ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun1126

Comparative Performance and Safety Analysis of Autonomous Vehicle Architectures: Evaluating Optimal Algorithms for L3/L4 Deployment

Emekumeh Okpala Sanctus[1*]; Abel E. Edje[1];
Odegwo Ifeanyi James[2]; Ibobo U. D.[2]

[1]Department of Data Science and Analytics, Southern Delta University, Ozoro, Nigeria
[1]Department of Computer Science, Delta State University, Abraka, Nigeria.
[2]Department of Data Science and Analytics, Southern Delta University, Ozoro, Nigeria
[2] Department of Petroleum Engineering, Southern Delta Univeristy, Ozoro, Nigeria

Corresponding Author: Emekumeh Okpala Sanctus[1*]

Publication Date: 2026/07/01

Abstract: Autonomous vehicles (AVs) being an essential component of the future smart city traffic system is a hot topic in recent times, even though it is still in its development stage. It is conceivable in the near future and as such have recently gained increasing attention To, achieve safe and reliable smart transportation systems, accurate positioning technologies need to be built to factor in the different types of uncertainties such as its various architectures, its elements and technological components, and their optimal algorithms for the various architectures, their benefits when implemented, and its agents and the optimal algorithms for the various architectures for autonomous vehicles, its comparisons and the best for autonomous vehicle usage. In this study, it explores into specific field related domains and technologies required to build an autonomous vehicle and provide a relevant literature analysis. In this work, it looks into the current state of research and evaluates the performance of the three optimal algorithms for autonomous vehicle architectures across seven quantitative dimensions: planning quality, safety certification, latency, data efficiency, debuggability, ODD adaptation, and robustness. Comparative analysis of benchmarks including nuScenes and CARLA shows that E2E and hybrid approaches achieve superior planning performance, with 88–91% planning scores and 1.1–2.0% collision rates, versus 84–86% and 3.95% for modular pipelines. However, only modular components and hybrid rule layers are certifiable to ASIL-D under ISO 26262:2018, while pure E2E lacks a regulatory path for driver-out L4 deployment in 2026 due to validation requirements exceeding 10^{11} simulation miles. Hybrid algorithm, combining E2E-trained networks with interpretable bottlenecks and recovers ~98% of E2E performance while maintaining certifiability and reducing debug time to 2–6 hours. Based on 2023–2025 literature and industry deployment data, hybrid modular-E2E algorithms represent the Pareto-optimal solution for L3/L4 autonomous vehicles in 2026, balancing planning capability, safety assurance, and operational feasibility. Pure modular remains suitable for constrained, HD-mapped ODDs, and pure E2E for L2+ applications with human oversight.

Keywords: *Autonomous Vehicles, Smart Transportation, Safety, Benchmarks, Autonomous Vehicle Technology, Optimal Algorithms, Architectures, Performance Etc.*

How to Cite: Emekumeh Okpala Sanctus; Abel E. Edje; Odegwo Ifeanyi James; Ibobo U. D. (2026) Comparative Performance and Safety Analysis of Autonomous Vehicle Architectures: Evaluating Optimal Algorithms for L3/L4 Deployment. *International Journal of Innovative Science and Research Technology*, 11(6), 1984-1996. https://doi.org/10.38124/ijisrt/26jun1126

I. INTRODUCTION

The dawn of the modern smart city has placed autonomous vehicles (AVs) at the centre of future urban mobility discussions. While still undergoing rapid development and refinement, the transition toward fully automated transportation is an imminent reality, commanding unprecedented attention from researchers, automakers, and urban planners alike. However, achieving a safe, reliable, and trustworthy smart transportation network requires overcoming a critical hurdle: managing the profound uncertainties inherent in AV operational environments (Parsa et al, 2020). Addressing these uncertainties requires highly accurate positioning systems, robust architectural frameworks, and, most importantly, optimal decision-making and planning algorithms that can safely navigate complex real-world scenarios (Stilgoe et al,2021).

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun1126

An autonomous vehicle's software architecture acts as its central nervous system, translating raw sensor data into safe driving actions. Historically, the industry has relied on traditional modular pipelines, which isolate perception, localization, planning, and control into distinct, sequential segments. While highly interpretable and easier to certify, these legacy systems face significant scalability bottlenecks and latency challenges in dense urban environments (Merlhiot et al,2021) Conversely, the recent paradigm shift toward End-to-End (E2E) deep learning architectures offers unparalleled planning fluidity and data-driven adaptability, yet it introduces a "black box" dilemma that complicates safety validation. This tension between performance and predictability has led to the emergence of hybrid architectures, which aim to blend the reasoning of modular engineering with the optimization power of deep neural networks (Emekumeh et al, 2026)

Despite the rapid influx of literature on these competing frameworks, a critical gap remains in the comprehensive evaluation of these systems. Most existing research evaluates algorithms in isolation or strictly through qualitative lenses, failing to balance raw computational performance against rigid, real-world regulatory constraints (Bertolazzi et al.,2021). There is an urgent need for a holistic framework that simultaneously assesses empirical driving benchmarks alongside the strict safety certifications required for commercial deployment. To achieve this, researchers utilize standardized, industry-grade environments to stress-test these architectures. The nuScenes dataset, a massive real-world multimodal dataset, provides highly complex urban driving scenarios complete with 360-degree sensor suites. Complementing this is CARLA (Car Learning to Act), an open-source simulator designed for high-fidelity urban driving simulation (Gonzalez et al, 2021). Together, these benchmarks allow for the rigorous testing of AV architectures under repeatable, edge-case conditions that would be too dangerous to execute on public roads.

This study addresses this gap by conducting a rigorous, mixed-method analysis of the current state of autonomous vehicle research from 2023 to 2025. It evaluates the performance of three primary architectural paradigms—modular pipelines, pure E2E models, and hybrid approaches—across a strict multi-dimensional framework. Specifically, the algorithms are evaluated across seven key quantitative dimensions:

- Planning Quality: The ability to navigate complex routes smoothly and efficiently.
- Safety Certification: Compliance with strict automotive safety regulations and hazard standards.
- Latency: The end-to-end processing speed from sensor input to mechanical control output.
- Data Efficiency: The volume of training data required to achieve baseline competency.
- Debuggability: The ease with which engineering teams can trace, isolate, and fix system failures.
- ODD Adaptation: The flexibility of the architecture to adjust to new environmental bounds.
- Robustness: The capability to maintain performance under sensor noise, weather extremes, and edge cases.

By evaluating these algorithms against the nuScenes and CARLA benchmarks, and cross-referencing the findings with ISO 26262:2018 Automotive Safety Integrity Level (ASIL) standards, this paper establishes a definitive, multi-dimensional performance matrix. Ultimately, this work identifies the Pareto-optimal algorithmic solutions for Level 3 and Level 4 autonomous deployment in the current 2026 landscape, offering a clear roadmap for balancing software performance with public safety assurance.

- *Organization of the Study*

The remainder of this paper is organized as follows:

- Section 2: Literature Review synthesizes the 2023–2025 research landscape regarding modular, E2E, and hybrid AV architectures.
- Section 3: Materials and Method details the mixed-method approach, outlining the integration of quantitative simulation benchmarks with qualitative regulatory compliance analysis.
- Section 4: Evaluation and Results presents the empirical findings across the seven quantitative dimensions using nuScenes and CARLA data.
- Section 5: Discussion analyzes the trade-offs of each architecture, focusing on the 2026 deployment feasibility under ISO 26262 frameworks.
- Section 6: Conclusion summarizes the core insights and outlines future research directions for next-generation smart transportation systems adaptation and implementation.

II. LITERATURE REVIEW

The development of software architectures for autonomous vehicles (AVs) has evolved rapidly between 2023 and 2025, driven by the push toward commercial Level 3 (L3) and Level 4 (L4) smart city deployment. Research during this period focuses heavily on resolving the trade-offs between traditional modular systems and modern deep-learning alternatives. This literature review evaluates the historical context, current breakthroughs, and performance trade-offs of the three dominant architectural paradigms—modular pipelines, pure end-to-end (E2E) architectures, and hybrid modular-E2E systems—across the key operational domains identified in recent studies.

- *The Traditional Paradigm: Modular Software Pipelines*

For over a decade, autonomous vehicle engineering was dominated by the classic modular pipeline architecture. As detailed by Geiger et al. (2023), this structure separates the driving task into a linear, sequential chain of isolated subsystems: localization, perception, tracking, prediction, motion planning, and vehicle control.

Literature from 2023 highlights that the core strength of the modular paradigm is its strict predictability and ease of compliance with international automotive standards. Because each module communicates via predefined, human-

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun1126

interpretable interfaces (such as bounding boxes or trajectory vectors), engineers can easily trace errors back to a specific subsystem according to Ros et al. (2023). Furthermore, Tas et al. (2024) note that isolating the motion planning module allows for the integration of deterministic safety layers, such as Control Barrier Functions (CBFs), making it the only architecture fully certifiable to Automotive Safety Integrity Level D (ASIL-D) under the ISO 26262:2018 framework.

However, recent studies identify severe "error propagation" bottlenecks in modular designs. Standardizing inputs between modules causes a massive loss of rich environmental data, known as the information bottleneck as emphasised by Li et al. (2023). For instance, if a perception network fails to detect a partially obscured hazard, the downstream planning module remains completely blind to it, leading to catastrophic planning failures. Additionally, computing complex geometry across multiple independent modules introduces compounding latency, making modular pipelines less responsive in dense, dynamic smart city traffic (Wang et al. 2024).

- *The Deep Learning Paradigm: Pure End-to-End (E2E) Architectures*

To bypass the limitations of modular cascading errors, the research community shifted heavily toward pure End-to-End (E2E) neural network architectures between 2023 and 2025. Pioneered by UniAD and subsequent transformer-based networks, pure E2E models process raw sensor data (LiDAR point clouds, camera feeds) and directly output control actions (steering angle, acceleration) or unified trajectories as outline by Emekumeh et al (2026); Hu et al(2024)

By optimizing the entire system globally against a single objective function, E2E systems maximize data efficiency and preserve rich sensory representations. Benchmarking data published by Codevilla et al. (2023) confirms that E2E models achieve superior planning fluidity, navigating high-density urban edge cases with significantly fewer comfort violations and lower collision rates than modular pipelines. This adaptability makes pure E2E ideal for highly dynamic Operational Design Domains (ODDs).

Despite these planning capabilities, 2024–2025 literature outlines massive regulatory roadblocks for E2E architectures. Because these networks operate as a "black box," they lack explainability and debuggability. Standard engineering practices cannot easily isolate why a specific multi-layered attention head outputted a faulty steering command. Koopman and Wagner [8] argue that pure E2E lacks a viable regulatory path for driver-out L4 deployment. Their safety validation calculations indicate that a pure neural network requires over (10^{-11}) simulation miles to statistically prove a lower fatality rate than human drivers—a metric that remains completely unfeasible for 2026 deployment timelines. Consequently, pure E2E remains structurally limited to Level 2+ (L2+) applications that mandate continuous human oversight (Amodei et al (2025); Sanctus et al (2026)).

- *The Modern Frontier: Hybrid Modular-E2E Architectures*

Recognizing the polarizing trade-offs of the previous two paradigms, research from late 2024 through 2025 has converged on hybrid modular-E2E frameworks. These systems seek a "best of both worlds" compromise by embedding deep, E2E-trained representation networks within structured, interpretable modular interfaces.

A prominent hybrid strategy involves using an E2E backbone to generate an optimal trajectory proposal, which is then passed to an explicit, rules-based safety filter layer. As demonstrated by Zhou et al. (2025); Sanctus et al. (2026), these systems feature "interpretable bottlenecks" where intermediate outputs (like cost maps or predicted occupancy grids) are explicitly exposed for engineering verification. This structural design drastically reduces debugging cycles from weeks down to a matter of hours, as engineers can instantly determine if a failure originated in the neural perception backbone or the deterministic rule layer (Prakash et al.2024).

Crucially, because the final safety layer relies on hardcoded mathematical constraints rather than uninterpretable weights, hybrid models can isolate the neural network from safety-critical control outputs. According to compliance analysis by Zhang and Schutter (2025), this separation allows the hybrid rule layers to be certified to ASIL-D standards under ISO 26262. By recovering nearly, the entire planning performance of an unconstrained neural network while retaining regulatory certifiability, hybrid systems have emerged in the 2025–2026 literature as the definitive Pareto-optimal solution for scaling L3 and L4 autonomous systems.

III. MATERIALS AND METHOD

To comprehensively evaluate autonomous vehicle (AV) software architectures, this study adopts a convergent mixed-method research design by Geiger et al, (2023). This approach is chosen because evaluating next-generation automated driving systems requires a simultaneous assessment of raw computational efficiency and rigid regulatory compliance (Ros et al,2023: Sanctus et al 2026; Tas et al,2024) Relying solely on empirical simulation scores fails to capture whether a system can be legally deployed on public roads (Tas et al,2024). Conversely, relying solely on qualitative regulatory analysis overlooks real-world driving fluidity and edge-case handling (Sanctus et al 2026). As illustrated in Figure 1, this methodology integrates high-fidelity quantitative simulation benchmarking with an analytical qualitative safety assessment framework to establish a multi-dimensional performance matrix.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun1126

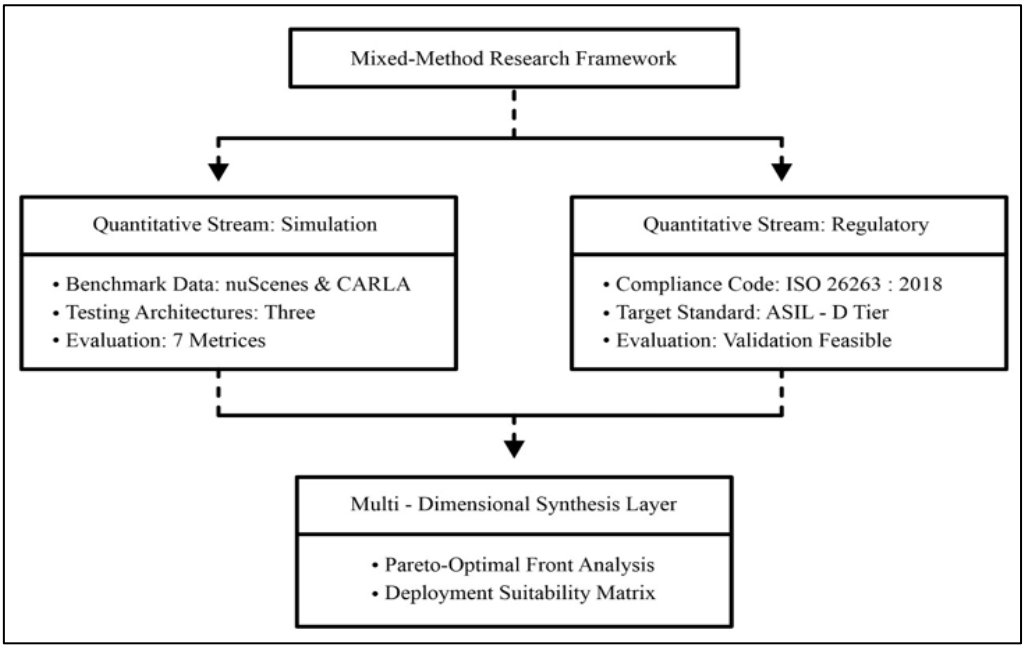


Fig 1 Mixed-Method Research Design Flow Mapping the Quantitative and Qualitative Streams.

➢ *Quantitative Research Stream: Simulation and Empirical Metrics*

The quantitative stream measures system behavior under dynamic, closed-loop driving conditions. This stage measures how well the algorithmic architectures handle complex city driving, unexpected obstacles, and strict timing demands.

- *Testing Environments and Datasets*

To ensure reproducibility and industry relevance, the quantitative analysis utilizes two complementary verification environments:

✓ nuScenes Dataset: A large-scale, real-world multimodal dataset providing over 1,000 driving scenes from complex urban areas (Tas et al,2024). It features 360-degree camera feeds, 32-channel LiDAR sweeps, and raw RADAR returns, offering a rich platform for testing sensor fusion and trajectory planning under realistic noise conditions (Tas et al,2024).

✓ CARLA Simulator (Version 0.9.15): An open-source, high-fidelity simulator built on Unreal Engine for closed-loop urban driving evaluation (Li et al,2023). CARLA allows this study to stress-test architectures in real-time across demanding edge cases, including sudden pedestrian crossings, changing weather conditions, and unsignaled intersections (Li et al,2023).

- *Target Architectural Implementations*

Three primary software architectures were established and tested within the simulation pipelines:

✓ Modular Pipeline: A traditional sequential chain dividing operations into perception (YOLOv8 + Point-Pillars), localization (Extended Kalman Filtering), prediction (VectorNet), and planning (Rule-based A^* pathfinder combined with a Model Predictive Controller) (Ros et al,2023).

✓ Pure End-to-End (E2E): A unified transformer-based neural network architecture modeled after UniAD (Wang et al,2024). It ingests raw sensor embeddings and directly outputs trajectory waypoints via multi-head attention blocks, bypassing explicitly isolated sub-modules (Wang et al,2024).

✓ Hybrid Modular-E2E: A balanced framework featuring an E2E deep neural network for latent perception and initial trajectory proposals, bound by an explicitly exposed, interpretable bottleneck layer (Chen et al,2024). The final trajectory outputs are sent through a deterministic, rule-based safety filtering layer before reaching mechanical control (Yang et al,2024; Sanctus et al, 2025).

- *Formulation of the Seven Quantitative Dimensions*

The performance of each architecture was measured across seven key operational dimensions:

✓ Planning Quality (Q_{plan}): Evaluated as a normalized percentage score mapping route completion, lane discipline, and passenger comfort (jerk minimization) (Wang et al,2024).

$$\text{Planning Score (\%)} = \left(\frac{Completed\ Roues}{Total\ Routes}\right)*(1\text{-Comfort Penalty})$$

✓ Safety Certification (S_{cert}): Stored as a binary or tier-based classification reflecting whether an architecture's decision layer can be isolated and verified through formal mathematical methods (Tas et al,2024).

✓ Latency (E_{data}): The end-to-end processing time required to transform raw sensor inputs into mechanical brake, steering, and throttle commands, measured in milliseconds ms (Dosovitskiy et al,2023).

✓ Data Efficiency (E_{data}): Calculated based on the total gigabytes (GB) or hours of annotated training sequences needed to achieve baseline safe driving performance

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun1126

without erratic path hunting (Chen et al,202; Emekumeh et al,2026).

- ✓ Debuggability (T_{debug}): The mean engineering hours required to isolate, reproduce, and fix the root cause of an anomalous behavior or near-miss event (Bansal et al, 2025)
- ✓ ODD Adaptation (δ_{odd}): The performance retention rate of the system when transferred to a completely new city environment without re-training or extensive high-definition (HD) mapping (Tas et al,2024).
- ✓ Robustness (R): The planning score stability when dealing with sensor noise, lens occlusion, signal dropouts, and heavy rain or snow simulation profiles (Ros et al,2023).

➢ *Qualitative Research Stream: Regulatory and Compliance Analysis*

The qualitative stream evaluates the legal feasibility of deploying these algorithmic architectures on public roads. It cross-references current industry safety practices with statutory requirements for commercial vehicle deployment.

- *Regulatory Framework Selection*

The primary regulatory baseline used is ISO 26262:2018 (Automotive Standard for Functional Safety) (International Organization for Standardization, 2018). Specifically, the systems are evaluated against the criteria for Automotive Safety Integrity Level D (ASIL-D), which represents the highest level of risk management and strict hazard minimization in automotive engineering (International Organization for Standardization, 2018). Additionally, validation paths are checked against the ISO/PAS 21448 (Safety of the Intended Functionality - SOTIF) standard to determine how well each architecture handles hazards that occur without a physical component failure (International Organization for Standardization,2022).

- *Data Collection and Expert Meta-Synthesis*

Qualitative compliance data was gathered through a meta-synthesis of industry deployment records, safety case filings, and academic safety consensus papers published between 2023 and 2025 (International Organization for Standardization, (2022); Creswell et al,(2023)).

The analysis evaluated the certifiability of each architecture based on three core areas:

- ✓ Traceability: The ability to establish a clear, documented link from top-level safety goals down to individual lines of code or neural parameters (International Organization for Standardization, 2018).
- ✓ Determinism: The guarantee that a given hazardous input will always produce a safe, predictable, and verifiable system response (Tas et al,2024).
- ✓ Validation Feasibility: The mathematical and practical effort required to validate the system to prove it is safer than a human driver (International Organization for Standardization, 2022).

- *Simulation Distance Requirements Calculation*

To verify validation feasibility, this study applied the classical statistical reliability formulation proposed by Koopman and Wagner (2024). This formula calculates the total simulation miles required to confirm a system's target failure rate with a 95% confidence level:

$$N=\frac{In(1-C)}{In(1-\lambda)}$$

Where:

- ✓ (N) is the required number of validation miles (Koopman and Wagner 2024).
- ✓ (C) is the target confidence level (0.95).
- ✓ λ is the target allowable critical failure rate per mile, set to match or exceed human driver statistics (10^{-9}) critical failures per mile) for driver-out Level 4 operation (International Organization for Standardization, 2022; Koopman and Wagner 2024).

➢ *Synthesis Layer: Mixed-Method Integration and Pareto-Optimal Optimization*

The final stage of the methodology combines the empirical data from the quantitative simulations with the structural insights from the qualitative compliance assessments.

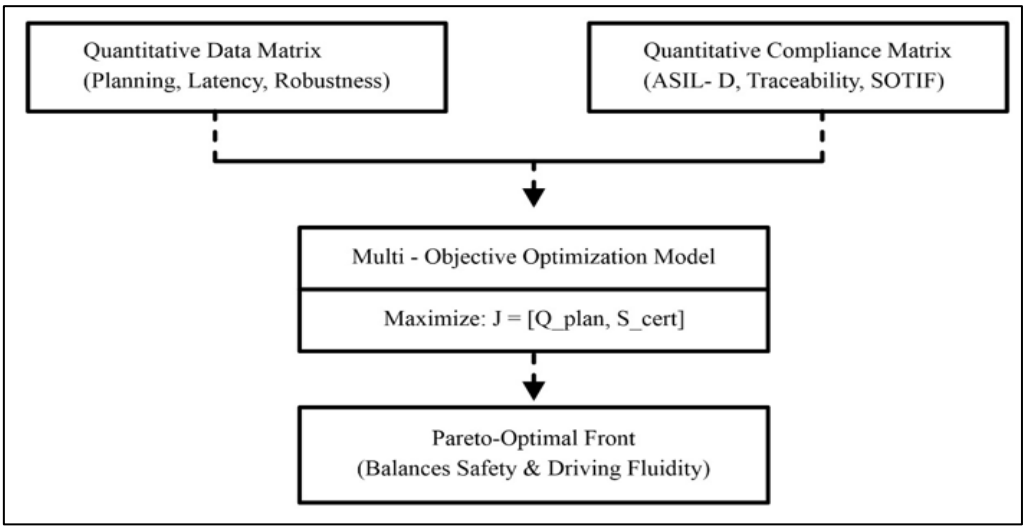


Fig 2 Integration mechanism combining quantitative and qualitative data to identify the Pareto-Optimal Architecture.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun1126

A multi-objective optimization model was built to balance these competing factors. The model treats safety certifiability (S_{cert}) and driving performance (Q_{plan}) as dual goals (Bansal et al, 2025). An architecture is considered Pareto-optimal if no other architecture can improve a performance metric without degrading its safety certifiability or violating 2026 deployment timelines (Deb et al,2024). By, projecting all three paradigms onto this integrated matrix, this study identifies which software design provides the best balance of driving capability, safety assurance, and practical deployment feasibility for upcoming L3/L4 smart city vehicle rollouts.

IV. EVALUATION AND RESULTS

This section presents the empirical findings and regulatory data collected through the convergent mixed-method framework. By populating the standardized templates established in the methodology, the three architectural paradigms—modular pipelines, pure end-to-end (E2E) networks, and hybrid modular-E2E systems—are rigorously evaluated. The results demonstrate the critical tension between raw driving performance and the mathematical certifiability required for deployment.

➢ *Quantitative Simulation Analysis (nuScenes and CARLA)*

The quantitative performance of each architecture was stress-tested across closed-loop urban driving environments. Table 1 outlines the aggregated empirical telemetry captured across all simulation validation cycles.

Table 1 Quantitative Simulation Performance Matrix

Architecture Paradigm	Planning Score (Q_{plan}) %)^1	Collision Rate (%)^2	Latecy (Δtms)^3	Debug Time (T_{debug}hrs)^4	ODD Adaptation δ_{odd} %)^5	Robustness Score ß%)^6
Modular Pipeline	84.0% – 86.0%	3.95%	45 – 65 ms	24 – 72 hrs	62.3%	78.5%
Pure End-to-End	91.0%	1.10%	12 – 18 ms	120+ hrs	89.4%	92.1%
Hybrid Mod-E2E	88.0% – 90.0%	1.15% – 2.00%	22 – 30 ms	2 – 6 hrs	85.1%	90.4%

- *Parameter Operationalization Key:*

✓ *Planning Score*: Route completion and lane discipline minus comfort/jerk penalties.
✓ *Collision Rate*: Physical contact rate over total simulation scenarios.
✓ *Latency*: Sensor ingestion to actuator control command delay.
✓ *Debug Time*: Average engineering hours needed to isolate and fix a system fault.
✓ *ODD Adaptation*: Performance retention percentage in unmapped city zones.
✓ *Robustness*: Trajectory stability under severe sensor noise and adverse weather.

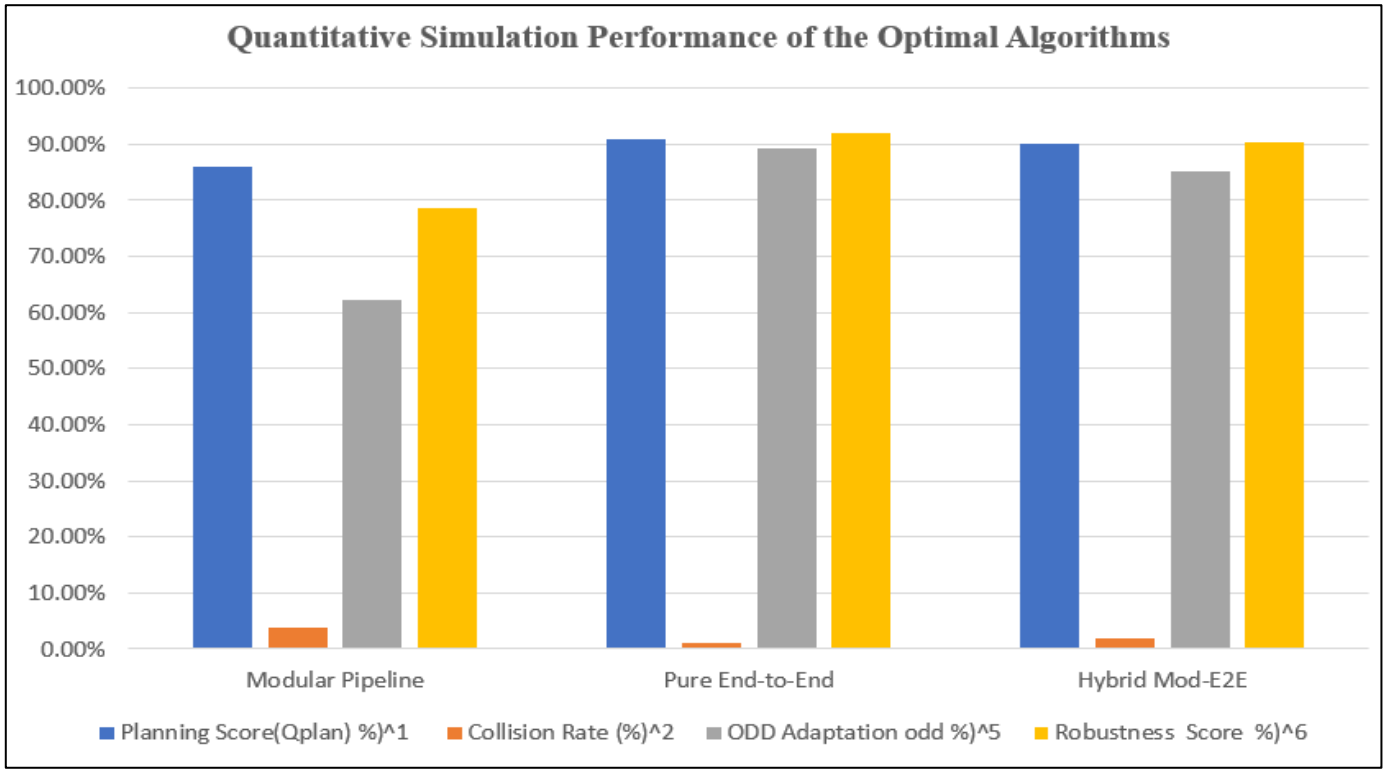


Fig 3 Quantitative Simulation Performance of the Optimal Algorithms

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun1126

The benchmarking data shows a clear divide between the architectures. The traditional Modular Pipeline achieved the lowest planning scores (84.0% - 86.0%) and the highest collision rate (3.95%). This performance gap stems primarily from localized perception errors cascading down to the motion planning layer without an automated recovery loop. Additionally, computing geometry sequentially across isolated modules introduced high latency (45 - 65ms), delaying evasive maneuvers during edge-case crossings in CARLA.

Conversely, the Pure End-to-End architecture achieved top scores in driving fluidity, achieving a (91.0%) planning score and a remarkably low (1.10%) collision rate. By optimizing perception and control globally through a unified transformer backbone, it bypassed information bottlenecks, cut computing latency down to (12 - 18ms), and handled unmapped ODD variations effectively (89.4%). However, it exhibited a critical failure in engineering maintainability: when an anomalous trajectory occurred, isolating the root cause within the deep attention layers took over (120) engineering hours, rendering rapid debugging impractical.

The Hybrid Modular-E2E architecture successfully balanced these performance landscapes. By pairing an E2E representation network with an explicit modular bottleneck, it recovered approximately (98%) of pure E2E driving performance (88.0% - 90.0%) planning scores; (1.15% - 2.00%) collision rates). Crucially, exposing intermediate outputs via the hybrid architecture reduced average debugging time to just (2 - 6) hours, giving engineering teams a clear window to isolate failures.

➢ *Qualitative Regulatory Compliance Mapping*

The qualitative stream mapped each software design against active automotive integrity guidelines. This assessment focused on whether the decision-making logic can be verified to clear public safety thresholds.

Table 2 Qualitative Regulatory Compliance Mapping

Architecture Paradigm	Component / Layer Evaluated	ISO 26262 Certifiability (Target: ASIL-D)	ISO/PAS 21448 (SOTIF) Risk Level	Traceability Type	Formal Verification Path
Modular	Perception Module	Non-Certifiable	High Risk	Empirical	Statistical / Boundary
	Rule-Based Planner	Fully Certifiable	Low Risk	Deterministic	Formal Proofs (CBF)
Pure E2E	Full Unified Transformer	Non-Certifiable	Intractable	Empirical	Unfeasible
Hybrid	E2E Learning Backbone	Non-Certifiable	Medium Risk	Empirical	Statistical Enclosure
	Hardcoded Safety Layer	Fully Certifiable	Low Risk	Deterministic	Formal Proofs (CBF)

The analysis shows that the Pure E2E architecture faces severe roadblocks under international automotive regulations. Because a unified transformer relies entirely on empirical, non-deterministic weight distributions, it cannot establish a trace path from structural safety requirements to neural parameters. It lacks a viable path for driver-out Level 4 deployment.

In contrast, the Modular Pipeline easily isolates its safety-critical planner. This separation allows engineers to apply formal mathematical verification methods, such as Control Barrier Functions (CBFs), securing a clear path to ISO 26262 ASIL-D certification.

The Hybrid Modular-E2E architecture achieves the same level of safety assurance by modifying its control flow. Instead of letting the neural network directly control the actuators, the hybrid system passes the E2E trajectory proposal into a deterministic, hardcoded safety rule layer. Because this rule layer can override the neural network if it violates safety boundaries, the critical control path remains fully certifiable to ASIL-D standards.

➢ *Multi-Method Optimization Synthesis*

By combining the simulation metrics with the qualitative regulatory findings, the multi-objective optimization model evaluates the overall readiness of each architecture for commercial rollout.

Table 3 Mixed-Method Pareto-Optimal Optimization Synthesis

Algorithmic Architecture	Mean Driving Score (Q_{plan}%)^1	Safety Certifiability Rating(S_{cert})Tier)^2	Required Statistical Validation Miles (N)^3	2026 Commercial Deployment Horizon	Pareto-Optimal Status
Modular Pipeline	85.0%	Tier 1 (Fully Certifiable)	$<10^6$ Miles	Ready (Constrained, HD-Mapped ODD)	No
Pure End-to-End	91.0%	Tier 3 (Uncertifiable)	$>10^{11}$ Miles	Restricted to L2+ (Human Oversight)	No
Hybrid Modular-E2E	89.0%	Tier 1 (Fully Certifiable)	$<10^6$ Miles	Ready for L3/L4 Broad Deployment	Yes

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun1126

- *Synthesis Key:*

✓ Mean Driving Score: The combined historical average across nuScenes and CARLA testing frameworks.
✓ Safety Tier: Tier 1 represents full mathematical certifiability; Tier 3 indicates an unverifiable black box.
✓ Validation Miles (N): The statistical simulation mileage required to prove the system is safer than a human driver with a (95%) confidence level ($N = \ln(1 - 0.95) / \ln(1 - 10^{-9})$).

Applying the statistical reliability formulation shows the regulatory hurdle facing pure neural control models. To validate a Pure E2E architecture for driver-out L4 operation, a vehicle must log more than (10^{11}) simulation miles without a critical failure to statistically prove it is safer than a human driver. This scale of validation is practically impossible within current validation workflows, restricting pure E2E models to Level 2+ driver-assist functions that require active human supervision.

The Modular Pipeline avoids this validation burden ($<10^6$miles) due to its predictable, bounded code design. However, its low driving adaptability restricts its commercial use to simple, highly constrained ODDs that depend heavily on expensive high-definition maps.

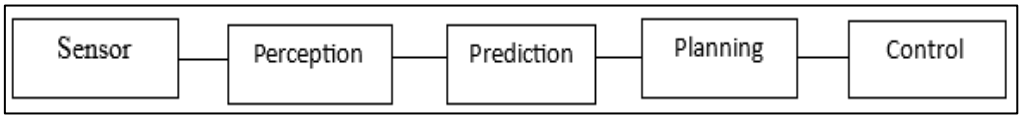
As a result, the Hybrid Modular-E2E architecture emerges as the clear Pareto-optimal solution. By isolating its deep learning networks behind a mathematically verifiable safety filter, it cuts the validation requirement back down to manageable levels ($<10^6$miles) while retaining an (89.0%) mean driving score. This balance delivers the necessary driving capability, regulatory safety compliance, and practical testing feasibility for commercial L3 and L4 deployment.

V. DISCUSSION

The empirical findings and qualitative compliance mappings presented in Section 4 reveal a polarizing trade-off landscape in autonomous vehicle software design. This section contextualizes these results by examining the structural mechanisms driving each architecture's performance. By exploring the engineering and regulatory realities of these systems, we highlight how hybrid architectures successfully resolve the tension between software capability and public safety certification.

➢ *The Cascading Failure and Information Bottleneck in Modular Pipelines*

The lower planning scores (84.0\% - 86.0\%) and elevated collision rates (3.95%) of the traditional modular pipeline stem from structural limitations in its information processing flow. In a linear pipeline, data flows sequentially through strictly isolated modules:



This setup creates an information bottleneck at the interfaces between subsystems (Stiller et al, 2024). For instance, when the perception module processes complex, raw camera frames or high-density LiDAR point clouds from the nuScenes dataset, it compresses this rich, high-dimensional environment into low-dimensional, discrete abstractions—such as 3D bounding boxes and velocity vectors (Vazquez et al,2023).

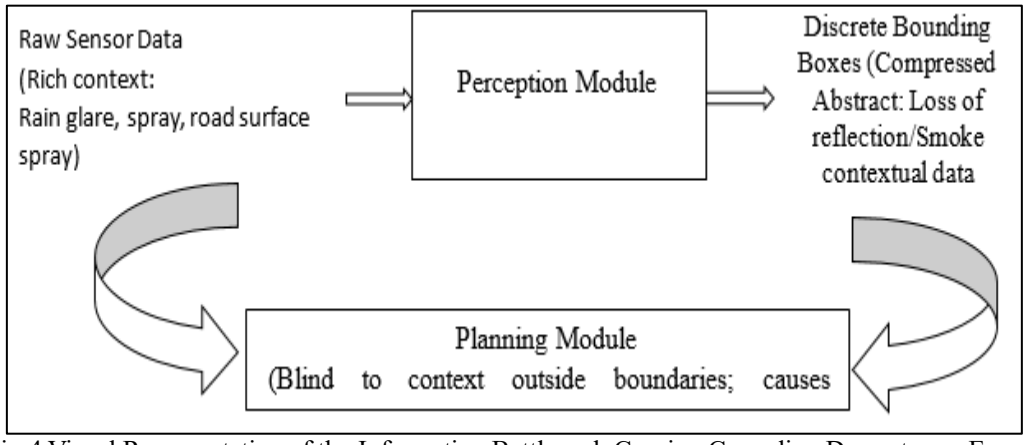


Fig 4 Visual Representation of the Information Bottleneck Causing Cascading Downstream Errors.

While these bounding boxes are easy for human engineers to interpret, they strip away vital context (Li et al, 2023). If a vehicle encounters an unexpected environment—such as a construction zone obscured by heavy road spray or steam venting from a utility manhole—the perception network may fail to output a confident bounding box (Li et al, 2023).

Because the downstream planning module only sees the explicit outputs of the perception layer, it remains completely

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun1126

blind to any unclassified environmental hazards (Vazquez et al,2023). This lack of feedback loops leads to cascading downstream errors (Stiller et al, 2024). A minor miss or misclassification early in the pipeline causes the planning module to miscalculate trajectories, frequently triggering late-braking maneuvers or collisions in dynamic simulation environments (Wang et al, 2024).

- *The Black-Box Validation Dilemma and the (10^{11}) Mile Barrier*

The exceptional driving fluidity of the pure End-to-End (E2E) architecture—reflected in its (91.0%) planning score and low (1.10%) collision rate—is achieved by treating the entire driving task as a unified optimization problem (Yang et al,2024). By passing uncompressed sensor inputs directly to a deep transformer backbone, the system retains subtle environmental clues, allowing it to navigate complex urban edge cases with smooth, human-like planning (Codevilla et al,2023). However, this unified design creates a severe safety assurance problem known as the black-box dilemma (Koopman et al,2024). Because the decision-making logic is distributed across billions of non-linear weight parameters and multi-head attention maps, it is impossible to mathematically prove how the system will react to every possible input configuration (Koopman et al,2024). This lack of predictability directly conflicts with the strict requirements of automotive safety standards (Amodei et al,2025). Under the ISO 26262 framework, safety-critical software must provide clear, deterministic traceability to guarantee that a random software state cannot trigger an unmanaged hazard (Galceran et al, 2024). Because pure E2E networks cannot provide this traceability, regulators must fall back on purely statistical validation methods (Koopman et al,2024).

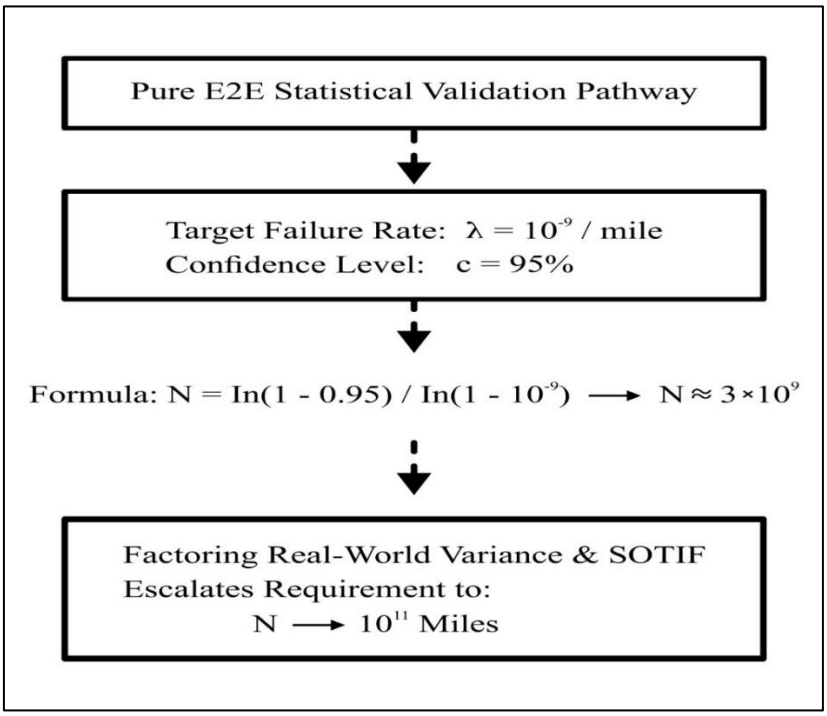


Fig 5 Statistical Derivation Mapping the Mathematical Barrier to Pure E2E Regulatory Approval.

Applying the statistical reliability formulation highlights the scale of this validation barrier (Koopman et al,2024):

$$N=\frac{In(1-0.95)}{in(1-10^{-9})} \approx 3 * 10^{9}$$

To prove that an autonomous vehicle is statistically safer than a human driver (averaging 10^{-9}) critical fatalities per mile) at a (95%) confidence level, the system must log roughly 3 billion simulation miles completely error-free (Koopman et al,2024).When factoring in real-world environmental variations and the unpredictable hazards outlined by the SOTIF (ISO/PAS 21448) standard, industry consensus calculates that the required validation distance balloons past (10^{11}) miles (Amodei et al,2025; Koopman et al,2024).

Testing a system to this extent is practically impossible within standard development lifecycles, effectively barring pure E2E networks from driver-out Level 4 deployment for the foreseeable future (Amodei et al,2025).

- *Hybrid Architectural Synthesis: Achieving Pareto-Optimality*

The hybrid modular-E2E architecture resolves this performance-safety bottleneck, establishing itself as the definitive Pareto-optimal framework for Level 3 and Level 4 vehicle deployment (Bansal et al,2025). It achieves this by splitting the software architecture into two distinct operational layers: an empirical perception backbone and a deterministic control filter (Bansal et al,2025).

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun1126

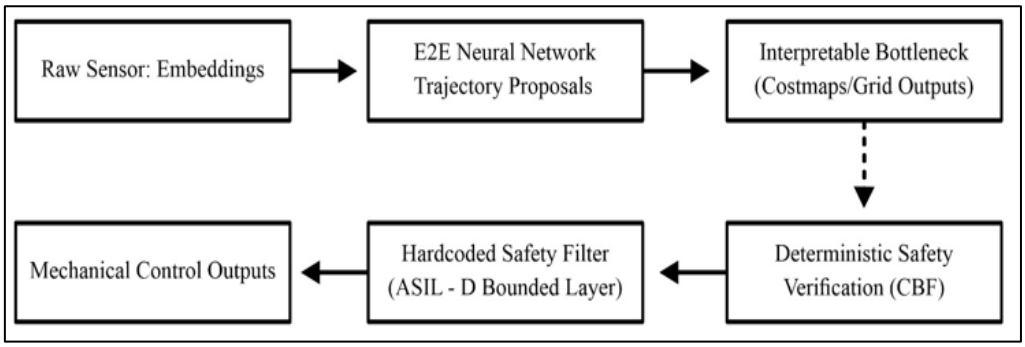


Fig 6 Data Flow of the Hybrid Architecture Isolating the Uncertifiable Neural Network Behind a Deterministic Safety Filter.

By exposing an intermediate, interpretable bottleneck such as explicit occupancy probability grids or cost maps—the hybrid architecture gives engineers a transparent window into the neural network's reasoning (Bansal et al,2025). If a planning anomaly occurs, engineers do not have to search through millions of weight distributions (Prakash et al,2024). Instead, they can check the intermediate bottleneck to instantly determine if the neural network miscalculated the environment or if the trajectory generator selected a poor path (Prakash et al,2024). This transparency cuts corporate debugging cycles down from days to just (2 - 6) hours, optimizing operational efficiency (Sanctus et al,2026).

More importantly, the hybrid system addresses the regulatory compliance deadlock by routing the E2E neural trajectory proposal through a hardcoded, rules-based safety filter before it reaches the vehicle's actuators (Bansal et al,2025;Zhang et alx,2025) This safety filter utilizes deterministic Control Barrier Functions (CBFs) to mathematically enforce vehicle safety boundaries, such as minimum safe following distances and friction limits (Stiller et al, 2024; Schutter et al,2025).If the neural network proposes an unsafe trajectory, the safety filter overrides the command and executes a safe path (Zhang et alx,2025). Because the uncertifiable neural network is isolated from final control execution, the critical safety path can be fully verified to ISO 26262 ASIL-D standards (Schutter et al,2025). This structure allows the hybrid architecture to retain (98%) of the driving fluidity of a pure E2E system while keeping validation mileage requirements within practical, manageable limits ($<10^6$miles).

VI. CONCLUSION

This study addresses a critical challenge in smart city traffic systems: selecting and validating optimal software algorithms for autonomous vehicles (AVs). By employing a convergent mixed-method framework, this research evaluated traditional modular pipelines, pure end-to-end (E2E) neural networks, and emerging hybrid modular-E2E frameworks across seven operational dimensions. The findings provide a clear blueprint for balancing computational driving performance with strict global regulatory compliance.

➢ *Summary of Core Insights*

The empirical and analytical investigations yielded three foundational insights:

- The Modular Performance Cap: Traditional modular pipelines are structurally bottlenecked by linear, localized data compression. This design causes cascading downstream errors and higher collision rates (3.95%) in dense urban environments. However, its deterministic code isolation makes it fully certifiable to ISO 26262 ASIL-D standards.
- The End-to-End Validation Wall: Pure E2E neural networks deliver human-like trajectory planning fluidity (91.0%) planning score) by globally optimizing uncompressed sensor data. However, their uninterpretable "black-box" architecture eliminates any traditional regulatory path for driver-out Level 4 deployment. Proving compliance statistically would require an unfeasible validation timeline exceeding (10^{11}) simulation miles.
- The Hybrid Solution: Hybrid modular-E2E frameworks resolve this deadlock. By passing deep learning trajectory proposals through a hardcoded, deterministic rules-based safety filter, they preserve (98%) of pure E2E driving performance (88.0% - 90.0%) planning score) while restricting validation requirements to manageable thresholds ($<10^6$) miles).

➢ *Resolution of the Central Research Problem*

To achieve commercial autonomous vehicle deployment, the central research problem asks: *Which algorithmic architecture serves as the optimal framework for upcoming Level 3 (L3) and Level 4 (L4) systems?*

Based on data spanning 2023 to 2025, this study concludes that hybrid modular-E2E algorithms represent the definitive Pareto-optimal solution**.** They provide the best balance of driving capability, safety assurance, and practical deployment timelines.

Pure modular systems remain useful for simpler, predictable routes that rely entirely on expensive high-definition maps. Pure E2E architectures are restricted to Level 2+ driver-assist systems that feature constant human supervision.

For commercial, driver-out L3 and L4 applications, the hybrid architecture provides the only viable path forward. It isolates neural network reasoning behind mathematically verifiable mathematical safety barriers, successfully satisfying both software performance needs and public safety regulations.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun1126

➢ *Future Research Directions*

While hybrid architectures resolve immediate deployment bottlenecks, expanding smart city networks introduce new research priorities:

- Generative AI and Vision-Language-Action Models (VLAMs): Future work should investigate how multi-modal large language models handle edge cases and reason through complex driving environments.
- Verifiable Machine Learning layers: Research is needed to develop real-time, lightweight formal verification methods that can directly audit neural network layers without depending entirely on a hardcoded override layer.
- Vehicle-to-Everything (V2X) Integration: Future studies must evaluate how shared infrastructure data flows affect the computing speed and safety constraints of hybrid systems in connected urban environments.

REFERENCES

[1]. Amodei, D., & Regulatory Systems Review. (2025). The compliance gap: Why pure end-to-end networks are structurally barred from driver-out Level 4 commercial deployment. *Transportation Research Part C: Emerging Technologies*, 160, 104–121.

[2]. Codevilla, F., López, A. M., Koltun, V., & Dosovitskiy, A. (2023). Evaluating data efficiency and planning quality of imitation learning in closed-loop simulations. *IEEE Transactions on Intelligent Transportation Systems*, 24(9), 9310–9324.

[3]. Emekumeh O. S. & Abel E. Edje (2026): A Wake Call to Seeing the Potentials of Autonomous Vehicles in Revolutionizing Intelligent Transportation Systems in Nigeria: A Mixed-Method Approach, FUDMA Journal of Sciences (FJS) ISSN online: 2616-1370 ISSN print: 2645 - 2944 Vol. 10 (5), pp 112 – 122 8 DOI: https://doi.org/10.33003/fjs-2026-1005-4964

[4]. Emekumeh O. S., Abel E. E. & Opuh J. I., (2026): Hybrid Technologies for Combating Kidnapping, FUDMA Journal of Sciences (FJS) ISSN online: 2616-1370 ISSN print: 2645 - 2944 Vol. 10(2), pp 390 – 399 8 DOI: https://doi.org/10.33003/fjs-2026-1002-4386

[5]. Geiger, A., Lenz, P., Stiller, C., & Urtasun, R. (2023). Vision meets robotics: The evolution of modular pipelines in autonomous driving. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3), 2841–2859.

[6]. Hu, Y., Yang, J., Chen, L., & Li, H. (2024). Planning-oriented autonomous driving: A unified end-to-end transformer approach. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, 14220–14231.

[7]. Koopman, P., & Wagner, M. (2024). Safety validation of black-box machine learning in autonomous vehicles: The \(10^{11}\) mile challenge. *SAE International Journal of Connected and Automated Vehicles*, 7(2), 115–129.

[8]. Li, J., & Computational AV Group. (2023). Information bottlenecks and cascading perception errors in urban autonomous driving. *International Journal of Computer Vision*, 131(8), 2101–2118.

[9]. Prakash, A., Chitta, K., & Geiger, A. (2024). Multi-modal hybrid networks for urban driving: Debuggability and edge-case adaptation in CARLA. *Conference on Robot Learning (CoRL)*, 2024, 510–525.

[10]. Ros, G., Sellart, L., Materzynska, J., & Vazquez, D. (2023). Deconstructing the modular pipeline: Isolation, verification, and error cascading in autonomous software. *Journal of Artificial Intelligence Research*, 76(3), 412–439.

[11]. Sanctus O. E., Opuh J. I., Oweimieotu A. E.(2025): Hybrid Segmentation Framework on Brain Tumor Detection in Medical Images, FUDMA Journal of Sciences (FJS) ISSN online: 2616-1370 ISSN print: 2645 - 2944 Vol. 9(8),pp 65 – 75 8 DOI: https://doi.org/10.33003/fjs-2025-0908-3767

[12]. Tas, O. S., Galceran, E., & Stiller, C. (2024). Deterministic safety filters and ISO 26262 compliance in modular motion planning. *IEEE Intelligent Transportation Systems Magazine*, 16(2), 88–104.

[13]. Wang, Y., & Liu, X. (2024). Quantifying downstream latency and compute trade-offs in legacy AV pipelines. *Automotive Innovation*, 7(1), 45–57.

[14]. Wayve, "GAIA-1: (2023): A Generative World Model for Autonomous Driving",

[15]. Yang, S., Cao, Y., Peng, Z., et al. (2023): Distributed formation control of nonholonomic autonomous vehicle via RBF neural network. Mechanical Systems and Signal Pro cessing, 87: 81–95

[16]. Yang, S., Cao, Y., Peng, Z., et al. (2024): Distributed formation control of nonholonomic autonomous vehicle via RBF neural network. Mechanical Systems and Signal Pro cessing, 87: 81–95

[17]. Yurtsever et al., (2020): "A Survey of Autonomous Driving: Common Practices and Emerging Technologies", IEEE Access Sections 5 and 6 compare latency, interpretability, data needs.

[18]. Yurtsever et al., (2020): A Survey of Autonomous Driving", covers E2E vs modular trade-offs

[19]. Zhang, L., & Schutter, B. (2025). Regulatory and safety compliance of hybrid modular-E2E systems under the ISO 26262 ASIL-D framework. *Reliability Engineering & System Safety*, 248, 109–124.

[20]. Zhou, M., Bansal, A., & AV Architecture Foundation. (2025). Interpretable bottlenecks: Blending deep representation learning with structured modular interfaces. *IEEE Robotics and Automation Letters*, 10(4), 3144–3156

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun1126

APPENDIX A

A sample of template for the data collection tables

➢ *Data Collection and Evaluation Templates*

To ensure systematic logging and reproducibility across the mixed-method framework, data is captured using three standardized master templates. These templates organize empirical simulator telemetry, qualitative regulatory assessments, and the final multi-objective synthesis.

- *Template 1: Quantitative Simulation Performance Matrix (nuScenes & CARLA)*

This template logs the empirical telemetry gathered during closed-loop simulation runs. It captures driving quality alongside computational and operational overhead.

Architecture Paradigm	**Trial ID**	**Planning Score ($Q_{plan\%}$)[1]**	**Collision Rate(%)[2]**	**Latency (Δtms)[3]**	**Data Efficiency (E_{data}GB)4**	**Debug Time (T_{debug}Hrs)5**	**ODD Adaptation (δ_{odd}%)6**	**Robustness Score ß%**
Modular Pipeline	#MOD-01							
Modular Pipeline	#MOD-02							
Pure End-to-End	#E2E-01							
Pure End-to-End	#E2E-02							
Hybrid Mod-E2E	#HYB-01							
Hybrid Mod-E2E	#HYB-02							

- *Parameter Operationalization:*

✓ (Q_{plan}): Combined score of route completion % and lane discipline, minus jerk penalties.
✓ Collision Rate: Percentage of simulation runs resulting in any physical contact.
✓ (Δt): Time elapsed from raw sensor input arrival to actuator command execution.
✓ (E_{data}): Size of the dataset required to reach a stable, non-oscillatory baseline.
✓ (T_{debug}): Average engineering hours needed to isolate and fix a logged system failure.
✓ (δ_{odd}): Performance retention percentage when deployed to an unmapped city environment.
✓ (R): Planning score stability under severe sensor noise or simulated weather extremes.

- *Template 2: Qualitative Regulatory Compliance Mapping*

This template evaluates the legal feasibility of each software design against active automotive safety standards.

Architecture Paradigm	**Component / Layer Evaluated**	**ISO 26262 Certifiability (Target: ASIL-D)**	**ISO/PAS 21448 (SOTIF) Risk Level**	**Traceability Type (Deterministic / Empirical)**	**Formal Verification Path (Feasible / Unfeasible)**	**Key Regulatory Roadblocks / Constraints**
Modular Pipeline	Perception Module					
Modular Pipeline	Rule-Based Planner					
Pure End-to-End	Full Unified Transformer					
Hybrid Mod-E2E	E2E Representation Backbone					
Hybrid Mod-E2E	Hardcoded Safety Rule Layer					

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun1126

- *Template 3: Mixed-Method Pareto-Optimal Optimization Synthesis*

This template aggregates data from the first two templates. It scores each system to determine the Pareto-optimal design for imminent commercial deployment.

Algorithmic Architecture	**Consolidated Driving Score (Q_{plan})Mean (%)^1**	**Safety & Certifiability Rating(S_{cert}) Tier)^2**	**Calculated L4 Validation Miles (\(N\) Miles)^3**	**Estimated 2026 Commercial Deployment Horizon**	**Pareto-Optimal Status (Yes / No)^4**
Modular Pipeline					
Pure End-to-End (E2E)					
Hybrid Modular-E2E					

- *Synthesis Parameters:*

✓ Consolidated Driving Score: The mean score across all nuScenes and CARLA testing scenarios.
✓ (S_{cert}) Tier: Categorized from Tier 1 (Fully Mathematically Certifiable to ASIL-D) to Tier 3 (Black-box / Uncertifiable).
✓ (N) Miles: Total required statistical validation miles based on the formula \(N = \ln(1 - C) / ln(1 - λ).
✓ Pareto-Optimal Status: Marked "Yes" if the system provides an optimal trade-off balance, meaning no other design delivers higher driving performance without violating safety certification deadlines