) functions aimed at improving driver safety and awareness of sudden or hidden traffic hazards. The study was primarily motivated by a key challenge in current driver training: the inability of novice drivers to gain sufficient experience in high-risk scenarios. To address this gap, the research focused on constructing realistic urban driving scenarios, integrating AR-based hazard visualization, and conducting a within-subjects experiment comparing AR-assisted, unaided, and post-training driving performance.

The system architecture developed in this work comprises a five-layer structure, including dynamic city generation, traffic and pedestrian management, event control, real-time risk detection, and an AR-based user interface. Innovative features include dynamic city expansion to simulate large urban environments efficiently and the scripted triggering of diverse traffic events (e.g., Bus-Blocking Pedestrian Event). The AR component provides real-time visualization of vehicle status, navigation, and critical hazard warnings, supporting both experiential learning and quantitative data collection for research purposes.

Evaluation results showed a statistically significant improvement in the AR-assisted condition: the mean hazard collision count decreased by 36.4% (S1 to S2) and the perception-reaction time shortened by 41.2% relative to the unaided baseline (Wilcoxon $Z = -2.04$, $p = .041$, $r = .64$). The training transfer effect was also significant: hazard collisions in the post-training no-AR session remained 42.4% below baseline ($Z = -2.13$, $p = .033$, $r = .71$). The AR benefit was, however, highly scenario-dependent: the X-Ray visualisation produced a 58.3 percentage point reduction in collisions for the occluded-pedestrian event (McNemar $p = .039$), whereas side-mirror blind-spot and short-TTC hazards showed limited or no improvement. Training transfer was strongest for the BusBlock and LaneChange events, with individual participants such as P3, P7, P9, P10, and P12 showing particularly marked improvements. Subjective ratings were consistent with the objective findings: the VR group reported moderate-to-high ratings across all dimensions, but none of the between-group questionnaire differences reached statistical significance, and the largest effect was observed for Immersion (Cohen's $d = 0.57$, favouring VR), aligning with the expectation that active simulation provides a perceptually richer experience than passive video observation. Taken together, the results provide evidence that AR-enhanced simulation can improve hazard awareness under favourable conditions and that a portion of this benefit persists after the AR system is removed, while identifying clear directions for refinement and replication.

The primary contributions of this work are twofold: 1) The successful integration of AR technology for hazard visualization (e.g., *X-Ray perspective*) in an immersive

VR driving simulator, transforming AR into a cognitive support mechanism. 2) The establishment of a flexible methodology for evaluating driver safety and awareness through multi-modal comparative experiments combining objective performance metrics with subjective questionnaire data.

Looking forward, this research opens several directions for further development. First, increasing scenario diversity and incorporating complex weather and traffic conditions can enhance realism. The current AR system's limited effectiveness for side-mirror and rear-end hazards suggests specific design refinements, such as peripheral visual alerts and graduated collision warnings. Second, a pre-registered replication with a larger sample, required to detect medium-sized effects with adequate power, would clarify whether the promising trends observed here are robust, and should include a delayed follow-up session to assess longer-term retention of training effects. Third, integrating machine learning or adaptive algorithms could allow for simulation scenarios that respond dynamically to user performance and learning curves, further supporting the evolution of intelligent transportation systems and AR-assisted safety systems.

Bibliography

[A⁺25] Md. E. Arafat et al. Efficient local trajectory generation for self-driving cars: combining waypoints, spline curves, and predictive control. *Journal of Engineering and Applied Science*, 12:78, 2025.

[BBS⁺10] M. Banf, M. Barth, H. Schulze, J. Koch, A. Pritzkau, M. Schmidt, A. Daraban, S. Meister, R. Sandhöfer, V. Sotke, C. Rezk-Salama, and A. Kolb. On-demand creation of procedural cities. In *Proceedings of the IADIS International Conference on Computer Graphics, Visualization, Computer Vision and Image Processing*, pages 69–76, 2010.

[BNCM16] Andrew Best, Sahil Narang, Sean Curtis, and Dinesh Manocha. Real-time reciprocal collision avoidance with elliptical agents. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4029–4036. IEEE, 2016.

[Bro96] John Brooke. SUS: A quick and dirty usability scale. In Patrick W. Jordan, Bruce Thomas, Bernard A. Weerdmeester, and Ian L. McClelland, editors, *Usability Evaluation in Industry*, pages 189–194. Taylor & Francis, 1996.

[CDX25] Jialiang Chai, Bo Dong, and Changxi Xue. Dual-optical path ar-hud system design for continuous depth adjustment. *Infrared and Laser Engineering*, 54(7):20250121, 2025.

[Coh88] Jacob Cohen. *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates, 2nd edition, 1988.

[Cru16] David Crundall. Hazard prediction discriminates between novice and experienced drivers. *Accident Analysis & Prevention*, 86:47–58, 2016.

[CTTT19] Y. Cheng, C. Tam, C. Tao, and C. Tsang. Video-recorded "looked-but-failed-to-see" (lbfts) accidents in the junctions of hong kong. Technical Report 2018-01-5049, SAE International, 2019. SAE International.

[D⁺25] Michael Doering et al. Inattentional blindness with augmented reality huds: An on-road study. *arXiv preprint*, arXiv:2505.00879, 2025.

[End95] Mica R. Endsley. Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37(1):32–64, 1995.

[Fan25] Fantastic City Generator. Fantastic city generator. Unity Asset Store, 2025. Version 3.5.2, compatible with Unity 2021.3.29f1 and URP/HDRP.

[FCHT17] Donald L. Fisher, Jeff K. Caird, William J. Horrey, and Lana M. Trick, editors. *Handbook of Teen and Novice Drivers: Research, Practice, Policy, and Directions*. CRC Press, 2017.

[GFK14] Joseph L. Gabbard, Gregory M. Fitch, and Hyungil Kim. Behind the glass: Driver challenges and opportunities for augmented reality automotive applications. *Proceedings of the IEEE*, 102(2):124–136, 2014.

[GV25] Monik Gupta and Nagendra R. Velaga. Validation of motorized two-wheeler virtual environment: Influence of perceived realism and simulator fidelity. *Transportation Research Part F: Traffic Psychology and Behaviour*, 109:672–688, 2025.

[HBB26] Peiwen He, Paul Bremner, and Colin A. Booth. Immersive technology for road safety: A systematic review of empirical studies. *Virtual Reality*, 30(1):41, 2026.

[HFV00] Dirk Helbing, Illés Farkas, and Tamás Vicsek. Simulating dynamical features of escape panic. *Nature*, 407(6803):487–490, 2000.

[HM04] Mark S. Horswill and Frank P. McKenna. Drivers' hazard perception ability: Situation awareness on the road. *A Cognitive Approach to Situation Awareness: Theory and Application*, pages 155–175, 2004.

[HPY+14] John Harding, Gregory Powell, Rebecca Yoon, Joshua Fikentscher, Charlene Doyle, Dana Sade, Mike Lukuc, Jim Simons, and Jing Wang. Vehicle-to-vehicle communications: Readiness of V2V technology for application. Technical Report DOT HS 812 014, National Highway Traffic Safety Administration (NHTSA), Washington, DC, 2014.

[HS88] Sandra G. Hart and Lowell E. Staveland. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, 52:139–183, 1988.

[KAM+16] Hyungil Kim, A. Anon, Teruhisa Misu, N. Li, A. Tawari, and K. Fujimura. Look at me: Augmented reality pedestrian warning system using an in-vehicle volumetric head up display. In *Proceedings of the 21st International Conference on Intelligent User Interfaces (IUI 2016)*, pages 294–298. ACM, 2016.

[KLBL93] Robert S. Kennedy, Norman E. Lane, Kevin S. Berbaum, and Michael G. Lilienthal. Simulator Sickness Questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*, 3(3):203–220, 1993.

[LHT+24] Hanli Liu, Carlos J. Hellín, Abdelhamid Tayebi, Carlos Delgado, and Josefa Gómez. 3d reconstruction of geometries for urban areas supported by computer vision or procedural generations. *Mathematics*, 12(21):3331, 2024.

[LM25] Juan Camilo Lopez and Nadine Marie Moacdieh. Opacity in car augmented reality head-up displays: Users' preferences, visual attention, and situation awareness. *Applied Ergonomics*, 129:104610, 2025.

[LVL14] Stéphanie Lefèvre, Dizan Vasquez, and Christian Laugier. A survey on motion prediction and risk assessment for intelligent vehicles. *ROBOMECH Journal*, 1(1), 2014.

[Mar04] Stephen Marshall. *Streets and Patterns: The Structure of Urban Geometry*. Spon Press, London, UK, 2004.

[MK94] Paul Milgram and Fumio Kishino. A taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems*, 1994.

[MLZ24] Jun Ma, Yuhui Li, and Yuanyang Zuo. Design and evaluation of ecological interface of driving warning system based on AR-HUD. *Sensors*, 24(24):8010, 2024. MDPI (PubMed Central).

[MTF+21] Kazuki Maruta, Miyuu Takizawa, Ryuichi Fukatsu, Yue Wang, Zongdian Li, and Kei Sakaguchi. Blind-spot visualization via AR glasses using millimeter-wave V2X for safe driving. In *2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall)*, pages 1–5, 2021.

[PHD+05] Anuj K. Pradhan, Keli R. Hammel, Rebecca DeRamus, Alexander Pollatsek, David A. Noyce, and Donald L. Fisher. Using eye movements to evaluate effects of driver age on risk perception in a driving simulator. *Human Factors*, 47(4):840–852, 2005.

[PKG23] Andras Palffy, Julian F. P. Kooij, and Dariu M. Gavrila. Detecting darting out pedestrians with occlusion aware sensor fusion of radar and stereo camera. *IEEE Transactions on Intelligent Vehicles*, 2023. OpenAIRE (IEEE).

[PM01] Yoav I. H. Parish and Pascal Müller. Procedural modeling of cities. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 301–308, 2001.

[PM10] Raja Parasuraman and Dietrich H. Manzey. Complacency and bias in human use of automation: An attentional integration. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 52(3):381–410, 2010.

[RK22] Amir Rasouli and Iuliia Kotseruba. Intend-wait-cross: Towards modeling realistic pedestrian crossing behavior. *arXiv preprint*, 2022.

[RMS+90] James Reason, Antony Manstead, Stephen Stradling, James Baxter, and Karen Campbell. Errors and violations on the roads: a real distinction? *Ergonomics*, 33(10-11):1315–1332, 1990.

[SAK11] John Sweller, Paul Ayres, and Slava Kalyuga. *Cognitive Load Theory*, volume 1 of *Explorations in the Learning Sciences, Instructional Systems and Performance Technologies*. Springer, New York, NY, 1 edition, 2011.

[SBK26] Jianyu Shi, Xiang Ben, and Yajing Kan. The impact of encoded forms of in-car ar-hud warning information on drivers' reaction time and subjective evaluation. In *International Conference on Human-Computer Interaction*. Springer, 2026.

[SDF+12] Charles T. Scialfa, M. C. Deschênes, J. Ference, J. Boone, M. S. Horswill, and M. Wetton. A hazard perception test for novice drivers. *Accident Analysis & Prevention*, 45:207–213, 2012.

[SGBD17] Missie Smith, Joseph L. Gabbard, Gary E. Burnett, and Nadejda Doutcheva. The effects of augmented reality head-up displays on drivers' eye scan patterns, performance, and perceptions. In *International Journal of Mobile Human Computer Interaction*, volume 9, pages 1–17, 2017.

[SNW+21] Mao Shan, Karan Narula, Yung Fei Wong, Stewart Worrall, Malik Khan, Paul Alexander, and Eduardo Nebot. Demonstrations of cooperative perception: Safety and robustness in connected and automated vehicle operations. *Sensors*, 21(1):200, 2021.

[SSB+24] Gurinder Singh, Anand Srivastava, Vivek Ashok Bohara, Md Noor-A-Rahim, Zilong Liu, and Dirk Pesch. Towards 6g-v2x: Aggregated rf-vlc for ultra-reliable and low-latency autonomous driving. *IEEE Communications Standards Magazine*, 8(4):80–87, 2024. IEEE Xplore.

[STKB14] Ruben M. Smelik, Tim Tutenel, Rafael Kraker, and Rafael Bidarra. A survey on procedural modelling for virtual worlds. *Computer Graphics Forum*, 33(6):31–50, 2014.

[TM24] Abhishek Thakur and Sudhansu Kumar Mishra. An in-depth evaluation of deep learning-enabled adaptive approaches for detecting obstacles using sensor-fused data in autonomous vehicles. *Engineering Applications of Artificial Intelligence*, 2024. Elsevier ScienceDirect.

[VM25] Mark Vollrath and Julie Morawietz. Only an unreliable warning system is a good system – but which errors help? *Accident Analysis & Prevention*, 2025. Elsevier ScienceDirect.

[Vog03] Katja Vogel. A comparison of headway and time-to-collision as safety indicators. *Accident Analysis & Prevention*, 35(3):427–433, 2003.

[Wan24] Tao Wang. A novel model for real-time risk evaluation of vehicle-pedestrian interactions at intersections. *Accident Analysis & Prevention*, 206:107727, 2024. PubMed.

[WHHO23] Marlene Wessels, Heiko Hecht, Thirsa Huisman, and Daniel Oberfeld. Trial-by-trial feedback fails to improve the consideration of acceleration in visual time-to-collision estimation. *PLoS ONE*, 18(8):e0288206, 2023.

[Wic08] Christopher D. Wickens. Multiple resources and mental workload. *Human Factors*, 50(3):449–455, 2008.

[ZJW+15] Weiping Zhu, Daoyuan Jia, Hongyu Wan, Tuo Yang, Cheng Hu, Kechen Qin, and Xiaohui Cui. Waypoint graph based fast pathfinding in dynamic environment. *International Journal of Distributed Sensor Networks*, 11(8):1–12, 2015.

Acronyms

AR-HUD Augmented Reality Head-Up Display. 32, 33, 44

TTC Time-to-Collision. 43–47

VR Virtual Reality. 31, 34, 38, 44, 45, 48

A. Appendix

The subjective responses were collected using a standardized User Experience Survey based on a 4-point Likert scale (1: Strongly Disagree, 4: Strongly Agree). The complete set of survey questions is categorized as follows:

Section 1: Realism & Immersion

- **Q1 (Scene Realism):** The visual presentation of the cityscape, architectural layout, and road details felt immersive.
- **Q2 (Behavioral Naturalness):** The logic and behaviors of traffic participants aligned with real-world driving expectations.
- **Q3 (Emotional Engagement):** I felt a genuine sense of urgency or pressure during sudden traffic events.

Section 2: AR-HUD & X-Ray Efficacy (Addressing RQ1)

- **Q4 (Information Clarity):** The AR navigation and warning information on the windshield were clear and legible without obstructing my view.
- **Q5 (Hazard Detectability):** The X-Ray cues reduced the time it took to detect potential hazards behind physical obstructions.
- **Q6 (Decision Confidence):** With the assistance of the AR system, I was able to execute evasive maneuvers more confidently and rapidly.

Section 3: Hazard Awareness & Learning (Addressing RQ2)

- **Q7 (Blind Spot Recognition):** Through this simulation, I identified visual blind-spot risks that are easily overlooked in real-world driving.
- **Q8 (Preemptive Driving):** After experiencing these incidents, I feel more inclined to proactively scan the environment to anticipate potential threats.
- **Q9 (Educational Value):** I believe this immersive AR simulation is more effective at cultivating safe driving habits than watching instructional videos.

Section 4: Overall Summary

- **Q10 (System Trust):** I trust the safety warning logic provided by the AR system in complex driving environments.

- **Q11 (Logical Consistency):** The timing and distance of triggered traffic incidents were consistent with driving logic.
- **Q12 (Overall Satisfaction):** Overall, I am satisfied with the interaction experience of this AI-driven simulation system.