

Do SAE levels reflect real-world capability? A supplemental environmental-situational-capability taxonomy for L3–L4 automated driving

Chengxi Hu¹, Sikai Chen^{2, ✉}, Samuel Labi³, Hongliang Ding⁴, Hyungchul Chung⁵, Tiantian Chen^{1, ✉}

Cite this article: <https://doi.org/10.26599/JICV.2026.9210100>

ABSTRACT: The widely adopted SAE (Society of Automotive Engineers) taxonomy categorizes driving automation primarily based on the division of operational responsibilities between the human driver and the system. This has worked well for several applications. However, it fails to distinguish the requisite versatility of autonomous vehicle (AV) performance across diverse operational environments. This structural limitation has spawned critical challenges, including safety risks, commercialization barriers, regulation challenges, and inadequate Operational Design Domain (ODD) development guidelines toward Level 5 autonomy. To address these gaps, this paper proposes a supplemental taxonomy for SAE Levels 3 and 4 using an Environmental-Situational Capability (ESC) evaluation framework across three core dimensions: the built environment, natural environment, and ambient traffic conditions. Using a two-stage hierarchical classification strategy, the ESC framework maps expert evaluations of a specific automated driving system (ADS) to corresponding levels of the proposed taxonomy. This novel classification bridges the ADS level and the ODD by correlating environmental complexity with the minimum capabilities required for safe autonomous operation. Operating orthogonally to SAE standards, this taxonomy categorizes systems based on their actual operational capabilities rather than driving automation task responsibility alone. Consequently, it establishes a crucial alignment across researchers, manufacturers, regulators, and consumers, providing a practical reference for product benchmarking, safety validation, and policymaking.

KEYWORDS: Autonomous vehicles, Automation level, Taxonomy, AV (Autonomous vehicle) situational capability, Operational design domain

1 Introduction

1.1. Structural gap of existing SAE taxonomy

Evidence from the literature suggests that autonomous vehicles (AVs) can significantly mitigate driver workload, eliminate human-error-induced accidents, enhance travel efficiency, and reduce emissions (Koopman and Wagner, 2017; H. Wang et al., 2019; Matin and Dia, 2022). However, despite their immense transformative potential in revolutionizing future mobility, the current technological maturity of AVs remains insufficient to support large-scale commercialization. Public acceptance of this technology continues to face formidable challenges, predominantly contingent upon demonstrable traffic safety and accountability transparency (Z. Xu et al., 2021; Naiseh et al., 2025). This is further compounded by the fact that existing legal, ethical, and regulatory frameworks persistently lag behind technological advancements (Taeihagh and Lim, 2019; Alqahtani, 2025).

In response to these complexities, the Society of Automotive Engineers (SAE) International, in collaboration with the International Organization for Standardization (ISO), developed a standardized framework to define automated driving system (ADS) (SAE, 2014; SAE, 2021). The SAE taxonomy classifies ADS technologies into six progressive levels, from Level 0 (L0) to Level 5 (L5) (SAE, 2021). As illustrated in Fig. 1, the primary distinctions among these levels depend on the allocation of the Dynamic Driving Task (DDT)—specifically lateral/longitudinal control (Steering, Lateral and Longitudinal Vehicle

Motion Control, SLLVMC) and Object and Event Detection and Response (OEDR)—as well as fallback responsibilities and Operational Design Domain (ODD) constraints. Level 1 (L1) and Level 2 (L2) function as driver support features (DSF), where the system manages partial vehicle control but the driver remains responsible for OEDR and fallback. The transition to ADS occurs at Level 3 (L3) where the system executes the entire DDT within a restricted ODD, though it still requires a human driver for fallback operations. The reliance on human intervention dictates that L3 operation still requires the driver to maintain the capacity for timely situational awareness recovery (H. Chen et al., 2024; Agrawal and Peeta, 2021). Higher automation levels fundamentally shift this task allocation: Level 4 (L4) systems independently handle all DDT fallbacks within a constrained ODD, while Level 5 (L5) represents unrestricted, fully autonomous operation (SAE, 2021).

The SAE taxonomy has been widely adopted following its endorsement by the National Highway Traffic Safety Administration (NHTSA, 2016). However, it still exhibits significant ambiguities and mismatches between ADS classification and their real-world capabilities, evaluations, and deployment (Hopkins and Schwanen, 2021; S. Kim et al., 2026). Fundamentally, the current SAE framework is a classification restricted solely to degrees of driving automation.

¹ Cho Chun Shik Graduate School of Mobility, Korea Advanced Institute of Science and Technology, Daejeon 34051, Republic of Korea. ² Department of Civil and Environmental Engineering, University of Wisconsin–Madison, Madison 53706, USA. ³ Lyles School of Civil and Construction Engineering, Purdue University, West Lafayette 47907, USA. ⁴ Institute of Smart City and Intelligent Transportation, Southwest Jiaotong University, Chengdu 611756, China. ⁵ Department of Urban Planning and Design, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China.

✉ Corresponding authors. E-mail: S. Chen, sikai.chen@wisc.edu; H. Ding, nicole.chen@kaist.ac.kr

Received: May 14, 2026; Revised: August 7, 2026; Accepted: September 14, 2026

© The Author(s) 2026. This is an open access article under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0, <http://creativecommons.org/licenses/by/4.0/>).

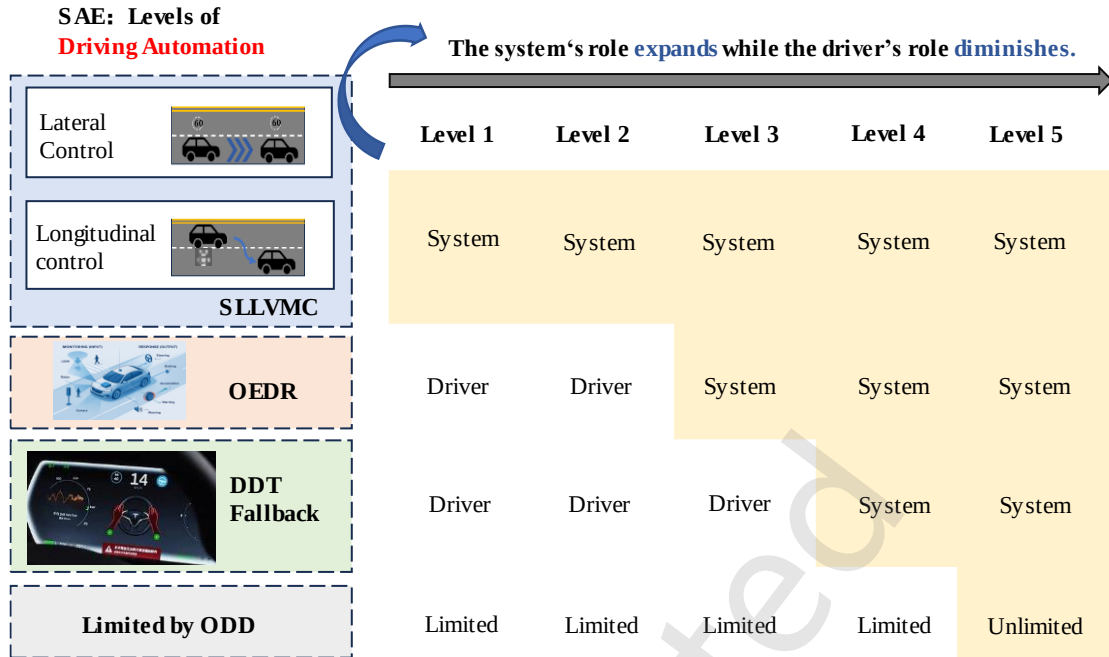


Fig. 1. SAE taxonomy based on the division of human vs. ADS responsibility.

Predicated primarily on the split in dynamic driving tasks between the human and the ADS and fallback responsibility, it fails to adequately characterize the actual operational proficiency, robustness, and reliability of an ADS across all natural and anthropogenic environments. Unlike the linear transfer of control evident from L1 to L3, the progression from L3 to L5 is characterized by a critical disconnect between the assigned automation level and actual operational performance in the field. Initially established by SAE, the ODD delineates the expected operating boundaries—encompassing spatial, environmental, and dynamic traffic parameters—required for the safe deployment of autonomous features (SAE, 2014; NHTSA, 2017; ISO, 2023). The underlying rationale is that by defining constrained ODDs, L3/L4 systems can operate safely and efficiently with the eventual goal of expanding these boundaries to enable functionality under any environmental condition. However, current ODD taxonomies are limited to merely cataloging ODD-related elements, lacking a systematic framework to quantify the operational complexity of these domains (Mendiboure et al., 2023; Salvi et al., 2022). Furthermore, no existing research or standards have correlated specific ODD boundaries with autonomous driving capabilities or automation levels. This points to a critical limitation: while L1-L3 clearly demarcate a progression in control transfer, the definitions for L3-L5 lack a corresponding progression in environmental complexity and ODD expansion. The distinction between ODD-restricted systems (L3/L4) and unrestricted Level 5 represents a categorical leap rather than an evolutionary continuum (Fig. 2).

The absence of ODD specifications within current SAE automation levels creates a critical and profound discrepancy between vehicle's assigned SAE level and its vehicle's practical capabilities. This distinction is illustrated in Fig. 3. The figure uses a coordinate system to visualize representative scenarios, demonstrating that a vehicle's environmental capability may be very different from its defined automation level. The fundamental reason for such dichotomy is the structural deficiency within the current SAE-based evaluation framework. Contemporary standards and classification frameworks

do not include comprehensive ODD specifications for L3/L4 systems. Although SAE standards require explicit ODD definition during vehicle design, they impose no comparative requirements between levels—an L4 system's ODD boundaries need not be stricter or superior to L3's. Similarly, ODD interpretations for L4 vary drastically, potentially rendering a manufacturer's certified L4 system inoperable within another manufacturer's ODD (S. Chen et al., 2023). Such regulatory flexibility has spawned industry-wide inconsistencies in manufacturers' ODD interpretations for the same automation level (Parker et al., 2023). Consumers often subconsciously associate higher automation levels with superior performance irrespective of the operating environment. This ambiguity fosters a critical disconnect between public perception and technical reality. For example, certain AVs designated as L4—such as point-to-point logistics vehicles operating within confined industrial zones, airports or robotaxis in dedicated lines—function under highly restricted conditions. In contrast, L2+ or L3 systems demonstrate autonomous functionality in complex urban driving scenarios. These systems must handle more dynamic environments, including intricate road infrastructure, heavy traffic, regulatory compliance, and variable weather conditions, which impose far greater demands on perception and decision-making capabilities (I. Kim et al., 2023). This suggests that L2+ or L3 system performance may surpass that of some L4 systems. Whenever drivers harbor inflated expectations of vehicle performance or lack a clear conceptualization of operational limits, they are prone to erroneously engaging the system in environments outside its specified ODD. Such ODD violations, where a system is activated beyond its competent domain, can be fatal. These underlying limitations severely complicate the formulation of testing protocols, regulatory oversight, and the establishment of coherent industry standards (Seppelt et al., 2018; Czarniecki, 2018).

1.2. Scope

1.2.1. Focus on SAE levels 3 and 4.

This paper focuses specifically on SAE L3 and L4 systems because they represent the critical transition from driver assistance technologies to

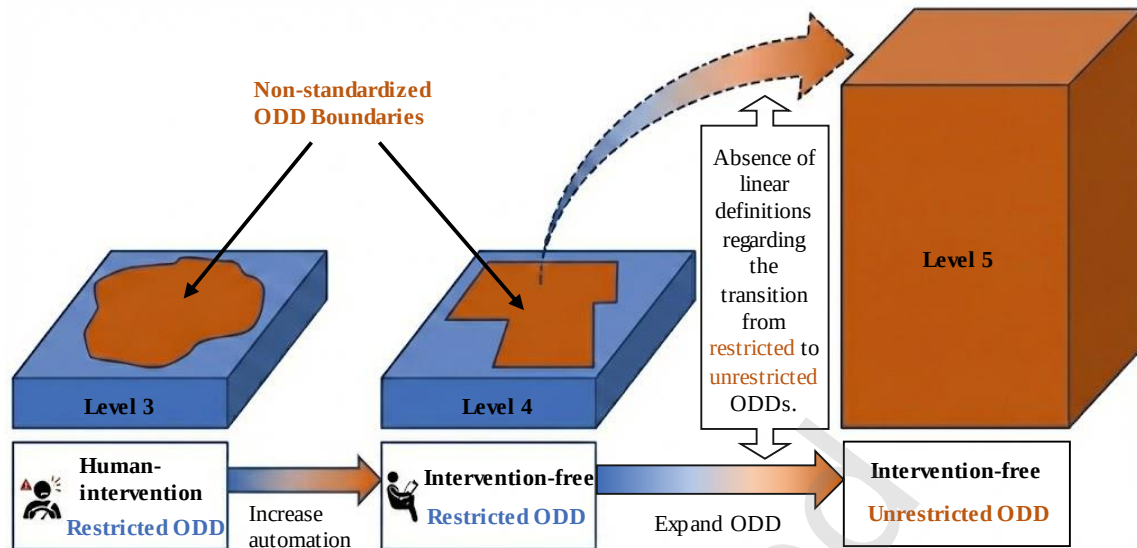


Fig. 2. Reasons for the non-linear nature of L4-to-L5 transition.

actual ADS, adhering to the SAE definition. Unlike L1-L2 systems that provide only longitudinal or lateral control, L3-L4 systems execute the complete DDT within their specific ODDs, differing structurally only in fallback responsibility. By focusing specifically on the transitional tiers of Levels 3 and 4, this paper aims to address the aforementioned limitations regarding how ODDs differentiate these systems and evolve toward unrestricted L5.

1.2.2. Autonomy vs automation

The ongoing terminological debate (Shladover, 2018; Araujo et al., 2023) regarding the terms “automated” and “autonomous” often distracts from more substantive issues. While still far from achieving general autonomy, L3-L4 systems demonstrate limited agency through tactical decision-making capabilities (e.g., deciding whether to overtake or yield). Methodological paradigms in ADS have evolved significantly, recently transitioning from end-to-end neural networks (L. Chen et al., 2024) to VLM (Vision language model)-driven scene understanding (You et al., 2026) and world model-based end-to-end planning architectures (Wang et al., 2024). These advancements enhance system consistency and semantic scene comprehension, as well as unifying semantic reasoning with direct action generation, equipping systems with the requisite world knowledge to navigate open-world challenges. Consequently, it can be argued that the classification of ADS must transcend the singular dimension of human-automation task split, to instead include a supplemental taxonomy that captures their actual operational performance in real-world environments.

1.3. Principal contributions

Current research has not yet established a unified framework capable of systematically and quantitatively characterizing the relationship between environmental complexity and autonomous driving capabilities across different ODDs and automation levels. The primary objective of this study is to provide a supplemental taxonomy to the existing SAE levels. To achieve this, we establish a quantitative correlation between autonomous driving performance and environmental complexity as our core evaluation mechanism. This approach draws inspiration from paradigms in robotics and artificial intelligence (Araujo et al., 2023; H. Chen et al., 2021), particularly the ALFUS (Autonomy Levels for Unmanned Systems) framework, which

evaluates system autonomy across human-robot interaction, task complexity, and environmental complexity (H. M. Huang et al., 2005). Given the highly coupled nature of driving tasks and operational environments during actual ADS deployment (Silva et al., 2022; H. Zhu et al., 2017), we dimensionally merge them to form a new ESC dimension. The ESC is designed to evaluate the system's proficiency in handling driving contexts characterized by aggregated and composite levels of complexity. To systematically characterize and quantify this dimension, we construct the environmental-situational capability (ESC) framework which establishes a hierarchical tree structure where three core environmental dimensions—the built environment, natural environment, and traffic conditions—are each stratified into specific sub-metrics. Building upon this structure, a two-stage hierarchical assessment methodology is developed to establish correlations between road environments and capability benchmarks required for safe autonomous operation. Then, using these quantitative outputs as a basis, we propose an operational-specific extended taxonomy specifically for SAE Levels 3 and 4. This supplementary taxonomy operates orthogonally to existing automation levels, allowing for a more precise classification of AVs in a manner that reflects their capabilities.

The proposed framework aims to fill the critical gap in current taxonomy by constructing a systematic bridge between automation grades and the ODD. Potentially, this approach better reflects the cognitive and decision-making challenges faced by autonomous systems in real-world conditions. From a practical perspective, the proposed supplementary classification framework offers direct utility. It provides system developers with a reference for product capability positioning, regulators with a basis for consistent ODD interpretation, and testing agencies with support for designing capability-matched evaluation procedures. By making environmental complexity explicit and comparable, this framework helps reduce ambiguity during system deployment, supports the progressive evolution of autonomous capabilities from Level 3/4 to Level 5, and ultimately facilitates the commercialization of safe ADS and transparent capability communication with the public.

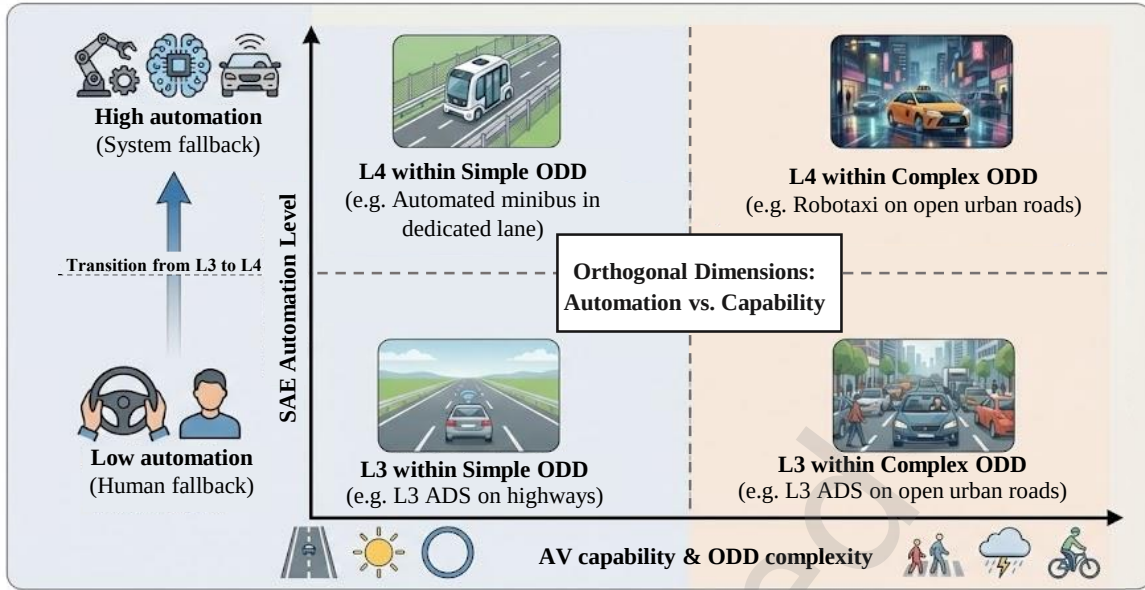


Fig. 3. Dimension orthogonality between automation level and operational capability.

The remainder of this paper is organized as follows: Section 2 reviews the existing literature on ODD taxonomies and environment-based ADS evaluation frameworks. Section 3 establishes the core evaluation metrics of the ESC framework by systematically analyzing ADS environmental vulnerabilities. Section 4 constructs the supplemental taxonomy by mapping evaluation scores to ESC levels, and demonstrates its application through seven illustrative use-case archetypes. Section 5 discusses the practical implications of the proposed classification, along with current limitations and future research directions. Finally, Section 6 concludes the paper.

2 Related works

2.1. ODD taxonomy

The concept of the ODD was initially introduced by SAE, representing the specific set of environmental and operational conditions under which an ADS is specifically designed to function safely (SAE, 2014). If these operational limits are exceeded, the ADS may cease to function safely. However, early conceptualizations of these external conditions lacked systematic categorization and direct practical usability. Given the critical importance of the ODD in performance expectations and safety assurance for autonomous driving, an increasing body of research has focused on establishing more precise ODD terminologies and taxonomies (Czarnecki, 2018; Koopman and Fratrik, 2019). Mendiboure et al. (2023) analyzed existing ODD categorizations to construct a low-overhead, deployable hierarchical framework that uniquely accommodates the constraints of both vehicular platforms and physical road elements. To address the continuously changing open-world operating conditions of autonomous driving, Salvi et al. (2022) introduced a fuzzy logic-based approach that fragments the ODD into smaller pieces, termed μ ODDs. They argued that current ODD designs require the ADS to guarantee compliance with static safety concepts, which may lead to overly conservative systems and a subsequent reduction in acceptable utility. Shakeri (2024) introduced a formalization method that rigorously defined and clarified the relationships among the Operational Domain (OD), ODD, and ODD Specification. Furthermore, a generalized, safety-driven ODD

framework was proposed and embedded into a five-step system engineering process—comprising ODD scoping, initialization, refinement, integration, and business relevance validation (Adedjouma et al., 2024). Another prominent direction for ODD definition is ontology, which establishes the logical relationships among ODD elements to render the methodology machine-readable and verifiable (Czarnecki, 2018). Ontology-based ODDs are widely utilized in ADS testing and validation, facilitating scenario generation (X. Zhang et al., 2024), and operational boundary identification (Sun et al., 2021).

Overall, current ODD research is undergoing a critical transition from abstract safety concepts toward structured and formalized representations. Nevertheless, existing ODD definitions often treat environmental parameters in isolation, lacking a deep integration with driving automation levels and the complex variability of real-world environments. Although the prospect of ODD version control has been envisioned (Mehlhorn et al., 2023), the field still lacks comprehensive frameworks and guidelines for dynamically broadening ODD boundaries during the transition from SAE L3 to L5.

2.2. Environment-based frameworks for ADS capability characterization

Currently, industry-wide consensus still lacks on standardized metrics for rigorously evaluating ADS performance (NHTSA, 2020). Existing studies have developed classification systems and evaluation frameworks from multiple perspectives to address the coupled complexity of driving tasks and operational environments in ADS. A few macro-level theoretical frameworks intended to quantify system autonomy (Antsaklis, 2020; Hrabia et al., 2015), however, remain overly abstract and are difficult to map directly onto specific ODDs.

Acknowledging the decisive role of environmental representation, the research focus has gradually shifted from early dichotomies of static and dynamic elements (Dickmanns and Graefe, 1988) and conventional road design considerations (Jameel and Evdorides, 2016) toward more systematic, hierarchical environmental frameworks tailored for autonomous driving. At the regulatory level, official standards (ISO, 2022; NHTSA, 2020) have defined seven core ODD elements—ranging from environmental conditions to infrastructure specifications—facilitating their systematic application throughout the development

lifecycle. To achieve greater representation, the contemporary PEGASUS framework further expanded environmental characteristics into a six-layer cognitive model (Scholtes et al., 2021; Sun et al., 2021). However, specific evaluation criteria for AV performance across different infrastructure environments remain scarce. Consequently, recent research has focused on assessing road infrastructure compatibility and adaptability specifically for AVs. For example, Radu et al. (2024) formulated a dynamic driving index for Level 4 systems, aggregating multidimensional factors such as infrastructure conditions, digital connectivity, and street object density. Similarly, the concept of automated drivability was introduced to assess the spatial deployment potential of Level 4 vehicles, categorizing road network segments based on their suitability for high automation (Soteropoulos et al., 2020). Expanding from assessment to implementation guidelines, infrastructure requirements for integrating AVs into urban and suburban road networks were considered in the development of a comprehensive infrastructure evaluation framework based on safety, efficiency, and accessibility across diverse land-use contexts (Manivasakan et al., 2021).

Beyond assessment frameworks, S. Chen et al. (2023) proposed a hierarchical system based on roadway infrastructure support to supplement the SAE Level 4 taxonomy and defined the maturity of autonomous capabilities through the readiness of physical road elements. However, a fundamental limitation of this taxonomy lies in its reliance on rigid, pre-defined combinations of infrastructure elements. This framework implicitly assumes a static correlation between infrastructure quality and operational complexity, overlooking the highly dynamic variability inherent in real-world scenarios, such as sudden shifts in weather conditions and traffic flows. Furthermore, another related study proposed a new taxonomy concerning automated driving readiness (Betz et al., 2024). Building upon NASA's framework, they introduced a series of descriptive tags and categorized the technology readiness level of AVs into 9 levels. However, they still failed to establish a concrete, linear correspondence between environmental requirements, capability levels, and technological maturity. Overall, there is still a lack of structured metrics for quantifying the intrinsic complexity of driving environments,

which has prevented the establishment of a critical mapping between ODD challenges and required the level of system intelligence.

3 ADS environmental vulnerabilities and the development of the ESC metric framework

Irrespective of whether the traditional modular architecture or the E2E paradigm (L. Chen et al., 2024) is adopted, the core objective of ADS is to continuously and independently execute the DDT. However, real-world driving environments are complex systems integrating physical, natural, and dynamic elements, which impose significant operational constraints (Z. Zhang et al., 2022). Restricted by physical sensing limits (Vargas et al., 2021), data generalization deficiencies (Yang et al., 2024; H. Li et al., 2025), and onboard computational bottlenecks (Howard et al., 2019), ADS are highly susceptible to performance degradation or systemic failure when encountering extreme conditions such as adverse weather, non-standard roadways, heterogeneous targets, and high-frequency dynamic interactions (Sana et al., 2023).

To systematically quantify the impact of environmental complexity on ADS, a tree-structured evaluation process is developed in the current study. This process categorizes the elements into three distinct groups according to their temporal variability: the built environment, natural environment, and ambient traffic conditions. Specifically, the built environment evaluates infrastructure readiness and supportiveness to delineate a deterministic operational baseline, the natural environment assesses the physical favorability of ambient conditions on perception reliability and control stability, and the traffic condition measures the impact of traffic states—whether homogeneously stable or complex and chaotic—on operational navigability. Within these three dimensions, based on the specific perception and planning vulnerabilities encountered by the ADS, each main metric is further stratified into three sub-metrics. Each sub-metric is qualitatively rated across four stringency levels—favorable, moderate, severe, and uncontrollable—depending on its relative threat to AV operations (Fig. 4). By considering a scenario spectrum ranging from ideal conditions (e.g., structured geometry, clear weather, and uniform traffic) to extreme edge cases, the ESC evaluation framework

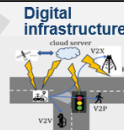
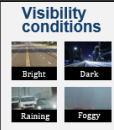
	 Roadway infrastructure  Traffic control infrastructure  Digital infrastructure	 Visibility conditions  Natural dynamic disturbance  Road surface friction	 Agent heterogeneity  Flow dynamics  Interaction complexity
Favorable	- Well-maintained freeways providing Predictable geometries, clear zones, and minimal roadside clutter - Highly visible, compliant, and well-maintained signals and signs - Ubiquitous, high-reliability connectivity enabling seamless V2V/V2I/V2P data exchange	- Clear daytime with ample sunlight and optimal sensor clarity - Smooth pavement with no obstructions and calm conditions - Dry pavement offering high and consistent friction coefficients	- Homogeneous traffic with uniform vehicle types and predictable movement patterns - Low-density traffic with stable, moderate speeds - Segregated roadways with minimal crossing conflicts and predictable movements
Moderate	- Urban roads with mixed-use roadside features and occasional obstructions - Partially obscured, faded, or non-standard devices requiring increased interpretation - Partial coverage or intermittent connectivity requiring fallback to onboard sensing	- Dawn/dusk or light rain/fog causing moderate sensor noise and range reduction - Moderate crosswinds or minor road debris requiring stability adjustments - Wet pavement or light contamination causing moderately reduced traction	- Mixed urban traffic combining cars, buses, and trucks with moderate behavioral diversity - Moderate-density traffic with occasional speed variations - Urban intersections with moderate interaction complexity and occasional yielding scenarios
Severe	- Irregular layouts, construction zones, or unsecured hazards causing localization conflicts - Missing, damaged, or conflicting control devices creating ambiguous right-of-way - Absence of V2X infrastructure, forcing complete reliance on autonomous vehicle perception	- Extreme low-light, heavy precipitation, or fog inducing severe sensor degradation - Severe crosswinds, sudden road intrusions, or deep potholes challenging vehicle control - Ice, snow, or heavy mud inducing severe traction loss and instability	- Highly heterogeneous traffic exhibiting varied and unpredictable movement characteristics - High-density traffic with volatile speed fluctuations - Uncontrolled mixed-traffic zones with dense, unpredictable interactions and frequent edge cases
	➤ Built environment	➤ Nature environment	➤ Traffic conditions

Fig. 4. ESC framework: evaluation matrix and severity assessment.

can serve as the evaluative baseline and reference for the proposed supplemental taxonomy.

3.1. Infrastructure supportiveness of the built environment

The built environment constitutes the fundamental physical space, semantic boundaries, and informational support for AV navigation. It encompasses static physical entities such as lanes and curbs, traffic control facilities that provide rule-based guidance, and digital infrastructure serving as a situational awareness compensation layer. Under ideal conditions, highly standardized static infrastructure, such as clear curbs and lane markings, can provide precise prior constraints, significantly facilitating robust algorithmic perception and operation (Carruth et al., 2022). Building upon this, dedicated AV guideways represent the highest tier of infrastructure support (S. Chen et al., 2023; Carreras et al., 2018) because they not only eliminate external dynamic interference via physical isolation but also can deeply integrate V2X cooperative systems, providing comprehensive information redundancy and augmentation (Tan et al., 2024). This highly deterministic environment minimizes the complexity of perception and decision-making. However, real-world roads exhibit significant heterogeneity. In actual deployment, ADS frequently encounter urban streets with highly variable topological structures or extreme operating conditions characterized by degraded or entirely absent infrastructure (Soteropoulos et al., 2021). For instance, on underdeveloped rural roads or off-road terrains, the disappearance of physical boundaries is often accompanied by the deprivation of digital augmentations, and may even suffer from GNSS signal denial caused by terrain occlusion, forcing the system to operate under highly constrained conditions (NHTSA, 2019; Ansarinejad et al., 2025a; Carruth et al., 2022). Field tests confirm that such adverse road conditions are the primary triggers for unplanned disengagements or phantom braking (Lv et al., 2017).

From the three dimensions of physical accommodation, regulatory guidance, and informational augmentation, this section refines the evaluation of the built environment into the following three core sub-indicators: Roadway infrastructure, Traffic control infrastructure and Digital infrastructure.

3.1.1. Roadway infrastructure

This metric evaluates the compatibility of geometric topological features and roadside facilities with ADS operations. Roadway geometric features encompass horizontal and vertical alignments, cross-sectional profiles (e.g., lane widths and curb boundaries), and intersection designs (Miaou and Lum, 1993). The road hierarchy directly reflects the comprehensiveness of the design. Higher-tier roads exhibit greater structural consistency, significantly mitigating the risk of lane departures (Y. Chen et al., 2018). However, existing conventional designs cannot be seamlessly leveraged by AVs. Standard roadway designs may require targeted adaptations for autonomous vehicles, particularly within mixed traffic flows that demand more stringent safety measures (Othman K., 2021). Narrow lanes significantly escalate the failure rates of Level 2-3 automated vehicles (García and Camacho-Torregrosa, 2020). Roadside facilities, comprising both structural and functional elements, collectively shape the roadway physical boundaries and the visual and spatial characteristics of the driving environment. Nevertheless, the density and spatial distribution of these standard elements simultaneously introduce inherent perceptual constraints. Vertical structures and vegetation frequently generate static occlusions (Ruan et al., 2023). These physical barriers, such as greenery along the road, actively

conceal Vulnerable Road Users (VRUs) and severely compromise the semantic coherence of the environment (M. Zhu et al., 2022). The intricate textures and overlapping geometries of vegetation and street furniture pose severe challenges to semantic segmentation algorithms, often resulting in boundary ambiguity and impairing scene comprehension capabilities (Muhammad et al., 2022). The most critical challenges within the roadway environment arise from unforeseen temporary alterations (J. Yang et al., 2024; H. Li et al., 2025), such as natural or anthropogenic roadway incursions, work zones, and temporary traffic control. These scenarios pose significant challenges to the perception and trajectory re-optimization capabilities.

3.1.2. Traffic control infrastructure

Traffic signs, pavement markings, and traffic signals are often conceptualized as the language of the road, conveying static and dynamic driving regulations. Under favorable conditions, highly standardized and well-maintained traffic control devices can furnish ADS with deterministic semantic constraints. However, deviations from this ideal state expose critical vulnerabilities in the system's semantic recognition capabilities.

First, pavement markings establish multidimensional spatial guidance. Longitudinal markings delineate navigable spaces, while transverse and special markings regulate intersection behavior. Clear lane markings are crucial for AVs to maintain stable trajectories and mitigate collision risks with adjacent vehicles (Zhang et al., 2021). However, suboptimal visibility caused by pavement wear, insufficient luminance, or degraded retro reflectivity introduces substantial perceptual uncertainty. When lane markings fade by more than 50%, the lane-keeping performance of AVs experiences severe degradation (Formosa et al., 2024). Second, traffic sign systems operate across regulatory, informational, and temporary hierarchies, playing a pivotal role in road operations by conveying critical visual semantics via distinct shapes, colors, backgrounds, and spatial placements (Serna and Ruichek, 2018). However, from a macroscopic generalization perspective, fundamental discrepancies in traffic sign appearances across different jurisdictions can pose a formidable challenge to the continuous training and domain adaptation of ADS (Aldoski and Koren, 2023). Semantic ambiguity induced by conflicting temporary signs or overlapping obsolete and current road surface markings can trigger failures related to critical decision-making. Finally, traffic signals allocate temporal right-of-way to ensure orderly traffic flow. In terms of traffic light detection and recognition, computer vision methods encompassing preprocessing, segmentation, feature extraction, classification, and tracking have been extensively investigated (Zhu et al., 2016; Houben et al., 2013). However, accurate recognition remains a severe challenge due to the persistence of confounding factors such as variance in scale, shape, and positional diversity, as well as orientation, bounding box precision, occlusion, and overlapping elements (Gautam and Kumar, 2023). Furthermore, the most fundamental and critical limitation is the complete absence of defined infrastructure, such as undeveloped rural roads or at unsignalized intersections where the system is forced to fallback entirely on ego-vehicle sensors for autonomous reasoning and planning.

3.1.3. Digital infrastructure

Digital infrastructure in ADS operation includes Vehicle-to-Everything (V2X, encompassing V2V, V2I, and V2P) communication, cloud platforms, wireless networks, and roadside units. Under this

standardized cooperative framework, digital infrastructure can theoretically function as a digital sensor and a critical situational awareness compensation layer. It can help overcome the physical limitations of line-of-sight (LOS) occlusions and computational constraints inherent in single-vehicle autonomy (Carreras et al., 2018; Tengilimoglu et al., 2023). In a fully interconnected environment characterized by ubiquitous 5G coverage and dense deployments of RSUs, vehicle-cloud synergy elevates the isolated motion planning of ego-vehicles into globally coordinated multi-agent decision-making. This enables advanced collaborative tasks such as collision warning, vehicle platooning, and dynamic work zone navigation (Zha et al., 2024; Z. Xu et al., 2025; Mo et al., 2024; Tan et al., 2024).

However, such infrastructure support inversely scales with the system's requisite onboard intelligence. (S. Chen et al., 2023). AVs exhibiting an over-reliance on intelligent infrastructure experience severe functional degradation when traversing areas with diminished external support. First, heterogeneous deployments driven by cost and geographical factors lead to severely fragmented network coverage (e.g., penetration in rural areas is significantly sparser compared to urban roads). Concurrently, V2P communication aimed at VRUs fails to reliably substantiate critical safety decisions due to device instability and low market penetration rates (Yusuf et al., 2024). Even within network coverage zones, data exchange is constrained by Quality of Service (QoS) metrics. Random signal latency and packet loss can easily induce the system to process stale data, thereby compromising decision-making logic (Kamtam et al., 2024). Furthermore, cybersecurity attacks such as malicious data spoofing, signal jamming, or message tampering constitute catastrophic failure modes, introducing intolerable safety hazards (X. Yang et al., 2022; Giannaros et al., 2023; Gyawali et al., 2020). Therefore, state-of-the-art ADS must proportionally mitigate their reliance on potentially vulnerable digital infrastructures, fostering robust, standalone perception and decision-making capabilities.

3.2. Physical favorability of the natural environment

The natural environment refers to the broad physical environment independent of human-engineered infrastructure, primarily encompassing meteorological conditions, atmospheric characteristics, illumination, and the natural physical state of the terrain. In stark contrast to the deterministically designed built environment, the natural environment is characterized by inherent stochasticity and highly dynamic variability. Its impact on autonomous driving predominantly targets the perception, vehicle control, and localization modules.

First, the most direct and fundamental impact of the natural environment on ADS is exerted on the sensor hardware layer. AVs universally rely on multi-modal sensor suites—comprising cameras (Janai et al., 2020; Asvadi et al., 2018; Rajapaksha et al., 2024), LiDAR (Aung et al., 2024; Li and Ibanez-Guzman, 2020; H. Wang et al., 2025; Zang et al., 2019), and millimeter-wave radars (Bilik et al., 2019; Fan et al., 2024)—to acquire raw environmental data. This acquisition forms the indispensable basis for subsequent trajectory planning and vehicle control. However, unpredictable fluctuations in illumination, precipitation, and aerosols (e.g., fog and haze) can rapidly alter the physical properties of atmospheric media. These phenomena induce severe signal refraction, energy absorption, and physical occlusion,

fundamentally degrading the detection integrity and sensing reliability of the perception system and potentially precipitating complete perceptual failure. (Aloufi et al., 2024; Z. Chen et al., 2025; Vargas et al., 2021). Empirical reports from the NHTSA explicitly identify perception anomalies as the leading cause of ADS operational failures (NHTSA, 2021). Statistically, adverse weather and low-light environments account for a disproportionate majority of environment-related crashes and pedestrian fatalities, respectively (Federal Highway Administration (FHWA), 2025). Furthermore, beyond directly degrading perception performance, the natural environment can also breach the physical limits of vehicle dynamics control. For example, dynamic factors including pavement water accumulation, icing, or severe crosswinds could drastically alter tire-road adhesion characteristics and the aerodynamic state of the vehicle body.

To comprehensively quantify the multi-dimensional constraints imposed by the natural environment on ADS operational capabilities, this section dissects the environmental assessment into three core sub-indicators: meteorological-luminous conditions, natural dynamic disturbances, and road surface friction.

3.2.1. Visibility conditions

This metric quantifies the signal interference and perceptual degradation inflicted upon ADS by severe atmospheric disturbances and abrupt illumination fluctuations. The underlying mechanisms of these impacts have been extensively validated in existing literature (Vargas et al., 2021; Carruth et al., 2022). Distinct environmental stressors exert differential, yet markedly adverse, effects across various sensing modalities. Vision systems demonstrate inherent advantages in extracting semantic and geometric features (Asvadi et al., 2018). However, inadequate lighting yields road environment images characterized by low luminance, poor contrast, and amplified noise (Y. Wang et al., 2024; H. Lin et al., 2024), which severely constrains their long-range depth estimation capabilities (Rajapaksha et al., 2024). Conversely, severe lens glare induced by direct sunlight or oncoming headlights at night not only blinds optical vision systems, but intense specular reflections can simultaneously disrupt the signal emission and reception of LiDAR systems (Vargas et al., 2021).

Precipitation (e.g., raindrops and snowflakes) and aerosols (e.g., fog and haze) rapidly alter the physical properties of the atmospheric transmission medium. Optical occlusion and refraction caused by rain severely corrupt image features, precipitating a drastic deterioration in the perception performance of deep neural networks (Hnewa and Radha, 2020). Although existing deep learning-based image enhancement techniques have achieved notable advancements (Tahir et al., 2024), the irreversible loss of physical signals is often difficult to fully mitigate. For example, de-raining algorithms, while effectively removing rain streaks, are highly susceptible to over-smoothing object edges, thereby inducing a secondary loss of critical geometric features (Hnewa and Radha, 2020). Simultaneously, suspended particulates such as fog and haze severely attenuate the signal propagation of short-wavelength LiDAR and ultrasonic sensors, culminating in resolution degradation, point cloud sparsity, and a drastically curtailed detection range (Zang et al., 2019). Although LiDAR is indispensable for high-precision SLAM localization, its technical bottlenecks under adverse weather conditions remain a formidable challenge (Aung et al., 2024; H. Wang et al., 2025).

3.2.2. Natural dynamic disturbances

The physical stability of autonomous vehicles confronts multifaceted challenges from natural environmental factors, which simultaneously compromise perception, decision-making, and control performance through complex interactions. Pavement irregularities (e.g., potholes, gravel, and water pooling) not only induce severe chassis vibrations that significantly degrade sensing reliability but also impair vehicle dynamics and escalate the steady-state risks of the control system (Lowe and Guvenç, 2024). Furthermore, the influence of crosswinds is critically significant. When wind speeds exceed 60 km/h, the substantial lateral aerodynamic forces generated can induce severe trajectory deviations. Such compound disturbances mandate that the system simultaneously execute sensor data calibration and vehicle stability control within stringent response timeframes, thereby imposing extreme demands on the real-time processing capabilities of the ADS (El-Demerdash et al., 2005).

3.2.3. Road surface friction

This metric focuses on the tire-road physical interaction, specifically the degradation of the pavement friction coefficient induced by environmental factors such as elevated moisture, icing, or snow accumulation (Zheng et al., 2019). Such traction loss poses critical safety hazards, as evidenced by the significant inverse correlation between friction coefficients and accident rates (Wallman and Åström, 2001). If the disturbances stemming from the loss of road adhesion are neglected, the vehicle will fundamentally fail to execute the intended trajectories formulated by the perception and planning modules. Therefore, ADS must seamlessly integrate real-time friction estimation with adaptive control strategies, ensuring that any maneuvers executed in adverse environments do not exceed the vehicle's physical stability limits.

3.3. Operational navigability of traffic conditions

The traffic environment is primarily constituted by diverse road users, encompassing a wide array of motorized vehicles and VRUs such as non-motorized vehicles and pedestrians. The instability of this environment manifests as highly dynamic fluctuations across both spatial and temporal dimensions (Stathopoulos and Karlaftis, 2003). A single operational trip may span a spectrum of traffic conditions, ranging from stable freeway flows to narrow residential streets and congested peak-hour traffic. In the foreseeable future, AVs and human-driven vehicles (HDVs) will inevitably continue to navigate shared physical infrastructures, especially within spatially constrained and highly interactive road environments. (Morando et al., 2018; Yan et al., 2023). This mixed traffic ecosystem, dominated by heterogeneous entities, not only exhibits extreme physical diversity but also poses severe challenges in macroscopic contexts (flow evolution) and microscopic contexts (individual vehicular interactions).

The impact of traffic conditions on AVs is concentrated primarily within the perception, decision-making, and planning modules. On the one hand, diverse target morphologies and highly cluttered traffic backgrounds significantly escalate the difficulty of robust detection and tracking (Ravindran et al., 2020). On the other hand, complex kinematic variations and multi-agent game-theoretic interactions impose prohibitive computational and decision-making burdens on intent inferencing, trajectory prediction, and cooperative planning (Badue et al., 2021). To comprehensively quantify the constraints imposed by the traffic environment on AV operational capabilities, this section dissects its evaluation into three core sub-metrics: agent heterogeneity, flow dynamics and interaction complexity.

3.3.1. Agent heterogeneity

This metric characterizes the diversity and variance of traffic participants in terms of physical morphology, kinematic profiles, and behavioral patterns. In low-heterogeneity scenarios, such as controlled-access highways or dedicated lanes for autonomous driving, traffic flow predominantly consists of standard motor vehicles with uniform features and predictable trajectories. Conversely, in high-heterogeneity scenarios like urban mixed traffic, the right-of-way is shared among passenger cars, commercial vehicles, two-wheelers, and pedestrians. Regarding standard vehicle detection, the system must overcome potential vast discrepancies in geometric features (e.g., shape, size, and color), multi-viewpoint perspective distortions, and the inherently cluttered background interference present in complex traffic scenes (Galvao et al., 2021). VRUs further exacerbate the difficulty of robust detection and intent prediction due to their high morphological variance, rapid kinematic state transitions, and diverse visual scales (Galvao et al., 2021; T. Li et al., 2025). The most critical threat posed by heterogeneity lies in the proliferation of OOD edge cases. These encompass occlusions (Ruan et al., 2023), small targets (Wei et al., 2024), and anomalously shaped objects (Zhao et al., 2024), such as pedestrians in atypical attire and dropped cargo. When confronted with these non-standard obstacles that fall outside the training distribution, conventional learning-based perception pipelines frequently suffer from severe false negatives, as they lack the fundamental world knowledge (Y. Wang et al., 2024; K. Wang et al., 2025).

3.3.2. Flow dynamics

This metric is primarily characterized by the dynamic coupling of traffic volume, traffic density, and operating speed, capturing the impact of the traffic environment's macroscopic spatiotemporal state on the microscopic operations of the ego-vehicle. Moderate density and stable mid-range speeds are considered optimal conditions (Cappellari and Weber, 2022). Under such conditions, the motion state identification and trajectory estimation of surrounding vehicles become highly tractable for the ego vehicle. However, drastic fluctuations in flow dynamics introduce heightened unpredictability. A surge in traffic volume and density exacerbates physical occlusions and feature overlaps. These not only drastically increase the difficulty of robust target detection but also cause the computational complexity of macroscopic scene understanding to escalate sharply (Hussain and Zeadally, 2018). High-speed operations drastically amplify the potential consequences of perception latency or decision-making errors. Statistical research indicates a positive correlation between traffic volume and collision frequency. However, accident severity exhibits a significant negative correlation with traffic volume due to the relative elevation of operating speeds under sparse traffic conditions (Cappellari and Weber, 2022). The most formidable challenge arises from the volatile confluence of high traffic density and drastic speed fluctuations, such as stop-and-go shockwaves.

3.3.3. Interaction complexity

This metric characterizes the frequency and unpredictability of dynamic engagements between AVs and other road users. Such interactive paradigms frequently materialize in manifest during lane changing (Ma and Li, 2023), ramp merging (X. Zhu et al., 2022), unsignalized intersections (T. Li et al., 2025), and roundabouts (Lin et al., 2025). Surrounding vehicles and pedestrians, much like the ego-vehicle, operate as active agents possessing independent motion planning capabilities. Such multi-agent interaction complexity dictates

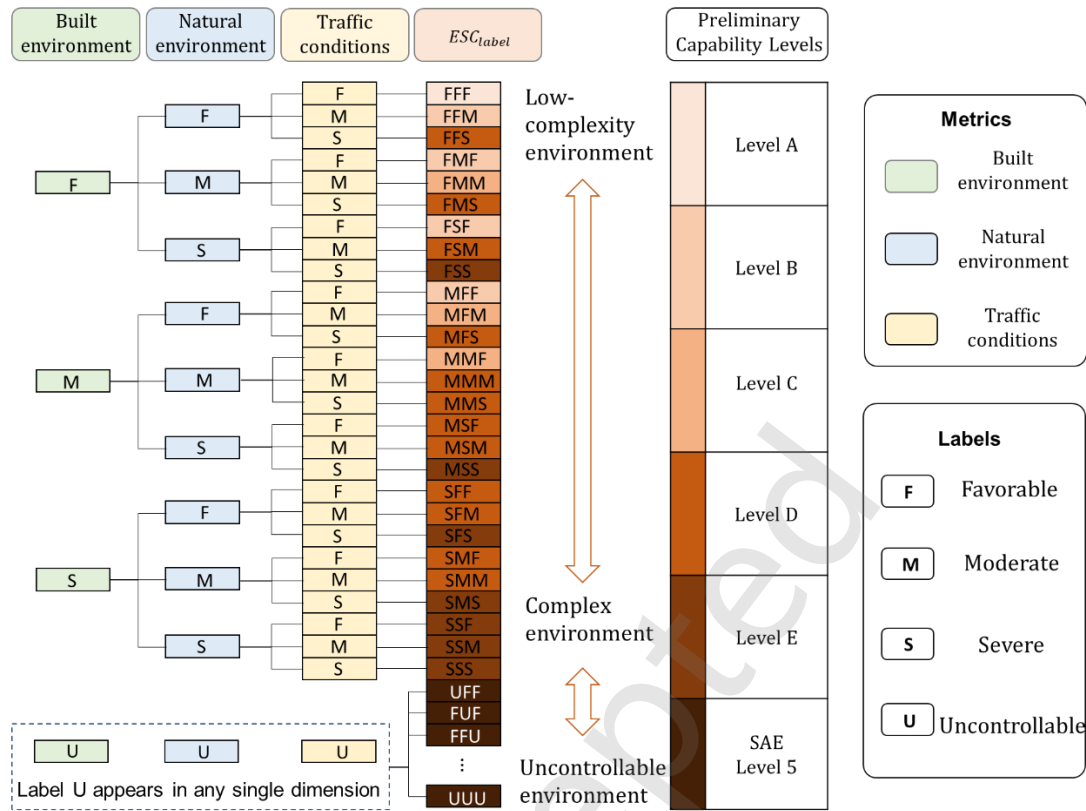


Fig. 5. Tree structure and corresponding preliminary levels based on expert evaluation labels

that the system must achieve a dynamic equilibrium within a continuous, coupled game-theoretic framework—simultaneously predicting the dynamic behaviors of others while adapting its own trajectory—to guarantee safe operational boundaries (H. Zhu et al., 2022). This interaction is further complicated by heterogeneous behavioral styles among agents (e.g., aggressive versus defensive) (H. Zhu et al., 2022) and the inherent bounded rationality of human behavior (T. Li et al., 2026). Vehicles may execute various evasive maneuvers, such as lane changes or deceleration, to prevent collisions (Sheykhsfard et al., 2023). Conversely, pedestrians can alternate between running, walking, or sudden dodging actions, abruptly altering their trajectories and velocity profiles within much shorter timeframes which renders interaction modalities exceptionally intricate (Li et al., 2025; Li et al., 2026). The most formidable challenges erupt within high-frequency interaction zones, such as unsignalized intersections, congested ramps, and mixed-use bustling streets. Research indicates that the ambiguous right-of-way rules in these areas generate frequent edge cases, which remain a common bottleneck across existing autonomous driving paradigms (Sana et al., 2023; Arvin et al., 2020).

As illustrated in Fig. 4, these three environmental categories—the built environment, natural environment, and traffic environment—each is rated as favorable, moderate, severe, uncontrollable, depending on their threat level to AVs. The built environment remains temporally stable, while the natural and traffic environments typically undergo significant temporal variation. On one extreme, a favorable condition is one characterized by well-defined infrastructure in the built environment, clear weather conditions in the natural environment, and uniform vehicle types, low and steady speeds, and light traffic flow in the traffic environment. On the other extreme, unfavorable conditions

is one characterized complex or unstructured road geometries combined with adverse weather phenomena such as heavy rain or fog and aggressive mixed-traffic flows featuring high densities and erratic speed fluctuations.

4 SAE supplemental taxonomy based on ESC framework

4.1. Establishing operational tiers via two-stage hierarchical mapping

The factors discussed in the previous section collectively impose requirements on AV environmental capability levels. However, many current approaches in this field treat these influencing factors as having merely additive effects (Manivasakan et al., 2021; Soteropoulos et al., 2020). For example, a Swedish road case study used the Analytic Hierarchy Process (AHP) methodology where experts assigned weights to individual indicators and simply summed all element scores (Radu et al., 2024). This additive approach inherently assumes that environmental factors are fully compensatory. For instance, if the three environmental dimensions are scored 7, 0, and 8, a simple additive model would yield a deceptively acceptable total score of 15. However, in real-world autonomous driving, a critical failure in a single domain—such as severe weather causing complete sensor degradation (a score of 0)—renders the vehicle fundamentally inoperable, regardless of how optimal the conditions in other dimensions might be.

The ESC framework therefore adopts a zero-preserving, partially compensatory multiplicative aggregation model. If any domain is assigned the Uncontrollable state, with $c_i = 0$, the aggregate favorability becomes zero and cannot be offset by the other dimensions. When all domain scores are positive, however, higher favorability in

one domain may partially offset lower favorability in another. Compared with a linear additive average, the multiplicative formulation remains more sensitive to low-scoring dimensions and imbalanced environmental profiles. To establish a structured mapping between the resulting scores and their corresponding operational tiers, we use a two-stage hierarchical classification strategy. Specifically, based on the environmental definitions and severity criteria established in Section 3 and the relative influence of sub-metrics established in the preceding framework, each environmental metric is evaluated on a discrete 0-9 scale. This evaluation is decomposed into two interconnected steps: preliminary classification based on evaluation labels, and boundary calibration driven by specific score products.

4.1.1. Preliminary classification

As defined in Fig. 4 of Section 3, each environmental domain is first assigned one of four severity categories according to its threat to ADS operation: Favorable, Moderate, Severe, or Uncontrollable. To provide greater resolution within the three non-extreme categories, each is divided into three ordered grades, yielding Favorable = 7-9, Moderate = 4-6, and Severe = 1-3, while Uncontrollable remains a separate zero state. In the absence of empirical support for unequal intervals, equal-width subdivision is adopted as the least-assumptive convention.

Let $I = \{B, N, T\}$ denote the set of environmental domains (Built, Natural, and Traffic), and $c_i \in [0, 9]$ represent the specific expert evaluation score for a given domain $i \in I$. We define a mapping function $\Phi(c_i)$ to transform the quantitative score into a qualitative label $l_i \in \{F, M, S, U\}$.

The mapping criteria are shown as Eq. (1).

$$l_i = \Phi(c_i) = \begin{cases} F, & \text{if } c_i \in [7, 9] \\ M, & \text{if } c_i \in [4, 6] \\ S, & \text{if } c_i \in [1, 3] \\ U, & \text{if } c_i = 0 \end{cases} \quad (1)$$

- Favorable (F): Environments where ADS do not require high capability to operate effectively. The corresponding score range is $c_i \in [7, 9]$.
- Moderate (M): Environments with intermediate challenges, requiring moderate capability. The corresponding score range is $c_i \in [4, 6]$.
- Severe (S): Environments with significant challenges, demanding high capability. The corresponding score range is $c_i \in [1, 3]$.
- Uncontrollable (U): Conditions under which road operation remains physically feasible, but require idealized L5-level capability: $c_i = 0$. Conditions under which safe road operation is physically infeasible are outside the scope of the ESC taxonomy.

Following this mapping, for any specific driving scenario, its preliminary capability label profile is defined as a 3-tuple ESC_{label} in Eq. (2).

$$ESC_{label} = (l_B, l_N, l_T) \quad (2)$$

Fig. 4 illustrates a tree structure organized according to these metrics, mapping the ESC_{label} combinations to their corresponding preliminary capability levels. Within this framework, a total of 27 distinct non-extreme combinations is generated (e.g., FFF, FFM) and categorized into five tiers, ranging from Level A to Level E. Combinations comprising exclusively F labels—such as well-structured spacious corridors, optimal illumination, and negligible traffic interactions—represent the lowest operational complexity, mapping directly to Level A. This indicates that the autonomous system requires only baseline environmental capabilities to maintain safe operation. Conversely, as

environmental conditions deteriorate, proportionally higher capabilities are demanded to mitigate the escalating complexity. Intermediate conditions span Levels B through D, while Level E encapsulates the most complex scenarios within the non-extreme operational domain. Crucially, whenever the label U appears in any single dimension, it acts as an absolute boundary condition and excludes the scenario from the 27 graded L3/L4 combinations. A U rating may represent a physically passable but highly challenging condition, such as an unstructured off-road environment without identifiable lane boundaries or external localization support. Such scenarios are mapped to the idealized L5 boundary, while conditions outside any safe operational domain are excluded from the ESC taxonomy.

4.1.2. Overlapping label intervals calibration

Preliminary categorization based solely on evaluation labels lacks sufficient granularity to capture subtle environmental variations. Therefore, a refined scoring approach that uses a multiplicative aggregation model to achieve finer discrimination, was adopted. Let $\alpha_i > 0$ represent the weighting factor for environmental-domain score, subject to $\sum_{i \in I} \alpha_i = 3$. During the multiplicative aggregation phase, the term $c_i^{\alpha_i}$ quantifies the weighted operational influence of environmental domain i on the final assessment. Thus, Eq. (3) calculates the Environmental Favorability Score (EF_{score}).

$$EF_{score} = \prod_{i \in I} c_i^{\alpha_i} \quad (3)$$

The initial EF_{score} is positively correlated with environmental favorability—higher scores indicate more supportive environments. This is intentionally designed to ensure that uncontrollable conditions are assigned a score of zero, allowing the multiplicative model to propagate their impact to the final evaluation. Taking the natural logarithm of both sides and isolating the contribution of domain i gives $\ln EF_{score} = \alpha_i \ln c_i + \sum_{j \neq i} \alpha_j \ln c_j$. Differentiating with respect to yields $\frac{\partial \ln EF_{score}}{\partial \ln c_i} = \alpha_i$. Thus, within the continuous representation of the aggregation function, is the model-based elasticity of with respect to the coded score.

To reflect the inverse relationship where higher complexity poses greater operational risk, we transform this metric. Assuming the theoretical maximum favorability is $EF_{score}^{max} = \prod_{i \in I} 9^{\alpha_i}$. The environmental-situational capability score ESC_{score} is derived by subtracting the calculated favorability from this maximum, as shown in Eq. (4). Under this mathematical treatment, as environmental hostility increases, the resulting ESC_{score} monotonically increases, ensuring consistent directionality in the risk assessment.

$$ESC_{score} = \prod_{i \in I} 9^{\alpha_i} - EF_{score} \quad (4)$$

In the present formulation, the three environmental domains are assigned equal weights, with $\alpha_B = \alpha_N = \alpha_T = 1$. Unless explicitly stated otherwise, all subsequent ESC score calculations, illustrative use-case evaluations, and tier assignments are conducted under this equal-

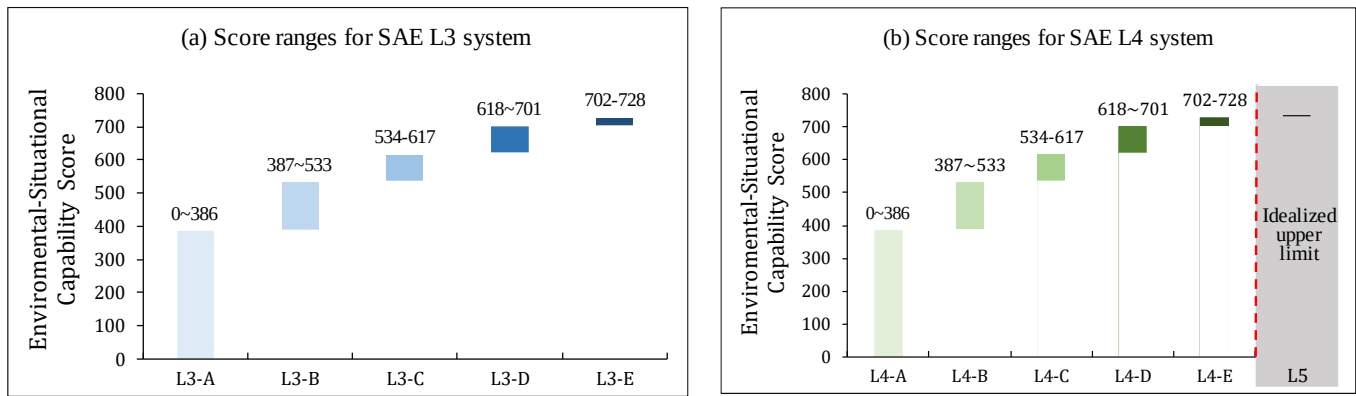


Fig. 6. Quantitative mapping of ESC scores to supplementary levels. (a) Score ranges for SAE Level 3 systems; (b) Score ranges for SAE Level 4 systems. The score of 729 is set as the idealized upper limit and is not assigned to any ESC level.

weight baseline. Equal weighting is adopted as the least-assumptive choice, since no empirical evidence currently supports systematically assigning greater importance to any one domain. The current model should therefore be interpreted as an equal-weight preliminary framework, while calibration of unequal weights is reserved for future empirical research.

While labels establish preliminary tiers, the multidimensional nature of the framework introduces boundary overlap between adjacent levels. In the equal weights baseline cases ($\alpha_i = 1$), for example:

- The worst-case scenario (minimum ESC_{score}) for the FFF label configuration (mapping to Level A) is calculated as $729 - EF_{score}[7,7,7] = 386$;
- The best-case scenario (minimum ESC_{score}) for the FFM label configuration (mapping to Level B) is calculated as $729 - EF_{score}[9,9,6] = 243$.

This produces an overlap between the semantic label combinations and the quantitative score intervals. The FFM profile is more favorable in two dimensions and less favorable in one dimension, and the lower score reflects the partial compensation permitted among positive domain scores in the multiplicative aggregation. To resolve this boundary overlap, we implement the following quantitative precedence rules: (1) Each autonomy level's boundary is strictly defined by the worst-case score of its anchoring label combination (e.g., the threshold for Level A is $ESC_{score} \leq 386$). (2) If a scenario's calculated score is mathematically lower than the threshold of the preceding tier, the quantitative score overrides the initial semantic constraint, and the scenario score is rigorously reclassified as that of the lower tier. This rule provides a consistent mapping between the coarse labels and the finer quantitative score intervals.

4.2. ESC supplementary levels and use cases

Fig. 6 presents the quantitative mapping between the calculated ESC scores and the ultimate supplementary levels. For each SAE automation level from L3 to L4, we have defined five subclasses (A through E) based on these scores. As visually depicted in the figure, the scoring intervals follow a non-linear distribution, forming a pyramid-like structure: wider score ranges are allocated to lower ESC levels, whereas progressively narrower intervals are assigned to higher classification tiers. This structure reflects the significantly increasing technical challenge associated with achieving higher levels of ADS robustness. Lower ESC subclasses (e.g., Level A) correspond to more

restricted operational domains and less demanding environmental conditions, whereas higher subclasses (e.g., Level E) demonstrate the capability to handle escalating environmental complexity. It is worth noting that the maximum theoretical ESC score of 729 is intentionally excluded from the L3 scoring intervals and the ideal SAE Level 5 is consequently evaluated separately within the subsequent two-dimensional matrix (Fig. 7). This distinction is rooted in the definitional prerequisites of the SAE taxonomy. Level 5 represents the absolute convergence of full automation and an unrestricted ODD, requiring zero human intervention under any circumstances. Because an L3 system is fundamentally predicated on the availability of a human fallback-ready user, granting it the maximum ESC score would imply completely unrestricted operational capability, thereby creating a definitional paradox with Level 5.

Building upon this quantitative foundation, Fig. 7 introduces the orthogonal two-dimensional classification framework. It should be noted that in the proposed system, the classification of driving automation levels will continue to adhere to the standards established by SAE, and our developed ESC evaluation criteria are orthogonal in nature to the SAE levels. This relationship is instantiated in two-dimensional matrix, clarifying the distinctions and connections between the ESC taxonomy and the traditional SAE system. For example, an L3-C system may require “a driver who remains ready to take over control after system activation, and the ability to operate in light rain, within structured urban environments with moderate interaction.” In contrast, an L4-A system denotes “no need for a human driver, with the system performing the DDT fallback, and is capable of operating in well-lit daytime conditions within fixed routes in confined areas.”

To concretely demonstrate this two-dimensional positioning, Table 1 and Fig. 7 annotate the locations of seven distinct autonomous driving use cases within the ESC taxonomy. These include five currently prevalent systems (L3 Highway Pilot System, L3 Urban Pilot System,

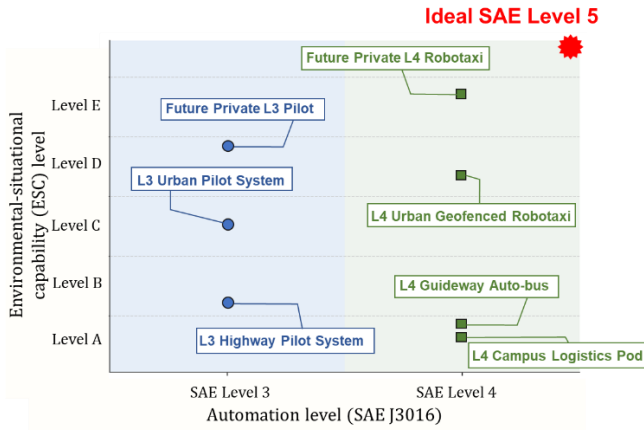


Fig. 7. The proposed ESC supplemental taxonomy is orthogonal to SAE automation levels and categorizes ADS use cases.

L4 Campus Logistics Pod, L4 Guideway Auto-bus, and L4 Urban Geofenced Robotaxi), alongside two envisioned future milestones (Future Private L3 Pilot and Future Private L4 Robotaxi). These use-case assignments are ESC classifications inferred from the specified environmental and operating conditions, rather than test-demonstrated capability ratings. The latter require the ADS to satisfy predefined performance criteria in matched validation scenarios. A more detailed elaboration is provided in the Section 5.2. Finally, as indicated by the upper-right red dot in the matrix, advancements in both ESC capabilities and SAE automation levels enable systems to steadily converge toward an ideal Level 5. While this framework does not provide a strict definition of Level 5, it aims to systematically illuminate the pathway to fully autonomous driving via the proposed taxonomy.

Table 1. Illustrative ESC classification of representative ADS use cases

Case number	System representative	ESC label	ESC score	ESC level
1	L3 Highway Pilot System	FMF (8, 6, 7)	393	L3-B
Scenario description: Deployed exclusively within closed-access, physically segregated intercity highway or expressway networks. The ODD is strictly confined to highly structured main carriageways devoid of at-grade intersections, pedestrian crossings, or non-motorized traffic. As an SAE Level 3 system, human drivers are not required to monitor the environment continuously but must remain receptive to issued TOR and perform the DDT fallback when approaching the spatial or environmental boundaries of the ODD.				
2	L3 Urban Pilot System	MMM (6, 5, 5)	579	L3-C
Scenario description: Operated within typical urban arterial and sub-arterial road networks characterized by complex intersections and mixed traffic zones. The ODD encompasses standard city driving but strictly requires human drivers to perform the DDT fallback when the system encounters unresolvable urban edge cases or traverses beyond defined operational boundaries.				
3	Future Private L3 Pilot	MSS (4, 3, 3)	693	L3-D
Scenario description: Operated within extremely complex wide-area urban road networks and unstructured urban-rural fringes. The ODD extends beyond standardized city grids to include poorly maintained				

suburban routes and weakly structured environments. As an advanced Level 3 system, it significantly broadens its operational envelope compared to conventional baseline systems. However, it still strictly mandates a human driver to execute the DDT fallback when traversing into absolute off-road terrain or encountering system-breaking limits.

4	L4 Campus Logistics Pod	FFF (8, 8, 9)	153	L4-A
Scenario description: Dedicated to highly structured and physically geofenced environments such as closed industrial parks. The ODD is strictly confined to predefined and repetitive low-speed logistics routes. As an SAE Level 4 system, it performs the complete DDT and DDT fallback autonomously without any expectation of human intervention within its designated operational boundaries.				
5	L4 Guideway Auto-bus	FMF (9, 6, 8)	297	L4-A
Scenario description: Deployed on dedicated autonomous driving tracks or physically isolated lanes. The system performs the complete DDT autonomously without human drivers within its predefined infrastructure.				
6	L4 Urban Geofenced Robotaxi	MMM (6, 5, 4)	609	L4-C
Scenario description: Deployed within the strict geofences of designated core and sub-core urban areas. The system performs the complete DDT autonomously without human intervention within its predefined operational boundaries.				
7	Future Private L4 System	SSS (3, 2, 2)	717	L4-E
Scenario description: Deployed as an open-road Level 4 system possessing cross-city and all-weather point-to-point autonomous driving capabilities. The ODD covers complex urban core networks and extends to edge areas including unmarked mountainous roads, unpaved rural dirt roads, and urban village alleys. The system autonomously performs the complete DDT and fallback. It reaches its operational boundaries when confronting completely topology-free natural wilderness, extreme weather exceeding physical limits, or chaotic scenarios with collapsed traffic order.				

Note: The domain scores and results are illustrative applications of the proposed framework under the equal-weight baseline. They are assigned according to the stated archetypal operating conditions to demonstrate the tier-mapping procedure. They are not empirically validated product ratings and were not derived from formal expert elicitation, inter-rater reliability assessment, independent ADS testing, or certification.

Furthermore, to illustrate the practical application of our proposed methodology, we select a representative system from the table—the L3 Urban Pilot System—to demonstrate the complete, step-by-step classification process.

Step 1: SAE level and scenario definition

Based on the hierarchical metrics defined in Section 3, illustrative scores are assigned the three core environmental dimensions on a 0–9 scale, where higher scores indicate more favorable and less complex conditions.

Step 2: Scoring and label assessment

- Built environment (Score: 6, Moderate): The system demonstrates real-time topological inference capabilities, enabling it to navigate

standard urban grids and complex physical intersections. It leverages partial V2X communication and high-definition (HD) map assistance to enhance perception. However, this external support remains unstable and is insufficient to guarantee absolute structural awareness. The system can effectively recognize region-specific traffic signs and signals within a designated city, as well as partially degraded lane markings. Conversely, severe structural anomalies breach its operational boundaries, as the system cannot be expected to reliably traverse construction work zones, shared-space roads lacking explicit curbs, or highly irregular geometric layouts such as multi-lane roundabouts.

- Natural environment (Score: 5, Moderate): The system exhibits a certain degree of perceptual robustness under conditions such as sunset glare, light showers, and urban night environments equipped with infrastructural illumination. However, sensor performance experiences severe degradation when encountering extreme scenarios, such as heavy rain, dense fog, snow-covered roads that completely obscure lane markings and physical boundaries, and pitch-black, unlit environments. Therefore, this level accounts for the majority of mild visual perturbations typically encountered in temperate urban environments, while strictly excluding sensor-degrading extremes.
- Traffic conditions (Score: 5, Moderate): The system adapts to mixed traffic flows exhibiting moderate heterogeneity within urban road networks. It properly handles conventional interactions involving diverse traffic participants and maintains a stable and safe kinematic profile during regular fluctuations in traffic speed. However, when confronted with highly adversarial dynamic interactions or scenarios characterized by high behavioral uncertainty among traffic participants (e.g., narrow, high-density streets lacking explicit right-of-way), the system may trigger a takeover request (TOR). Consequently, this traffic environment is accurately rated as moderate in complexity.

Step 3: ESC score calculation and classification.

The ESC_{label} provides a preliminary grouping. For the L3 Urban Pilot System, the assigned label of MMM initially categorizes the operational environment into preliminary Level C, based on the semantic tree structure. Assuming equal weights of the factors across all three environmental dimensions ($\alpha_i = 1$ each) for this theoretical demonstration, the ESC_{score} is calculated as follows:

- L3 Urban Pilot System: $ESC_{score} = 729 - (6 \times 5 \times 5) = 579$.

Finally, the calculated score of 579 falls squarely within the mathematical interval defined for subclass C (534 to 617). It is eventually classified as an L3-C system. Contrastively, for the Case 5 (L4 Guideway Auto-bus) in Table 1, its calculated score of 297 is mathematically lower than the worst-case boundary threshold for Level A (386). Because the extreme advantages in the built environment (9) and traffic conditions (8) compensate for the moderate natural environment (6) at a systemic level, the quantitative score overrides the FMF semantic constraint. Consequently, it is classified as an L4-A system.

This approach enables developers to position their products within a two-dimensional matrix defined by the achieved automation level, and the operational safety bounds under adverse environmental conditions. By mathematically resolving boundary ambiguities, this framework

supports companies in refining product categorization and encourages diversified product development while mitigating potential regulatory disputes.

5 Discussions

5.1. Framework evaluation and capability characterization

The proposed ESC taxonomy bridges the gap between merely ODD-related element lists and concrete ADS capability requirements. Serving as a practical and theoretical supplement to the SAE standards, it facilitates capability quantification for ADS during real-world deployment. Fundamentally, the ESC framework introduces a performance-based evaluation paradigm that focuses on deterministic capability constraints rather than internal algorithmic complexity. As autonomous driving undergoes rapid commercialization, this structured environmental grading system provides a consistent baseline to objectively benchmark and compare diverse industry solutions.

Simultaneously, this framework aims to provide actionable quantitative boundaries to support the SOTIF (ISO, 2022). This systematic approach can assist system developers in parameterizing and isolating known-unsafe compound environmental scenarios within the SOTIF framework, thereby providing a structured reference for constructing safety-critical scenarios and establishing safety performance boundaries (Song et al., 2022). While the framework does not identify unknown-unsafe scenarios directly, its contribution here is indirect. The declared ESC profile makes the boundary of the parameterized region explicit, and this operates in two ways within a SOTIF-oriented workflow. During testing, it organizes scenario coverage within the claimed ODD and links environmental classifications to validation evidence, allowing validation scenarios to be concentrated near that boundary so that previously unidentified unsafe conditions can be exposed before deployment. During field operation, conditions encountered at or beyond the boundary can be located on specific environmental domains and sub-metrics, converted from unknown-unsafe into known-unsafe conditions within the existing classification structure, and used to update the ODD declaration or the scoring criteria.

Furthermore, the framework incorporates intelligent driving assistance infrastructure into metrics, which is typically absent in conventional ODD definitions. Although comprehensive digital infrastructure can significantly elevate the upper bound of operational capabilities for autonomous driving systems (Zha et al., 2024; Z. Xu et al., 2025; Mo et al., 2024), its utility is strictly constrained by the spatial extent of such deployments. This consequently reveals that systems heavily reliant on external infrastructural support still exhibit inherent limitations regarding stand-alone intelligence and overall autonomy (S. Chen et al., 2023).

5.2. Practical deployment, commercialization, and regulation

The proposed ESC supplemental taxonomy demonstrates significant potential for facilitating commercial deployment and managing consumer expectations. Currently, relying solely on the existing SAE levels, consumers often find it challenging to fully differentiate the actual operational capabilities among various automated driving products. Consequently, consumers' current Willingness to Pay (WTP) for advanced autonomous driving systems remains generally low

(Morita and Managi, 2020).

In practice, the ESC score should be interpreted together with its three-domain and nine-sub-metric profile. A manufacturer-declared ODD is used to define the claimed ESC profile and the corresponding testing scope, but does not itself constitute evidence of demonstrated capability. Manufacturers can use this profile to identify capability limitations, compare system versions, and determine which environmental conditions require additional testing before ODD expansion. Regulators and testing agencies can classify validation scenarios using the same ESC metrics and derive a tailored test suite for the claimed ESC profile. Such a suite may include:

- representative scenarios within the claimed tier,
- scenarios at its upper boundary,
- challenge scenarios immediately beyond that boundary.

This provides a structured basis for capability-verification programs, jurisdiction-specific minimum requirements, and reassessment following major software updates or ODD expansion. For type-approval applications, the manufacturer could include the claimed ESC profile and its corresponding ODD in the technical documentation. The approval authority or designated technical service could then evaluate the submitted evidence using the validation procedure described above. A claimed ESC profile can be regarded as test-demonstrated only after the ADS satisfies predefined performance criteria across the matched scenario set.

A further consideration in application of the framework is cross-regional comparability. Jurisdiction-specific certification thresholds should be empirically calibrated and determined by regulators according to local risk tolerance. Such jurisdiction-specific calibration is necessary because infrastructure standards, road characteristics, climatic exposure, traffic composition, and driving behavior can differ substantially across regions (Edelmann et al., 2021; Sun et al., 2023). A region with severe winters may require a high natural-environment score, a dense metropolitan area may prioritize the traffic-condition dimension, and a rural or less developed region may set its requirement on the built-environment dimension. The scoring criteria themselves remain identical across regions, so an ESC label carries the same meaning wherever it is issued, and comparability is preserved at the level of the label while the pass or fail decision remains a local policy choice. At its current stage, the framework supports comparative benchmarking, capability-profile communication, and test design, while formal certification thresholds require further empirical calibration.

Following such empirical calibration and regulatory adoption, a standardized capability rating based on the ESC framework is expected to alleviate the inherent information asymmetry among manufacturers, consumers, and regulatory bodies. For example, it could provide regulatory authorities with a quantified reference dimension when exploring market access standards and conducting routine supervision. Based on the ESC taxonomy, the industry might consider establishing a standardized operational permit or product capability label for ADS. In practice, vehicles would undergo standardized certification prior to market entry to receive an ESC sticker—conceptually similar to a New Car Assessment Program (NCAP) (NHTSA, 2024) crash safety rating or an energy efficiency label. ESC could be implemented as a supplementary consumer-information layer. Following standardized scenario testing, the published result should include the SAE level, the test-demonstrated ESC tier, the three-domain profile, and the

associated ODD limitations. Consequently, instead of broadly marketing a vehicle merely as SAE Level 4, it could be precisely annotated as, for example, L4-C. To avoid potential misinterpretation, ESC labels should always be communicated together with their underlying capability profile rather than as an isolated tier. For example, an ADS should be represented as L4-B (6, 5, 7), rather than only L4-B, where the three values correspond to the built environment, natural environment, and traffic condition dimensions. Such a representation prevents ESC from being interpreted as a single safety ranking and preserves the distinction between environmental capability and SAE-defined fallback responsibility. This would help differentiate a system restricted to operating in restricted guideways under clear weather from one capable of navigating complex chaotic urban traffic in adverse weather conditions, for example.

Such an intuitive labeling approach helps translate broad system descriptions into more transparent reference indicators, thereby assisting consumers in making purchasing decisions that better align with their actual needs, usage preferences, and budget constraints. For example, urban residents might place a higher premium on a system's ability to handle high-density traffic flow to assist with inner-city commuting. Conversely, consumers with long-distance commutes or those residing in suburban or rural areas—whose broader geographical range exposes them to a higher exposure to adverse weather and unstructured roads—might be more inclined to select systems with superior adaptability to natural environments and degraded road conditions. Overall, enhancing system transparency through the introduction of a standardized capability evaluation framework may serve as an effective pathway for manufacturers to steadily maintain consumer trust.

Although the ESC rating system provides a quantitative baseline, its practical implementation within judicial and regulatory frameworks faces complex challenges. The primary challenge lies in standardizing the ESC certification process and accurately aligning the real-time environmental complexity at the exact moment of a crash with the vehicle's nominal ESC label. For example, if a vehicle certified for moderate traffic complexity (M, 6) experiences a collision caused by a sudden pedestrian intrusion amidst otherwise normal traffic, it creates a highly contentious gray area: should the crash be attributed to a system capability failure, or a transient environmental perturbation? Therefore, integrating this framework into future legal and insurance systems requires not only establishing standardized data acquisition and forensic verification mechanisms to accurately reconstruct the environmental complexity at the time of the crash, but also formulating clear legal adjudication guidelines regarding dynamic environmental shifts and liability attribution boundaries.

From a macro-level traffic management perspective, the ESC taxonomy shows significant potential as a supplementary reference for future autonomous driving legislation and traffic law enforcement. Traffic authorities could explore integrating the ESC framework into Intelligent Transportation Systems (ITS) to conceptually pilot dynamic right-of-way allocation mechanisms driven by environmental complexity. For example, upon forecasts of sudden extreme rainstorms or widespread road icing, ITS could dynamically adjust or recommend restriction of road access to L4-A and L4-B automated vehicles temporarily while recommending uninterrupted operations of vehicles whose tested operating profiles cover the prevailing conditions. Furthermore, for L3 systems requiring human fallback or teleoperation,

the ESC standards could potentially be incorporated into AV operator licensing manuals as a reference framework, facilitating the development of differentiated emergency takeover training requirements tailored to specific environmental tiers.

5.3. Limitations and future work

A primary limitation of the proposed ESC framework is that identical performance scores may correspond to vastly distinct combinations of environmental conditions. For example, two vehicles both rated as L4-B might exhibit entirely disparate capability profiles: a radar-centric vehicle might excel in mitigating adverse conditions such as glare but rely heavily on well-equipped infrastructure; conversely, a vision-centric vehicle might navigate degraded infrastructure with ease but strictly require clear weather conditions. Due to this inherent heterogeneity, the ESC taxonomy cannot replace traditional ODD definitions. To ensure safety, high-level ESC labels must be tightly coupled with detailed, verifiable, and machine-readable ODD formats (Ito, 2021). Furthermore, the specific coupling relationship between the ESC rating and the underlying ODD must be clearly and transparently communicated to consumers through official documentation or promotional materials.

The current preliminary framework provides a structured pathway for linking environmental assessment and institutional testing evidence to ESC assignments. However, this framework does not replace the need for empirically validated weighting and score calibration. Although the current multiplicative approach effectively models the cumulative effect of environmental complexity, deriving these parameters from subjective expert judgments may fail to fully capture the complex, non-linear interdependencies between different factors. Given the fundamentally distinct nature of various failure mechanisms and the inherent biases associated with expert evaluations, determining the sensitivity and coupled weighting among environmental metrics remains an open research gap. This necessitates extensive expert surveys to further quantify the relative importance of these factors, and the development of more sophisticated composite evaluation functions, such as the comprehensive safety metrics for autonomous system perception proposed by Volk et al. (2020). The equal-width subdivision adopted in Section 4.1.1 should likewise be interpreted as a transparent design convention. Accordingly, the current version should be interpreted as an equal-weight preliminary framework, while empirical calibration of domain-specific weights remains a priority for future work.

To address these limitations, future research should prioritize empirical calibration of both the weighting relationships and the ordinal score intervals. First, establishing standardized terminology and definitions for typical scenarios, operational conditions, and unforeseen events is crucial for achieving accurate system classification and subsequent impact assessment. Second, to overcome these limitations and transition from subjective scoring to a robust quantitative paradigm, future studies must comprehensively integrate objective, multi-source datasets. Evaluations must leverage empirical data collected from restricted proving grounds and public road trials to benchmark operational performance, alongside the systematic mining of large-scale disengagement reports (e.g., the California DMV). Leveraging emerging techniques that utilize pre-trained large language models to analyze and extract insights from these reports (Ansarinejad

et al., 2025b), researchers can map the extracted failure patterns onto the ESC metrics to determine their relative impact weights. Collectively, these efforts will enable a rigorous, data-driven calibration of the entire ESC framework.

To enhance the evaluation of the deployment capability and reliability of ADS, future research should also incorporate more measurable dimensions, similar to the tri-axis framework proposed by H. M. Huang et al. (2005). The integration of LLM-driven human-like reasoning and other cognitive dimensions suggests that future evaluation standards should evolve into multi-dimensional radar charts or multi-axis architectures. Such a multi-dimensional assessment paradigm can decouple complex system performance into independent or synergistic capability benchmarks, such as cognitive human-like decision-making capabilities and system safety redundancy. Furthermore, with the widespread deployment of V2X communications, cyber-physical security and adversarial threats—such as physical adversarial patches, adversarial camouflage, and sensor spoofing attacks—are emerging as critical dimensions of system vulnerability (Girdhar et al., 2023; Duo et al., 2022). Therefore, future ODD definitions and ESC evaluations should incorporate a dedicated digital robustness or cyber resilience dimension to guarantee operational safety at higher automation levels.

6 Conclusion

This study introduces a proposed ESC framework to address the critical quantization gap inherent in current ODD definitions and SAE automation taxonomies. While the transition from SAE Level 3 to Level 4 represents a fundamental transfer of fallback responsibility, the traditional taxonomy inherently fails to capture the corresponding environmental robustness and intelligence required for actual real-world deployment. By systematically decoupling the operational environment into built, natural, and traffic dimensions, the proposed ESC taxonomy transforms fragmented physical parameters into a structured, performance-based capability grading system. Leveraging a multiplicative aggregation model alongside a two-stage hierarchical classification strategy, we established an orthogonal capability matrix that stratifies L3 and L4 systems into five progressive operational tiers (A through E). This multi-dimensional assessment paradigm attempts to further explore the actual operational performance of systems across diversified ODD baselines, building upon the existing automation grading. Ultimately, the ESC framework aims to offer a standardized and quantifiable baseline reference, seeking to facilitate a connection between macro-level capability claims and micro-level engineering constraints. We hope this approach can provide a useful methodological perspective for enhancing system transparency, supporting SOTIF-oriented scenario structuring, and assisting in the safe commercialization and regulatory oversight of advanced autonomous vehicles.

Author contributions

Chengxi Hu: Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Sikai Chen:** Validation, Writing – review & editing, Supervision, Methodology. **Samuel Labi:** Writing – review & editing, Methodology. **Hongliang Ding:** Writing – review & editing,

Methodology. **Hyungchul Chung**: Writing – review & editing, Methodology, Supervision. **Tiantian Chen**: Writing – review & editing, Supervision, Methodology, Conceptualization, Funding acquisition.

Acknowledgements

This work was supported by AI+ KCIR Global Research Center from the National Research Foundation of Korea ERC Project (RS-2025-02223612).

Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article.

Just Accepted

References

- Adedjouma, M., Botella, B., Ibanez-Guzman, J., Mantissa, K., Proum, C. M., Smaoui, A., 2024. Defining operational design domain for autonomous systems: A domain-agnostic and risk-based approach. In: Proceedings of the 2024 19th Annual System of Systems Engineering Conference (SoSE), 166-171.
- Agrawal, S., Peeta, S., 2021. Evaluating the impacts of situational awareness and mental stress on takeover performance under conditional automation. *Transp. Res. Part F Traffic Psychol. Behav.* 83, 210-225.
- Aldoski, Z. N., Koren, C., 2023. Impact of traffic sign diversity on autonomous vehicles: a literature review. *Period. Polytech. Transp. Eng.* 51, 338-350.
- Ali, M. L., Zhang, Z., 2024. The YOLO framework: A comprehensive review of evolution, applications, and benchmarks in object detection. *Comput.* 13, 336.
- Aloufi, N., Alnori, A., Basuhail, A., 2024. Enhancing autonomous vehicle perception in adverse weather: A multi objectives model for integrated weather classification and object detection. *Electron.* 13, 3063.
- Alqahtani, T., 2025. Recent trends in the public acceptance of autonomous vehicles: A review. *Veh.* 7, 45.
- Ansarinejad, M., Ansarinejad, K., Lu, P., Huang, Y., Tolliver, D., 2025a. Autonomous vehicles in rural areas: A review of challenges, opportunities, and solutions. *Appl. Sci.* 15, 4195.
- Ansarinejad, M., Gaweesh, S. M., Ahmed, M. M., 2025b. Assessing the efficacy of pre-trained large language models in analyzing autonomous vehicle field test disengagements. *Accid. Anal. Prev.* 220, 108178.
- Antsaklis, P., 2020. Autonomy and metrics of autonomy. *Annu. Rev. Control.* 49, 15-26.
- Araujo, H., Mousavi, M. R., Varshosaz, M., 2023. Testing, validation, and verification of robotic and autonomous systems: a systematic review. *ACM Trans. Softw. Eng. Methodol.* 32, 1-61.
- Arvin, R., Khattak, A. J., Kamrani, M., Rio-Torres, J., 2020. Safety evaluation of connected and automated vehicles in mixed traffic with conventional vehicles at intersections. *J. Intell. Transp. Syst.* 25, 170-187.
- Asvadi, A., Garrote, L., Premebida, C., Peixoto, P., Nunes, U. J., 2018. Multimodal vehicle detection: fusing 3D-LIDAR and color camera data. *Pattern Recognit. Lett.* 115, 20-29.
- Aung, N. H. H., Sangwongngam, P., Jintamethasawat, R., Shah, S., Wuttisittikulkij, L., 2024. A review of lidar-based 3d object detection via deep learning approaches towards robust connected and autonomous vehicles. *IEEE Trans. Intell. Veh.* 10, 526-547.
- Badue, C., Guidolini, R., Carneiro, R. V., Azevedo, P., Cardoso, V. B., Forechi, A., Jesus, L., Berriel, R., Paixão, T. M., Mutz, F., Veronese, L. P., Oliveira-Santos, T., De Souza, A. F., 2021. Self-driving cars: A survey. *Expert Syst. Appl.* 165, 113816.
- Betz, J., Lutwitz, M., Peters, S., 2024. A new taxonomy for automated driving: Structuring applications based on their operational design domain, level of automation and automation readiness. In: Proceedings of the 2024 IEEE Intelligent Vehicles Symposium (IV), 1-7.
- Bilik, I., Longman, O., Villeval, S., Tabrikian, J., 2019. The rise of radar for autonomous vehicles: Signal processing solutions and future research directions. *IEEE Signal Process. Mag.* 36, 20-31.
- Zheng, B. S., Chen, J. Y., Zhao, R. M., Huang, X. M., 2019. Skid resistance demands of asphalt pavement during the braking process of autonomous vehicles. In: Proceedings of MATEC Web of Conferences (Vol. 275), 04002.
- Carreras, A., Daura, X., Erhart, J., Ruehrup, S., 2018. Road infrastructure support levels for automated driving. In: Proceedings of the 25th ITS World Congress, 12-20.
- Carruth, D. W., Walden, C. T., Goodin, C., Fuller, S. C., 2022. Challenges in low infrastructure and off-road automated driving. In: Proceedings of the 2022 Fifth International Conference on Connected and Autonomous Driving (MetroCAD), 13-20.
- Cappellari, P., Weber, B. S., 2022. An analysis of the New York City traffic volume, vehicle collisions, and safety under COVID-19. *Journal of safety research*, 83, 57-65.
- Chen, H., Wen, Y., Zhu, M., Huang, Y., Xiao, C., Wei, T., Hahn, A., 2021. From automation system to autonomous system: An architecture perspective. *J. Mar. Sci. Eng.* 9, 645.
- Chen, H., Zhao, X., Li, H., Gong, J., Fu, Q., 2024. Predicting driver's takeover time based on individual characteristics, external environment, and situation awareness. *Accid. Anal. Prev.* 203, 107601.
- Chen, L., Wu, P., Chitta, K., Jaeger, B., Geiger, A., Li, H., 2024. End-to-end autonomous driving: Challenges and frontiers. *IEEE Trans. Pattern Anal. Mach. Intell.* 46(12), 10164-10183.
- Chen, S., Zong, S., Chen, T., Huang, Z., Chen, Y., Labi, S., 2023. A taxonomy for autonomous vehicles considering ambient road infrastructure. *Sustainability.* 15, 11258.
- Chen, Y., Chen, S., Zhang, T., Zhang, S., Zheng, N., 2018. Autonomous vehicle testing and validation platform: Integrated simulation system with hardware in the loop. In: Proceedings of the 2018 IEEE Intelligent Vehicles Symposium (IV), 949-956.
- Chen, Z., Zhang, Z., Su, Q., Yang, K., Wu, Y., He, L., Tang, X., 2025. Object detection for autonomous vehicles under adverse weather conditions. *Expert Syst. Appl.* 128994.
- Czarniecki, K., 2018. Operational design domain for automated driving systems-taxonomy of basic terms. Waterloo Intelligent Systems Engineering (WISE) Lab, University of Waterloo, Canada, 1.
- Dickmanns, E. D., Graefe, V., 1988. Dynamic monocular machine vision. *Mach. Vis. Appl.* 1, 223-240.
- Duo, W., Zhou, M., Abusorrah, A., 2022. A survey of cyberattacks on cyber physical systems: Recent advances and challenges. *IEEE/CAA J. Autom. Sin.* 9, 784-800.
- Edelmann, A., Stümper, S., Petzoldt, T., 2021. Cross-cultural differences in the acceptance of decisions of automated vehicles. *Appl. Ergon.* 92, 103346.
- El-Demerdash, S. M., ORABY, W., Selim, A. M., 2005. Four-wheel steering vehicle stability under random. In: Proceedings of the Total Vehicle Technology: Finding the Radical, Implementing the Practical (3rd International Conference), 15.
- Fan, L., Wang, J., Chang, Y., Li, Y., Wang, Y., Cao, D., 2024. 4D mmWave radar for autonomous driving perception: A comprehensive survey. *IEEE Trans. Intell. Veh.* 9, 4606-4620.
- Federal Highway Administration (FHWA), 2025. How Do Weather Events Affect Roads? Road Weather Management Program, U.S. Department of Transportation.
- Formosa, N., Quddus, M., Man, C. K., Singh, M. K., Morton, C., Masera, C. B., 2024. Evaluating the impact of lane marking quality on the operation of autonomous vehicles. *J. Transp. Eng. Part A Syst.* 150,

- Galvao, L. G., Abbod, M., Kalganova, T., Palade, V., Huda, M. N., 2021. Pedestrian and vehicle detection in autonomous vehicle perception systems—A review. *Sensors*. 21, 7267.
- García, A., Camacho-Torregrosa, F. J., 2020. Influence of lane width on semi-autonomous vehicle performance. *Transp. Res. Rec.* 2674, 279-286.
- Gautam, S., Kumar, A., 2023. Image-based automatic traffic lights detection system for autonomous cars: a review. *Multimed. Tools Appl.* 82, 26135-26182.
- Giannaros, A., Karras, A., Theodorakopoulos, L., Karras, C., Kranias, P., Schizas, N., Kalogeratos, G., Tsoilis, D., 2023. Autonomous vehicles: Sophisticated attacks, safety issues, challenges, open topics, blockchain, and future directions. *J. Cybersecur. Priv.* 3, 493-543.
- Girdhar, M., Hong, J., Moore, J., 2023. Cybersecurity of autonomous vehicles: A systematic literature review of adversarial attacks and defense models. *IEEE Open J. Veh. Technol.* 4, 417-437.
- Gyawali, S., Xu, S., Qian, Y., Hu, R. Q., 2020. Challenges and solutions for cellular based V2X communications. *IEEE Commun. Surv. Tutor.* 23, 222-255.
- Hnewa, M., Radha, H., 2020. Object detection under rainy conditions for autonomous vehicles: A review of state-of-the-art and emerging techniques. *IEEE Signal Process. Mag.* 38, 53-67.
- Hopkins, D., Schwanen, T., 2021. Talking about automated vehicles: What do levels of automation do? *Technol. Soc.* 64, 101488.
- Houben, S., Stallkamp, J., Salmen, J., Schlipsing, M., Igel, C., 2013. Detection of traffic signs in real-world images: The German Traffic Sign Detection Benchmark. In: Proceedings of the 2013 international joint conference on neural networks (IJCNN), 1-8.
- Howard, A., Sandler, M., Chu, G., Chen, L. C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q. V., Adam, H., 2019. Searching for mobilenetv3. In: Proceedings of the IEEE/CVF international conference on computer vision, 1314-1324.
- Hrabia, C. E., Masuch, N., Albayrak, S., 2015. A metrics framework for quantifying autonomy in complex systems. In: Proceedings of the German Conference on Multiagent System Technologies, 22-41.
- Huang, H. M., Pavsek, K., Novak, B., Albus, J., Messin, E., 2005. A framework for autonomy levels for unmanned systems (ALFUS). In: Proceedings of the AUVSI's unmanned systems North America, 849-863.
- Hussain, R., Zeadally, S., 2018. Autonomous cars: Research results, issues, and future challenges. *IEEE Commun. Surv. Tutor.* 21, 1275-1313.
- ISO, 2022. Road Vehicles — Safety of the Intended Functionality (SOTIF). ISO 21448.
- ISO, 2023. Road Vehicles - Test scenarios for automated driving systems - Specification for operational design domain. ISO 34503.
- Ito, M., 2021. ODD description methods for automated driving vehicle and verifiability for safety. *J. Univers. Comput. Sci.* 27.
- Jameel, A. K., Evdorides, H., 2016. An investigation for an all-encompassing iRAP road Star Rating index. In: Proceedings of the Functional Pavement Design, 1607-1616.
- Janai, J., Güney, F., Behl, A., Geiger, A., 2020. Computer vision for autonomous vehicles: Problems, datasets and state of the art. *Found. Trends Comput. Graph. Vis.* 12, 1-308.
- Kamtam, S. B., Lu, Q., Bouali, F., Haas, O. C., Birrell, S., 2024. Network latency in teleoperation of connected and autonomous vehicles: A review of trends, challenges, and mitigation strategies. *Sensors*. 24, 3957.
- Kim, I., Kang, D., Jeong, H., Lee, S., Yun, I., 2023. Method of evaluating multiple scenarios in a single simulation run for automated vehicle assessment. *Sensors*. 23, 8271.
- Kim, S., Novakazi, F., Shi, E., Harms, I. M., Oviedo-Trespalcacios, O., 2026. A Taxonomic Odyssey: Evolution, Criticisms, and Future Directions of Driving Automation Taxonomies—The Case of SAE J3016. *Transp. Res. Interdiscip. Perspect.* 36, 101858.
- Koopman, P., Fratrick, F., 2019. How many operational design domains, objects, and events? *Safeai@aaai*. 4.
- Koopman, P., Wagner, M., 2017. Autonomous vehicle safety: An interdisciplinary Challenge. *IEEE Intell. Transp. Syst. Mag.* 9, 90-96.
- Li, Y., Ibanez-Guzman, J., 2020. Lidar for autonomous driving: The principles, challenges, and trends for automotive lidar and perception systems. *IEEE Signal Process. Mag.* 37, 50-61.
- Li, H., Wang, X., Zhang, Z., Zhu, W., 2025. Out-of-distribution generalization on graphs: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* 47, 10490 – 10512.
- Li, T., Hu, C., Sze, N. N., Sun, Z., Ding, H., Chen, T., 2025. Modelling pedestrian-vehicle interaction behaviours at non-signalised crosswalks using a game theory approach incorporating the risk perception. *Transportmetrica A Transp. Sci.* 1-31.
- Li, T., Sun, Z., Zhou, M., Sze, N. N., Zhou, Y., 2026. Modeling bounded rationality in pedestrian-vehicle interactions at non-signalized crosswalks: A game theoretic quantal response equilibrium approach. *Accid. Anal. Prev.* 231, 108502.
- Lin, H., Parsi, A., Mullins, D., Horgan, J., Ward, E., Eising, C., Denny, P., Deegan, B., Glavin, M., Jones, E., 2024. A study on data selection for object detection in various lighting conditions for autonomous vehicles. *J. Imaging*. 10, 153.
- Lin, Z., Tian, Z., Lan, J., Zhang, Q., Ye, Z., Zhuang, H., Zhao, X., 2025. A conflicts-free, speed-lossless kan-based reinforcement learning decision system for interactive driving in roundabouts. *IEEE Trans. Intell. Transp. Syst.* 26, 16377-16390.
- Lowe, E., Guvenc, L., 2024. General Obstacle Avoidance Capability Assessment for Autonomous Vehicles. *Electron.* 13, 4901.
- Lv, C., Cao, D., Zhao, Y., Auger, D. J., Sullman, M., Wang, H., Dutka, L. M., Skrypchuk, L., Mouzakitis, A., 2017. Analysis of autopilot disengagements occurring during autonomous vehicle testing. *IEEE/CAA J. Autom. Sin.* 5, 58-68.
- Ma, C., Li, D., 2023. A review of vehicle lane change research. *Phys. A Stat. Mech. Appl.* 626, 129060.
- Manivasakan, H., Kalra, R., O'Hern, S., Fang, Y., Xi, Y., Zheng, N., 2021. Infrastructure requirement for autonomous vehicle integration for future urban and suburban roads—Current practice and a case study of Melbourne, Australia. *Transp. Res. Part A Policy Pract.* 152, 36-53.
- Matin, A., Dia, H., 2022. Impacts of connected and automated vehicles on road safety and efficiency: A systematic literature review. *IEEE Trans. Intell. Transp. Syst.* 24, 2705-2736.
- Mehlhorn, M. A., Richter, A., Shardt, Y. A., 2023. Ruling the operational boundaries: A survey on operational design domains of autonomous driving systems. *IFAC-PapersOnLine*. 56, 2202-2213.
- Mendiboure, L., Benzagouta, M. L., Gruyer, D., Sylla, T., Adedjouma, M., Hedhli, A., 2023. Operational design domain for automated driving systems: Taxonomy definition and application. In: Proceedings of the 2023 IEEE Intelligent Vehicles Symposium (IV), 1-6.
- Miaou, S. P., Lum, H., 1993. Modeling vehicle accidents and highway geometric design relationships. *Accid. Anal. Prev.* 25, 689-709.
- Mo, Y., Vijay, R., Rufus, R., Boer, N. D., Kim, J., Yu, M., 2024. Enhanced

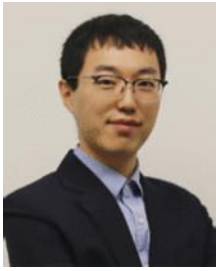
- perception for autonomous vehicles at obstructed intersections: An implementation of vehicle to infrastructure (v2i) collaboration. *Sensors*. 24, 936.
- Morando, M. M., Tian, Q., Truong, L. T., Vu, H. L., 2018. Studying the safety impact of autonomous vehicles using simulation-based surrogate safety measures. *J. Adv. Transp.* 6135183.
- Morita, T., Managi, S., 2020. Autonomous vehicles: Willingness to pay and the social dilemma. *Transp. Res. Part C Emerg. Technol.* 119, 102748.
- Muhammad, K., Hussain, T., Ullah, H., Del Ser, J., Rezaei, M., Kumar, N., Hijji, M., Bellavista, P., De Albuquerque, V. H. C., 2022. Vision-based semantic segmentation in scene understanding for autonomous driving: Recent achievements, challenges, and outlooks. *IEEE Trans. Intell. Transp. Syst.* 23, 22694-22715.
- Naiseh, M., Clark, J., Akarsu, T., Hanoch, Y., Brito, M., Wald, M., Webster, T., Shukla, P., 2025. Trust, risk perception, and intention to use autonomous vehicles: an interdisciplinary bibliometric review. *AI Soc.* 40, 1091-1111.
- NHTSA, 2016. Federal automated vehicles policy - Accelerating the Next Revolution In Roadway Safety. NHTSA 12507-091216-v9.
- NHTSA, 2017. Automated driving systems 2.0: a vision for safety. NHTSA 13069a-090617-v9a.
- NHTSA, 2019. Rural/urban comparison of traffic fatalities: 2017 data. DOT HS 812 74.
- NHTSA, 2020. Framework for Automated Driving System Safety. NHTSA-2020-0106.
- NHTSA, 2021. Incident Reporting for Automated Driving Systems (ADS) and Level 2 Advanced Driver Assistance Systems (ADAS).
- NHTSA, 2024. New Car Assessment Program Final Decision Notice—Crashworthiness Pedestrian Protection. NHTSA-2024-0078.
- Othman, K., 2021. Impact of autonomous vehicles on the physical infrastructure: Changes and challenges. *Designs*. 5, 40.
- Parker, J. I., Lee, J. D., McDonald, A., Rau, P., Philips, B., 2023. Do researchers use the levels of automation appropriately? How automation is discussed in research. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 1671-1677.
- Radu, A. D., Chen, L., Leffler, D., 2024. Autonomous Drivability: A Case Study of Swedish Roads. In: Proceedings of the KES-STs International Symposium, 13-23.
- Rajapaksha, U., Soheli, F., Laga, H., Diepeveen, D., Bennamoun, M., 2024. Deep learning-based depth estimation methods from monocular image and videos: A comprehensive survey. *ACM Comput. Surv.* 56, 1-51.
- Ravindran, R., Santora, M. J., Jamali, M. M., 2020. Multi-object detection and tracking, based on DNN, for autonomous vehicles: A review. *IEEE Sensors J.* 21, 5668-5677.
- Ruan, J., Cui, H., Huang, Y., Li, T., Wu, C., Zhang, K., 2023. A review of occluded objects detection in real complex scenarios for autonomous driving. *Green Energy Intell. Transp.* 2, 100092.
- SAE International, 2014. Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems. SAE-J3016-2014.
- SAE International, 2021. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. SAE-J3016-2021.
- Salvi, A., Weiss, G., Trapp, M., Oboril, F., Buerkle, C., 2022. Fuzzy interpretation of operational design domains in autonomous driving. In: Proceedings of the 2022 IEEE Intelligent Vehicles Symposium (IV), 1261-1267.
- Sana, F., Azad, N. L., Raahemifar, K., 2023. Autonomous vehicle decision-making and control in complex and unconventional scenarios—A review. *Machines*. 11, 676.
- Scholtes, M., Westhofen, L., Turner, L. R., Lotto, K., Schuldes, M., Weber, H., Neurohr, C., Bollmann, M., H., Körtke, F., Hiller, J., Hoss, M., Bock, J., Eckstein, L., 2021. 6-layer model for a structured description and categorization of urban traffic and environment. *IEEE Access*. 9, 59131-59147.
- Seppelt, B., Reimer, B., Russo, L., Mehler, B., Fisher, J., Friedman, D., 2018. Towards a human-centric taxonomy of automation types. White Paper, 1723.
- Serna, C. G., Ruichek, Y., 2018. Classification of traffic signs: The european dataset. *IEEE Access*. 6, 78136-78148.
- Shakeri, A., 2024. Formalization of operational domain and operational design domain for automated vehicles. In: Proceedings of the 2024 IEEE 24th International Conference on Software Quality, Reliability, and Security Companion (QRS-C), 990-997.
- Sheykhsfard, A., Haghighi, F., Das, S., Fountas, G., 2023. Evasive actions to prevent pedestrian collisions in varying space/time contexts in diverse urban and non-urban areas. *Accid. Anal. Prev.* 192, 107270.
- Shladover, S. E., 2018. Connected and automated vehicle systems: Introduction and overview. *J. Intell. Transp. Syst.* 22, 190-200.
- Silva, Ó., Cordera, R., González-González, E., Nogués, S., 2022. Environmental impacts of autonomous vehicles: A review of the scientific literature. *Sci. Total Environ.* 830, 154615.
- Song, Y., Chitturi, M.V., Noyce, D.A., 2022. Intersection Two-Vehicle Crash Scenario Specification for Automated Vehicle Safety Evaluation Using Sequence Analysis and Bayesian Networks. *Accid. Anal. Prev.* 176, 106814.
- Soteropoulos, A., Mitteregger, M., Berger, M., Zwirchmayr, J., 2020. Automated drivability: Toward an assessment of the spatial deployment of level 4 automated vehicles. *Transp. Res. Part A Policy Pract.* 136, 64-84.
- Soteropoulos, A., Berger, M., Mitteregger, M., 2021. Compatibility of automated vehicles in street spaces: Considerations for a sustainable implementation. *Sustainability*. 13, 2732.
- Stathopoulos, A., Karlaftis, M. G., 2003. A multivariate state space approach for urban traffic flow modeling and prediction. *Transp. Res. Part C Emerg. Technol.* 11, 121-135.
- Sun, C., Deng, Z., Chu, W., Li, S., Cao, D., 2021. Acclimatizing the operational design domain for autonomous driving systems. *IEEE Intell. Transp. Syst. Mag.* 14, 10-24.
- Sun, X., Jiang, Y., Burnett, G., Bai, J., Bai, R., 2023. A cross-cultural analysis of driving styles for future autonomous vehicles. *Adv. Des. Res.* 1, 71-77.
- Taeiagh, A., & Lim, H. S. M., 2019. Governing autonomous vehicles: emerging responses for safety, liability, privacy, cybersecurity, and industry risks. *Transp. Rev.* 39(1), 103-128.
- Tahir, N. U. A., Zhang, Z., Asim, M., Chen, J., ELAffendi, M., 2024. Object detection in autonomous vehicles under adverse weather: A review of traditional and deep learning approaches. *Algorithms*. 17, 103.
- Tan, J., Lyu, F., Li, L., Hu, F., Feng, T., Xu, F., Zhang, Z., Yao, R., Wang,

- L., 2024. Dynamic V2X perception from road-to-vehicle vision. *IEEE Trans. Intell. Veh.* 1-14.
- Tengilimoglu, O., Carsten, O., Wadud, Z., 2023. Implications of automated vehicles for physical road environment: A comprehensive review. *Transp. Res. Part E Logist. Transp. Rev.* 169, 102989.
- Vargas, J., Alsweiss, S., Toker, O., Razdan, R., Santos, J., 2021. An overview of autonomous vehicles sensors and their vulnerability to weather conditions. *Sensors*. 21, 5397.
- Volk, G., Gamerding, J., von Bernuth, A., Bringmann, O., 2020. A comprehensive safety metric to evaluate perception in autonomous systems. In: Proceedings of the 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), 1-8.
- Wallman, C. G., Åström, H., 2001. Friction measurement methods and the correlation between road friction and traffic safety: A literature review. VTI meddelande 911A.
- Wang, H., Huang, Y., Khajepour, A., Zhang, Y., Rasekhipour, Y., Cao, D., 2019. Crash mitigation in motion planning for autonomous vehicles. *IEEE Trans. Intell. Transp. Syst.* 20, 3313-3323.
- Wang, H., Li, J., Dong, H., 2025. A Review of Vision-Based Multi-Task Perception Research Methods for Autonomous Vehicles. *Sensors*. 25, 2611.
- Wang, K., Ma, Q., Shen, C., Lu, J., 2025. Application of uncertainty to out-of-distribution detection for autonomous driving perception safety. *IEEE Trans. Intell. Transp. Syst.* 26, 11276-11293.
- Wang, Y., Abd Rahman, A. H., Nor Rashid, F. A., Razali, M. K. M., 2024. Tackling Heterogeneous Light Detection and Ranging-Camera Alignment Challenges in Dynamic Environments: A Review for Object Detection. *Sensors*. 24, 7855.
- Wang, Y., He, J., Fan, L., Li, H., Chen, Y., & Zhang, Z., 2024. Driving into the future: Multiview visual forecasting and planning with world model for autonomous driving. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 14749-14759.
- Wei, W., Cheng, Y., He, J., Zhu, X., 2024. A review of small object detection based on deep learning. *Neural Comput. Appl.* 36, 6283-6303.
- Xu, Z., Jiang, Z., Wang, G., Wang, R., Li, T., Liu, J., Zhang, Y., Liu, P., 2021. When the automated driving system fails: Dynamics of public responses to automated vehicles. *Transp. Res. Part C Emerg. Technol.* 129, 103271.
- Xu, Z., Wang, X., Wang, X., Zheng, N., 2025. Safety validation for connected autonomous vehicles using large-scale testing tracks in high-fidelity simulation environment. *Accid. Anal. Prev.* 215, 108011.
- Yan, Y., Peng, L., Shen, T., Wang, J., Pi, D., Cao, D., Yin, G., 2023. A multi-vehicle game-theoretic framework for decision making and planning of autonomous vehicles in mixed traffic. *IEEE Trans. Intell. Veh.* 8, 4572-4587.
- Yang, J., Zhou, K., Li, Y., Liu, Z., 2024. Generalized out-of-distribution detection: A survey. *Int. J. Comput. Vis.* 132, 5635-5662.
- Yang, X., Shi, Y., Xing, J., Liu, Z., 2022. Autonomous driving under V2X environment: state-of-the-art survey and challenges. *Intell. Transp. Infrastruct.* 1, liac020.
- You, J., Jiang, Z., Huang, Z., Shi, H., Gan, R., Wu, K., Cheng, X., Li, X. P., Ran, B., 2026. V2x-vlm: End-to-end v2x cooperative autonomous driving through large vision-language models. *Transp. Res. Part C Emerg. Technol.* 183, 105457.
- Yusuf, S. A., Khan, A., Souissi, R., 2024. Vehicle-to-everything (V2X) in the autonomous vehicles domain—A technical review of communication, sensor, and AI technologies for road user safety. *Transp. Res. Interdiscip. Perspect.* 23, 100980.
- Zang, S., Ding, M., Smith, D., Tyler, P., Rakotoarivelo, T., Kaafar, M. A., 2019. The impact of adverse weather conditions on autonomous vehicles: How rain, snow, fog, and hail affect the performance of a self-driving car. *IEEE Veh. Technol. Mag.* 14, 103-111.
- Zha, Y., Shangguan, W., Chai, L., Chen, J., 2024. Hierarchical perception enhancement for different levels of autonomous driving: A review. *IEEE Sensors J.* 24, 17366-17386.
- Zhang, L., Jiang, F., Kong, B., Yang, J., Wang, C., 2021. Real-time lane detection by using biologically inspired attention mechanism to learn contextual information. *Cogn. Comput.* 13, 1333-1344.
- Zhang, X., Khastgir, S., Tiele, J. K., Takenaka, K., Hayakawa, T., Jennings, P., 2024. Odd and behavior based scenario generation for automated driving systems. *IEEE Access.* 12, 10652-10663.
- Zhang, Z., Ning, H., Zhou, F., 2022. A systematic survey of driving fatigue monitoring. *IEEE Trans. Intell. Transp. Syst.* 23, 19999-20020.
- Zhao, W., Li, J., Dong, X., Xiang, Y., & Guo, Y., 2024. Segment every out-of-distribution object. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 3910-3920.
- Zhu, H., Yuen, K. V., Mihaylova, L., Leung, H., 2017. Overview of environment perception for intelligent vehicles. *IEEE Trans. Intell. Transp. Syst.* 18, 2584-2601.
- Zhu, H., Alhajyaseen, W., Iryo-Asano, M., Nakamura, H., Dias, C., 2022. Defensive or competitive Autonomous Vehicles: Which one interacts safely and efficiently with pedestrians? *Phys. A Stat. Mech. Appl.* 606, 128083.
- Zhu, M., Sze, N. N., Newnam, S., 2022. Effect of urban street trees on pedestrian safety: A micro-level pedestrian casualty model using multivariate Bayesian spatial approach. *Accid. Anal. Prev.* 176, 106818.
- Zhu, X., Hu, W., Deng, Z., Zhang, J., Hu, F., Zhou, R., Li, K., Wang, F. Y., 2022. Interaction-aware cut-in trajectory prediction and risk assessment in mixed traffic. *IEEE/CAA J. Autom. Sin.* 9, 1752-1762.
- Zhu, Z., Liang, D., Zhang, S., Huang, X., Li, B., Hu, S., 2016. Traffic-sign detection and classification in the wild. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2110-2118.

Author biography



Chengxi Hu received the B.S. and M.S. degrees in transportation from Southwest Jiaotong University, in 2023 and 2025, respectively. He is currently pursuing the Ph.D. degree with the Cho Chun Shik Graduate School of Mobility, Korea Advanced Institute of Science and Technology (KAIST). His main research interests include autonomous vehicle, traffic safety and human-machine interaction.



Sikai (Sky) Chen received the Ph.D. degree in civil engineering with a focus on computational science and engineering from Purdue University in 2019. He is an Assistant Professor with the Department of Civil and Environmental Engineering, Department of Mechanical Engineering (courtesy), University of Wisconsin–

Madison. His research centers around three major themes, such as human users, AI, and transportation. He aims to innovate and develop safe, efficient, sustainable, and human-centered transportation systems using cutting-edge methods and technologies.



Samuel Labi received the Ph.D. degree in civil engineering from Purdue University in 2001. He is currently a Professor of transportation and infrastructure systems engineering in Purdue University's Lyles School of Civil Engineering, where he also serves as the director of the Next

Generation Transportation Systems (Nextrans) Center and associate director of the University Transportation Center for Connected and Autonomous Vehicles (CCAT). His research interests include infrastructure management, transportation finance, project evaluation, data analytics, and the economic evaluation of connected and automated transportation.



Hongliang Ding received the Ph.D. degree in Civil and Environmental Engineering from The Hong Kong Polytechnic University in 2022. He is currently an Associate Researcher with the School of Smart City and Transportation, Southwest Jiaotong University. His research interests include intelligent road traffic safety, big

data mining in transportation, and human-machine interaction in connected and mixed traffic environments.



Hyung-Chul Chung received his Ph.D. degree in Urban and Regional Planning from the University of Florida in 2013. He is currently an Associate Professor of Urban Planning and Design at Xi'an Jiaotong-Liverpool University (XJTLU). His research interests include urban policy

and governance, transportation and mobility, smart city and AI, and socio-spatial segregation.



Tiantian Chen received the Ph.D. degree in Civil and Environmental Engineering from the Hong Kong Polytechnic University. She is now an Associate Professor at the Cho Chun Shik Graduate School of Mobility, Korea Advanced Institute of Science and Technology (KAIST). Her research interests include

traffic safety, driving simulation, travel behavior analysis, smart interaction, human factor concepts of future transport systems, and transport policy.