

Study on Data-Driven Scenario Construction for Autonomous Driving Testing

Arif Hossain

Abstract—Guaranteeing autonomous vehicle (AV) safety requires scalable evaluations that capture both everyday driving and critical edge cases. Scenario-based testing has become a key strategy, with scenario generation serving as its core. We classify generation techniques into three categories: rule-oriented, data-driven, and learning-enabled. For each, we examine representative approaches, simulation tools, description standards, and assessment metrics addressing fidelity, diversity, and risk exposure. Persistent issues include the simulation–reality gap, limited transferability, and the challenge of modeling rare events. Emerging directions such as language-guided generation, hybrid architectures, and open scenario repositories point toward more reliable and certifiable testing of AV systems.

Index Terms—Simulation, autonomous driving systems, autonomous vehicles.

I. INTRODUCTION

Autonomous Driving Systems (ADS) have rapidly advanced in recent years, fueled by breakthroughs in artificial intelligence, sensor technologies, and vehicular connectivity [1]–[3]. These systems are envisioned to reduce traffic accidents, improve efficiency, and enhance mobility across diverse populations. Yet, before large-scale deployment, ensuring their safety and reliability remains a formidable challenge [4]–[8]. A major barrier lies in the need for rigorous testing strategies capable of validating system performance under the diversity, uncertainty, and risk inherent in real-world conditions [9]–[12].

Public perception underscores these concerns. According to AAA's 2024 annual autonomous vehicle survey [13], 66% of U.S. drivers reported being afraid of fully self-driving vehicles, while only 9% expressed trust, with fears largely driven by highly publicized accidents and safety incidents. As illustrated in Fig. 1, such skepticism has remained persistent, signaling that safety concerns are not only technical but also societal barriers to adoption.

Testing therefore plays a central role in the development, validation, and regulatory approval of autonomous vehicles (AVs) [14]–[17]. Unlike conventional vehicles, AVs must independently perceive, predict, and act in dynamic environments, where unpredictable human behaviors, ambiguous road infrastructure, and rare but safety-critical events—such as sudden occlusions, erratic maneuvers, or sensor degradation—may occur [?, [18]–[23]. Traditional metrics based on mileage accumulation and disengagement rates are increasingly recognized as inadequate: real-world testing rarely captures statistically rare edge cases and often wastes resources on uneventful scenarios [24]–[27].

As a result, **scenario-based testing** has emerged as a more systematic and safety-oriented paradigm. This approach centers on constructing and executing targeted scenarios that represent both common driving contexts and rare but hazardous

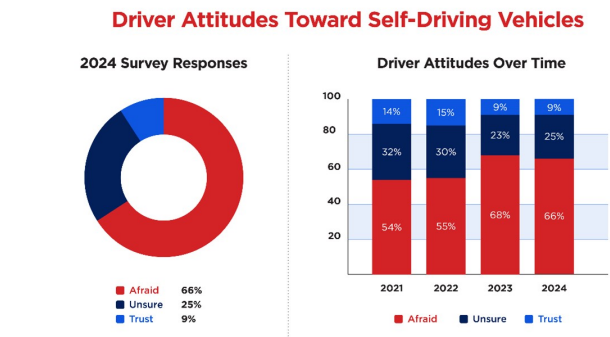


Fig. 1: Public attitudes toward self-driving vehicles in 2024. Source: AAA Newsroom, “Fear of Self-Driving Cars Persists,” March 28, 2024.

corner cases [28]–[30]. These include aggressive lane changes, pedestrian dart-outs, or multi-agent negotiation in dense traffic [31], [32]. Scenario-based testing enables structured evaluation of AV capabilities under uncertainty, while simulation platforms and formal scenario description languages (e.g., ASAM OpenSCENARIO, Scenic) facilitate repeatable, scalable, and transferable validation [33]–[35]. Regulatory frameworks such as ISO 21448 (SOTIF) and UNECE WP.29 are increasingly incorporating scenario-driven safety validation, underscoring its role in certification.

Within this context, **scenario generation** has become a critical enabler. Automated generation methods aim to synthesize diverse, realistic, and safety-critical test cases, going beyond manually curated or replayed datasets. These synthetic scenarios are especially valuable for stress-testing ADS against rare but high-risk events—such as near-miss collisions, unprotected turns, or aggressive cut-ins—that are unlikely to be encountered through brute-force road testing [36]–[39]. Recent research has explored reinforcement learning, Bayesian optimization, and other advanced algorithms to produce edge-case scenarios at scale [?, [?, [16], [40]–[51]. Such approaches enable closed-loop simulations that iteratively challenge AV decision-making under uncertainty [52]–[55].

This survey provides a comprehensive review of scenario generation techniques for AV safety testing. We classify existing methods into three main paradigms: **rule-based**, **data-driven**, and **learning-based** approaches. For each, we analyze their algorithmic foundations, scenario description languages, evaluation metrics (e.g., coverage, criticality, diversity), and supporting simulation tools. We then highlight research gaps, including generalization across domains, realism of simu-

lated environments, and the modeling of complex multi-agent interactions. Finally, we discuss emerging directions, such as semantic generation with large language models (LLMs), interactive multi-agent testing, and standardized scenario libraries for benchmarking [56], [57]. We aim for this survey to serve as a resource for researchers and practitioners striving to enhance the safety, reliability, and public trust of autonomous driving systems.

II. AUTONOMOUS DRIVING TESTING

Ensuring the safety and reliability of Autonomous Driving Systems (ADS) requires rigorous and systematic testing pipelines. Due to the complexity and unpredictability of real-world environments, conventional automotive testing methodologies are insufficient to evaluate the operational safety of AVs across their intended domain. To address this limitation, researchers and regulators have established multi-tiered testing frameworks that combine simulation, controlled track experiments, and on-road trials, thereby enabling comprehensive assessment under diverse conditions.

Testing strategies are commonly divided into three major categories: *simulation-based testing*, *closed-track testing*, and *on-road testing*, each with distinct contributions to the validation process.

Simulation-based testing leverages virtual environments to evaluate AV perception, planning, and control modules across a wide range of conditions. Platforms such as CARLA [58] and LGSVL [59] allow scalable, low-cost testing of scenarios including sudden occlusions, erratic maneuvers, and adverse weather or lighting [60]. Simulation ensures coverage of rare or hazardous cases that are often infeasible in physical testing. However, challenges remain in closing the sim-to-real gap, particularly in sensor modeling, environmental fidelity, and behavioral realism [61], [62].

Closed-track testing provides controlled, repeatable environments at facilities such as Mcity [63] and AstaZero [64]. It serves as a bridge between simulation and on-road validation, offering safe conditions to evaluate critical functionalities like automated emergency braking, lane changes, and V2X communication systems [65]. While track testing ensures physical realism and safety, it remains costly and limited in terms of environmental diversity and traffic complexity.

On-road testing represents the final and most realistic stage, exposing AVs to real traffic agents, unstructured environments, and regulatory contexts. Companies including Waymo, Cruise, and Baidu Apollo have reported millions of kilometers of road testing to validate their systems. Despite its necessity, on-road testing is constrained by the rarity of safety-critical events, limited statistical coverage, and significant liability risks [24], [66]. For this reason, it is typically preceded by extensive simulation and track-based testing.

Recent efforts emphasize hybrid validation pipelines that integrate these three levels: simulation for exploratory testing, closed tracks for controlled validation, and on-road deployment for certification. Table ?? provides a comparative summary of their trade-offs.

In the context of AV safety validation, it is essential to distinguish between three core concepts foundational to testing and scenario generation [67]:

- **Scene:** A *static snapshot* of the driving environment at a particular time, including road layout, infrastructure, traffic participants, environmental conditions, and the states of agents (e.g., position, velocity, heading). Scenes capture the spatial and contextual setup without temporal evolution.
- **Scenario:** A *temporal sequence* of scenes that models interactions among agents and unfolding events (e.g., overtaking, merging, emergency braking). Scenarios provide the basis for behavior modeling and safety testing [68].
- **Test Case:** A *parameterized instantiation* of a scenario, specifying initial configurations (e.g., positions, velocities, traffic density) along with measurable evaluation criteria. Test cases enable structured verification outcomes such as pass/fail or quantitative safety margins.

Such distinctions help formalize testing logic, improve comparability across datasets and tools, and ensure consistent interpretation in regulatory frameworks.

To support reproducible and certifiable ADS testing, a number of international standards and research initiatives have been proposed. These frameworks define methodologies for scenario modeling, simulation integration, and safety validation.

PEGASUS Project (Project for the Establishment of Generally Accepted Quality Criteria, Tools and Methods as well as Scenarios and Situations) is a German-led initiative that formalizes scenario-based testing pipelines. It emphasizes data-driven scenario derivation, parameterization using statistical distributions, execution in simulation, and coverage-driven evaluation [69]. PEGASUS has become a reference model in Europe for integrating simulation, track, and road testing with quantifiable safety arguments.

ISO 21448, also known as SOTIF (Safety of the Intended Functionality), complements ISO 26262 by addressing safety risks arising from limitations of perception and decision-making in complex or uncertain scenarios. It focuses on hazard identification and mitigation for failures not caused by hardware/software faults but by incomplete environment modeling or AI reasoning [69].

ASAM (Association for Standardisation of Automation and Measuring Systems) has developed open standards that enable interoperability across testing toolchains:

- **OpenSCENARIO:** Defines dynamic scenario logic, including actors, maneuvers, triggers, and actions.
- **OpenDRIVE:** Provides detailed road network geometry, topology, and infrastructure models.
- **OpenXOntology:** Establishes a shared semantic vocabulary to ensure consistent tool integration.

Additional efforts, such as ISO 34501/34502, introduce common terminologies, scenario taxonomies, and auditability criteria to ensure traceability in AV certification processes. Collectively, these frameworks mark a global shift toward

scenario-centric testing that emphasizes reproducibility, coverage, and proactive hazard exposure.

A. Rule-based Scenario Generation

Rule-based scenario generation constructs test cases based on predefined rules, templates, or expert knowledge. These methods often rely on regulatory conditions, structured parameter combinations, or safety-critical templates, offering high interpretability and traceability.

A classical line of work focuses on parameterized templates constrained by domain knowledge. Klischat and Althoff [70] proposed an approach that minimizes the solution space of the ego vehicle's motion planner using evolutionary algorithms such as Differential Evolution and Particle Swarm Optimization. Their framework effectively generates high-risk scenarios in multi-vehicle interactions and intersections, though computational efficiency decreases in high-dimensional settings due to collision constraints.

Althoff and Lutz [71] combined reachability analysis with optimization methods to synthesize collision-avoidance scenarios. Their method rapidly constructs scenarios within seconds, focusing on the safety envelope of the ego vehicle. However, it does not yet extend to multi-step trajectories or adversarial agent modeling.

To improve diversity and efficiency, Feng et al. [72] introduced the Adaptive Testing Scenario Library Generation (ATSLG) framework. Leveraging Bayesian Optimization with Gaussian Process Regression, ATSLG incrementally refines the scenario space, reducing required test iterations by one to two orders of magnitude. While effective in discovering critical scenarios, its scalability in high-dimensional feature spaces remains limited.

Gao et al. [73] proposed a combinatorial test generation method (CTBC) that integrates test matrices with a complexity-driven algorithm. By incorporating hierarchical influence modeling through AHP, their framework balances coverage and complexity. However, the accuracy of complexity estimation is constrained by underlying approximations.

More recently, Zhang et al. [74] extended rule-based methods by incorporating knowledge-enhanced scenario synthesis with large language models (LLMs). This approach aligns natural language intent with parameterized generation, enhancing flexibility while maintaining the structured backbone of rule-based design.

Overall, rule-based approaches remain widely adopted for their interpretability, repeatability, and suitability for regulatory testing. Nonetheless, they struggle to capture rare, emergent, or long-tail interactions and often lack adaptability in open-ended driving environments.

B. Data-driven Scenario Generation

Data-driven scenario generation exploits large-scale naturalistic driving datasets to extract, learn, and synthesize scenarios that reflect realistic traffic patterns, driver behaviors, and rare safety-critical cases. Unlike rule-based methods that rely on expert-crafted templates, this approach leverages statistical

analysis, machine learning, and data mining to uncover latent dynamics directly from empirical driving records. The availability of rich datasets such as NGSIM, Argoverse, and SHRP2 enables scalable creation of test cases that closely mirror real-world complexity.

A dominant line of work is **trajectory-based learning**. Zhang et al. [75] proposed *DP-TrajGAN*, a GAN-based framework augmented with differential privacy to synthesize realistic vehicle trajectories while preserving data confidentiality. Their method achieved strong fidelity and privacy trade-offs on NGSIM and Argoverse. Similarly, Krajewski et al. [76] combined GANs and VAEs to model diverse vehicle maneuvers, improving both scenario realism and simulation variability.

To generate more safety-critical conditions, researchers have explored **adversarial perturbation**. Wang et al. [77] introduced *AdvSim*, which perturbs trajectories and LiDAR signals to create adversarial test cases. Their results demonstrated the effectiveness of adversarial replay in uncovering system vulnerabilities.

Another important research thread focuses on **criticality assessment and targeting**. Westhofen et al. [78] provided a systematic review of criticality metrics and proposed a framework to evaluate their applicability for AV testing. Kang et al. [79] applied voxel-based 3D modeling and vision transformers to identify latent safety threats in LiDAR data, reporting an F1 score of 98.26% for risky zone detection.

Driving instability has also been employed as a data-driven indicator of crash risk. Arvin et al. [80] analyzed SHRP2 crash and near-crash events, demonstrating a strong correlation between pre-crash instability and severity. Such findings can guide the synthesis of high-risk scenarios for testing.

Rare and corner-case detection represents another active frontier. Bolte et al. [81] proposed a hybrid offline-online anomaly detection framework to identify low-frequency, high-impact events. This enables automatic harvesting of rare driving situations from continuous driving logs.

From a parameterization perspective, Muslim et al. [82] introduced a cut-out scenario generation method that abstracts highway traffic into interpretable parameter boundaries. This ensures both plausibility and controllable variability. Along similar lines, Huang et al. [83] presented the *CaDRE* framework, which integrates

III. EVALUATION

A critical component of scenario-based testing for autonomous driving is not only the generation of test scenarios but also their systematic evaluation. Without rigorous assessment, generated scenarios may fail to meaningfully contribute to safety validation. To this end, researchers have proposed a range of quality criteria and quantitative metrics to judge whether scenarios are realistic, diverse, comprehensive, and safety-critical. These evaluation dimensions determine how effectively scenarios expose weaknesses in the system under test and ensure their value in regulatory and industrial validation pipelines.

Scenario evaluation is generally structured around four core dimensions:

- **Realism:** Assesses how closely generated scenarios reflect real-world driving conditions, including plausible agent behavior, physically consistent motion dynamics, and compliance with traffic rules. Realism is crucial for external validity and is typically verified through statistical comparison with naturalistic datasets or human expert annotation [84]–[86].
- **Diversity:** Measures the variability among generated scenarios in terms of maneuvers, traffic densities, environmental conditions, and multi-agent interactions. High diversity enhances the likelihood of uncovering unexpected system failures and ensures broader stress-testing of ADS [87]–[91].
- **Coverage:** Captures how thoroughly scenarios span the operational design domain (ODD) and functional safety requirements. Coverage can be quantified by semantic labels (e.g., highway merging, unprotected left turns) or by parameter-space sampling strategies, ensuring systematic exploration of safety-relevant conditions [43], [45], [90], [92].
- **Criticality:** Evaluates the level of risk or difficulty posed by a scenario. Metrics include time-to-collision (TTC), minimum distance, required deceleration, and collision probability, all of which help identify scenarios most likely to trigger unsafe ADS responses [34], [88], [93]–[98].

To operationalize these metrics, effective evaluation requires advanced simulation environments that are realistic, flexible, and extensible. Such simulators allow controlled testing of perception, planning, and control modules at scale, while reducing the risks and costs of physical trials. In parallel, scenario description languages (e.g., OpenSCENARIO, Scenic) provide structured mechanisms to encode, manipulate, and standardize test cases. Together, high-fidelity simulators and formalized scenario languages form the backbone of scenario-based testing pipelines, ensuring that evaluation is both reproducible and transferable across different platforms and stakeholders.

IV. CHALLENGES AND RESEARCH GAPS

Despite significant progress in scenario generation for autonomous driving testing, several challenges and open research questions remain. These issues span technical, practical, and regulatory domains, highlighting the need for more robust, scalable, and standardized solutions. This section outlines the most critical challenges faced by current approaches and identifies gaps that present opportunities for future work.

A. Limited Data Diversity and Generalization

Most data-driven and learning-based generation techniques rely on real-world driving datasets, which, while extensive, often exhibit distributional bias. They tend to overrepresent common urban driving patterns and underrepresent edge cases such as near-misses, rare weather conditions, or unusual traffic interactions. As a result, generated scenarios may lack diversity and fail to generalize across unseen domains or geographies. Bridging this gap requires improved domain adaptation methods, cross-city or cross-country datasets, and techniques for transferring knowledge between different driving contexts.

Reality Gap in Synthetic Scenarios

A persistent challenge is the so-called *reality gap*—the mismatch between synthetic, simulator-generated scenarios and real-world driving environments. Even high-fidelity simulators may fail to capture subtle behaviors of pedestrians, occlusions, sensor noise, or infrastructure imperfections. This gap can lead to overestimation of AV performance in simulated testing and underpreparedness in real-world deployment. Addressing this issue involves combining real-world and simulated data, applying domain randomization, and improving simulator realism both at the perception and decision-making levels.

Scalability of Scenario Space

The scenario space for AV testing is practically infinite, encompassing a large number of interacting variables—agent types, behaviors, road layouts, environmental conditions, and temporal sequences. Exhaustive exploration of this space is infeasible. Thus, existing generation methods often sample from constrained subspaces, risking incomplete validation. New scalable methods for scenario space abstraction, semantic scenario clustering, and combinatorial coverage optimization are needed to ensure high testing efficiency without sacrificing thoroughness.

Modeling Safety-Critical but Rare Events

Safety-critical scenarios—such as sudden pedestrian crossings, aggressive merges, or multi-agent collisions—are rare but essential for validating robustness. However, they are underrepresented in data and difficult to model without manual engineering or adversarial optimization. Existing methods either rely on hand-crafted triggers or heuristic optimization, both of which may fail to cover unknown failure modes. Learning methods that can autonomously discover and amplify rare, high-risk patterns are still in their infancy and represent a vital area for advancement.

Standardization and Regulatory Alignment

A major obstacle in the deployment of scenario-based testing is the lack of unified standards for scenario modeling, evaluation, and exchange. While efforts like ASAM OpenSCENARIO and ISO 21448 (SOTIF) provide foundational standards, their integration into automated generation pipelines remains limited. Furthermore, regulatory acceptance of synthetic and generated scenarios for AV certification is still evolving. Research is needed on formal scenario verification, traceability, and compliance to ensure generated scenarios meet safety assurance requirements in a legally verifiable way.

V. EMERGING TRENDS AND FUTURE DIRECTIONS

As autonomous driving technologies mature, the demands on scenario generation frameworks are becoming more sophisticated. To support next-generation AV testing, recent research efforts have begun to explore novel paradigms that go beyond conventional rule-based or data-centric approaches. This section highlights several emerging trends and outlines promising future directions that are expected to shape the field in the coming years.

Semantic and Language-Driven Scenario Generation

The integration of large language models (LLMs) such as GPT and Claude into the AV testing pipeline has opened a new frontier: **semantic scenario generation**. Instead of specifying low-level scene parameters manually or learning them from data, users can now describe high-level scenarios in natural language—e.g., “a pedestrian suddenly crosses at a dimly lit intersection during rain”—which are then automatically translated into structured scenario code (e.g., OpenSCENARIO or Scenic). This paradigm enables more intuitive, human-centered interaction and lowers the barrier for specifying complex or rare situations.

Several early-stage systems now link language models with simulation backends (e.g., CARLA + LLM), enabling real-time scenario synthesis and editing. Future work may focus on integrating commonsense reasoning, legal constraints, and safety specifications directly into the generation pipeline through natural language interfaces.

Multi-modal and Multi-agent Scene Synthesis

Another active trend is the development of **multi-modal scenario synthesis** that incorporates visual, spatial, and behavioral information from multiple sources—such as video, LiDAR, maps, and text—to construct comprehensive test scenes. Generative models are being trained to combine these modalities into coherent environments, which better reflect the sensor fusion-based perception systems in real AVs.

In parallel, there is increasing interest in **multi-agent interaction modeling**. Modern urban scenarios often involve complex interactions among multiple agents with varying intent (e.g., pedestrians, cyclists, autonomous and human-driven vehicles). Modeling these interactions realistically, and generating coordinated behavior trajectories, remains a significant challenge. Multi-agent reinforcement learning, game-theoretic approaches, and diffusion-based generative models are emerging tools for tackling this complexity.

Hybrid Data-Driven and Rule-Based Approaches

Recognizing the limitations of pure rule-based or data-driven methods, researchers are moving towards **hybrid frameworks** that combine both strengths. Rule-based constraints provide safety and structure, while data-driven models contribute realism and diversity.

In practice, this might involve using data-driven models to sample base scenes and agents, with rule-based logic applied to inject specific intent, constraints, or triggers. Alternatively, hybrid approaches may operate in a layered architecture—where a symbolic planner outlines scenario semantics, and a learned module fills in the low-level details. These combinations are particularly promising for balancing interpretability with expressive power.

Towards Standardized Scenario Repositories and Benchmarks

As the field grows, there is a growing need for **open, standardized scenario repositories and benchmarking protocols**. Currently, many datasets and scenarios are either proprietary or fragmented, making reproducibility and comparative evaluation difficult.

Initiatives such as the **ASAM OpenX** family (OpenSCENARIO, OpenDRIVE, OpenLABEL) and projects like **GENEVA** or **SAFETAG** aim to unify scenario description formats and provide comprehensive libraries of validated test cases. Benchmarking tools that evaluate scenario quality, failure exposure, and coverage are also under active development.

In the future, publicly maintained scenario banks—similar to ImageNet for vision or GLUE for NLP—may become the cornerstone for training, testing, and certifying autonomous driving systems under global safety standards.

VI. CONCLUSION

Scenario generation has emerged as a central pillar in the validation and verification of autonomous driving systems. This survey has reviewed the landscape of scenario generation techniques from multiple perspectives, including rule-based, data-driven, and learning-based approaches. We have discussed key simulation platforms and scenario description languages that underpin automated testing pipelines, as well as the metrics used to evaluate scenario quality in terms of realism, diversity, criticality, and reproducibility.

From the reviewed literature and practices, several important takeaways can be drawn. Rule-based generation provides structure and standardization but struggles with diversity and scalability. Data-driven approaches benefit from realism grounded in real-world observations but are constrained by dataset limitations and rarity of critical events. Learning-based methods offer promising adaptability and automation, especially for generating adversarial or failure-triggering scenarios, but face challenges related to safety, interpretability, and validation.

Scenario quality evaluation remains a non-trivial task, requiring multidimensional metrics and feedback from simulation environments. Tools like Scenic and general coverage metrics are gaining traction in quantifying scenario space exploration. Moreover, the growing integration of scenario generation with closed-loop simulators enables dynamic, intelligent testing strategies that evolve alongside AV system development.

Looking forward, advancing scenario generation will be key to achieving safe and efficient autonomous driving. Future research should prioritize hybrid and semantic methods that balance structure with adaptability, develop standardized scenario libraries and benchmarking protocols, and close the realism gap between synthetic and real-world driving environments. Through collaborative efforts in research, tool development, and regulation, scenario-based testing will continue to evolve as a robust framework for ensuring safety in increasingly complex autonomous systems.

REFERENCES

- [1] Q. Zhang, K. Hua, Z. Zhang, Y. Zhao, and P. Chen, “Acnet: An attention-convolution collaborative semantic segmentation network on sensor-derived datasets for autonomous driving,” *Sensors*, vol. 25, no. 15, p. 4776, 2025.
- [2] Z. Wei, C. A. Gutierrez, J. Rodr, and J. Wang, “6g-enabled vehicle-to-everything communications: Current research trends and open challenges,” *IEEE Open Journal of Vehicular Technology*, 2025.

- [3] H. Kumar, P. Mamoria, and D. K. Dewangan, "Vision technologies in autonomous vehicles: progress, methodologies, and key challenges," *International Journal of System Assurance Engineering and Management*, 2025.
- [4] X. Liu, H. Huang, J. Bian, R. Zhou, and Z. Wei, "Generating intersection pre-crash trajectories for autonomous driving safety testing using transformer time-series gans," *Engineering Applications of Artificial Intelligence*, 2025.
- [5] H. X. Liu and S. Feng, "Curse of rarity for autonomous vehicles," *nature communications*, vol. 15, no. 1, p. 4808, 2024.
- [6] M. Zhao, C. Liang, T. Wang, J. Guan, and L. Wan, "Scenario hazard prevention for autonomous driving based on improved stpa," in *Safety, Reliability, and Security*. Springer, 2025.
- [7] R. Zhou, H. Huang, J. Lee, X. Huang, J. Chen, and H. Zhou, "Identifying typical pre-crash scenarios based on in-depth crash data with deep embedded clustering for autonomous vehicle safety testing," *Accident Analysis & Prevention*, vol. 191, p. 107218, 2023.
- [8] H. Huang, X. Huang, R. Zhou, H. Zhou, J. J. Lee, and X. Cen, "Pre-crash scenarios for safety testing of autonomous vehicles: A clustering method for in-depth crash data," *Accident Analysis & Prevention*, vol. 203, p. 107616, 2024.
- [9] A. A. B. da Costa, P. Irvine, T. Dodoiu, and S. Khashtgir, "Building a robust scenario library for safety assurance of automated driving systems: A review," *IEEE Transactions on Intelligent Transportation Systems*, 2025.
- [10] J. Lahikainen, "Ai-driven inverse method for identifying mechanical properties from small punch tests," Ph.D. dissertation, Aalto University, 2025.
- [11] R. Zhou, G. Zhang, H. Huang, Z. Wei, H. Zhou, J. Jin, F. Chang, and J. Chen, "How would autonomous vehicles behave in real-world crash scenarios?" *Accident Analysis & Prevention*, vol. 202, p. 107572, 2024.
- [12] J. Gu, M. Bellone, and A. Lind, "Camera-lidar fusion based object segmentation in adverse weather conditions for autonomous driving," in *2024 19th Biennial Baltic Electronics Conference (BEC)*. IEEE, 2024. [Online]. Available: <https://ieeexplore.ieee.org/document/10737955>
- [13] American Automobile Association, "Fear of self-driving cars persists," <https://newsroom.aacg.aaa.com/michigan-fear-of-self-driving-cars-persists/>, Mar. 2024, accessed: 2025-08-29.
- [14] S. Feng, H. Sun, X. Yan, H. Zhu, Z. Zou, S. Shen, and H. X. Liu, "Dense reinforcement learning for safety validation of autonomous vehicles," *Nature*, vol. 615, no. 7953, pp. 620–627, 2023.
- [15] S. Feng, X. Yan, H. Sun, Y. Feng, and H. X. Liu, "Intelligent driving intelligence test for autonomous vehicles with naturalistic and adversarial environment," *Nature communications*, vol. 12, no. 1, p. 748, 2021.
- [16] X. Huang, X. Cen, M. Cai, and R. Zhou, "A framework to analyze function domains of autonomous transportation systems based on text analysis," *Mathematics*, vol. 11, no. 1, p. 158, 2022.
- [17] J. M. Scanlon, K. D. Kusano, T. Daniel, C. Alderson, A. Ogle, and T. Victor, "Waymo simulated driving behavior in reconstructed fatal crashes within an autonomous vehicle operating domain," *Accident Analysis & Prevention*, vol. 163, p. 106454, 2021.
- [18] W. Huang, K. Wang, Y. Lv, and F. Zhu, "Autonomous vehicles testing methods review," in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2016, pp. 163–168.
- [19] H. Hungar, "Scenario-based validation of automated driving systems," in *International Symposium on Leveraging Applications of Formal Methods*. Springer, 2018, pp. 449–460.
- [20] N. Deldari, "Scenario annotation in autonomous driving: An outlier detection framework," Ph.D. dissertation, Uppsala University, 2025.
- [21] D. Pisinger and S. Ropke, "Large neighborhood search," in *Handbook of Metaheuristics*, M. Gendreau and J.-Y. Potvin, Eds. Boston, MA: Springer US, 2010, pp. 399–419.
- [22] Y. Zhou, Y. Sun, Y. Tang, Y. Chen, J. Sun, C. M. Poskitt, Y. Liu, and Z. Yang, "Specification-based autonomous driving system testing," *IEEE Transactions on Software Engineering*, vol. 49, pp. 3391–3410, 2023.
- [23] H. Zhang, J. Sun, and Y. Tian, "Accelerated risk assessment for highly automated vehicles: Surrogate-based monte carlo method," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, pp. 5488–5497, 2024.
- [24] N. Kalra and S. M. Paddock, "Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability?" *Transportation Research Part A*, 2016.
- [25] W. Ding, H. Lin, B. Li, and D. Zhao, "Causalaf: Causal autoregressive flow for safety-critical driving scenario generation," in *Conference on robot learning*. PMLR, 2023, pp. 812–823.
- [26] C. Li, J. Sifakis, Q. Wang, R. Yan, and J. Zhang, "Simulation-based validation for autonomous driving systems," in *Proceedings of the 32nd ACM SIGSOFT International Symposium on Software Testing and Analysis*, 2023, pp. 842–853.
- [27] B. Klamann, M. Lippert, C. Amersbach, and H. Winner, "Defining pass-/fail-criteria for particular tests of automated driving functions," *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, vol. null, pp. 169–174, 2019.
- [28] D. Nalic, T. Mihalj, M. Bäuml, M. Lehmann, A. Eichberger, and S. Bernsteiner, "Scenario based testing of automated driving systems: A literature survey," in *FISITA web Congress*, vol. 10, 2020.
- [29] "Finding critical scenarios for automated driving systems: A systematic mapping study," *IEEE Transactions on Software Engineering*, vol. 49, pp. 991–1026, 2023.
- [30] N. Sahu, A. Bhat, and R. Rajkumar, "Saferoute: Risk-minimizing cooperative real-time route and behavioral planning for autonomous vehicles," in *2024 IEEE 27th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2024. [Online]. Available: <https://ieeexplore.ieee.org/document/10919756>
- [31] R. Nagy and I. Szalai, "Development of an unsupervised learning-based annotation method for road quality assessment," *Transportation Engineering*, 2025.
- [32] T. Menzel, G. Bagschik, and M. Maurer, "Scenarios for development, test and validation of automated vehicles," in *2018 IEEE intelligent vehicles symposium (IV)*. IEEE, 2018, pp. 1821–1827.
- [33] T. Menzel, M. Bagschik, and M. Maurer, "Scenarios for safety validation of highly automated vehicles," *Transportation Research Part F*, 2020.
- [34] J. Sun, H. Zhang, H. Zhou, R. Yu, and Y. Tian, "Scenario-based test automation for highly automated vehicles: A review and paving the way for systematic safety assurance," *IEEE transactions on intelligent transportation systems*, vol. 23, no. 9, pp. 14088–14103, 2021.
- [35] M. Fiorino, M. Naem, M. Ciampi, and A. Coronato, "Defining a metric-driven approach for learning hazardous situations," *Technologies*, vol. 12, no. 7, p. 103, 2024.
- [36] M. Mammen, Z. Kayatas, and D. Bestle, "Evaluation of different generative models to support the validation of advanced driver assistance systems," *Applied Mechanics*, 2025. [Online]. Available: <https://www.mdpi.com/2673-3161/6/2/39>
- [37] X. Zhang, L. Xiong, P. Zhang, J. Huang, and Y. Ma, "Real-world troublemaker: A 5g cloud-controlled track testing framework for automated driving systems in safety-critical interaction scenarios," *arXiv preprint arXiv:2502.14574*, 2025.
- [38] J. Wang and X. Wang, "Safety-critical scenario generation for self-driving systems based on domain models," in *2nd International Conference on Intelligent Robotics and Control Engineering*, 2025. [Online]. Available: <https://ieeexplore.ieee.org/document/11011061>
- [39] R. Zhou, H. Huang, J. Lee, X. Huang, J. Chen, and H. Zhou, "Identifying typical pre-crash scenarios based on in-depth crash data with deep embedded clustering for autonomous vehicle safety testing," *Accident Analysis & Prevention*, vol. 191, p. 107218, Oct 2023.
- [40] C. Xu, Z. Ding, C. Wang, and Z. Li, "Statistical analysis of the patterns and characteristics of connected and autonomous vehicle involved crashes," *Journal of Safety Research*, vol. 71, pp. 41–47, Dec 2019.
- [41] J. Lenard, "Typical pedestrian accident scenarios for the development of autonomous emergency braking test protocols," *Accident Analysis & Prevention*, p. 8, 2014.
- [42] R. Zhou, Y. Liu, K. Zhang, and O. Yang, "Genetic algorithm-based challenging scenarios generation for autonomous vehicle testing," *IEEE Journal of Radio Frequency Identification*, vol. 6, pp. 928–933, 2022.
- [43] B. Zhu, Y. Sun, J. Zhao, J. Han, P. Zhang, and T. Fan, "A critical scenario search method for intelligent vehicle testing based on the social cognitive optimization algorithm," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 8, pp. 7974–7986, 2023.
- [44] R. Zhou, "Efficient safety testing of autonomous vehicles via adaptive search over crash-derived scenarios," *arXiv preprint arXiv:2508.06575*, 2025.
- [45] J. Bian, H. Huang, Q. Yu, and R. Zhou, "Search-to-crash: Generating safety-critical scenarios from in-depth crash data for testing autonomous vehicles," *Energy*, p. 137174, 2025.
- [46] J. Sun, H. Zhang, H. Zhou, R. Yu, and Y. Tian, "Scenario-based test automation for highly automated vehicles: A review and paving the way for systematic safety assurance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 9, pp. 14088–14103, Sep 2022.
- [47] Y. Tian, W. Zheng, Y. Shao, H. Zhang, and J. Sun, "Mjtg: A multi-vehicle joint trajectory generator for complex and rare scenarios," *IEEE Transactions on Vehicular Technology*, 2025.

- [48] A. Mondelli, Y. Li, A. Zanardi, and E. Frazzoli, "Test automation for interactive scenarios via promptable traffic simulation," *arXiv preprint arXiv:2506.01199*, 2025.
- [49] Y. Mei, T. Nie, J. Sun, and Y. Tian, "Llm-attacker: Enhancing closed-loop adversarial scenario generation for autonomous driving with large language models," *arXiv preprint arXiv:2501.15850*, 2025.
- [50] Z. Zeng, Q. Shi, W. Zhuang, X. Wang, and X. Fan, "Adversarial generation for autonomous vehicles in safety-critical ramp merging scenarios," in *International Conference on Electric Vehicle and Vehicle Engineering*. Springer, 2024, pp. 427–434.
- [51] S. Feng, Y. Feng, H. Sun, Y. Zhang, and H. X. Liu, "Testing scenario library generation for connected and automated vehicles: An adaptive framework," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 2, pp. 1213–1222, 2020.
- [52] S. Tang, Z. Zhang, J. Zhou, L. Lei, Y. Zhou, and Y. Xue, "Legend: A top-down approach to scenario generation of autonomous driving systems assisted by large language models," in *Proceedings of the 39th IEEE/ACM International Conference on Automated Software Engineering*, 2024, pp. 1497–1508.
- [53] M. Arnav, "Scenario generation methods for functional safety testing of automated driving systems," 2025.
- [54] Y. Huai, S. Almane, Y. Chen, X. Wu, Q. A. Chen, and J. Garcia, "scenorta: Generating diverse, fully-mutable, test scenarios for autonomous vehicle planning," *IEEE Transactions on Software Engineering*, pp. 1–21, 2023.
- [55] W. Ding, H. Lin, B. Li, and D. Zhao, "Generalizing goal-conditioned reinforcement learning with variational causal reasoning," *Advances in Neural Information Processing Systems*, vol. 35, pp. 26 532–26 548, 2022.
- [56] X. Cai, X. Bai, Z. Cui, D. Xie, D. Fu, H. Yu, and Y. Ren, "Text2scenario: Text-driven scenario generation for autonomous driving test," *arXiv preprint arXiv:2503.02911*, 2025.
- [57] C. Ricotta, S. Khzym, A. Faron, and A. Emadi, "Property optimized gnn: Improving data association performance using cost function optimization for sensor fusion in high density environments," in *2024 IEEE Smart World Congress (SWC)*. IEEE, 2024, pp. 1871–1877.
- [58] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "Carla: An open urban driving simulator," in *Conference on robot learning*. PMLR, 2017, pp. 1–16.
- [59] G. Rong, B. H. Shin, H. Tabatabaee, Q. Lu, S. Lemke, M. Mozeiko, E. Boise, G. Uhm, M. Gerow, S. Mehta *et al.*, "Lgsvl simulator: A high fidelity simulator for autonomous driving," in *2020 IEEE 23rd International conference on intelligent transportation systems (ITSC)*. IEEE, 2020, pp. 1–6.
- [60] R. Maier, L. Grabinger, D. Urlhart, and J. Mottok, "Causal models to support scenario-based testing of adas," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 2, pp. 1815–1831, 2023.
- [61] D. J. Fremont, E. Kim, Y. V. Pant, S. A. Seshia, A. Acharya, X. Bruso, P. Wells, S. Lemke, Q. Lu, and S. Mehta, "Formal scenario-based testing of autonomous vehicles: From simulation to the real world," in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2020, pp. 1–8.
- [62] H. Y. Sonetta, "Bridging the simulation-to-reality gap: Adapting simulation environment for object recognition," Master's thesis, University of Windsor (Canada), 2021.
- [63] Y. Dong, Y. Zhong, W. Yu, M. Zhu, P. Lu, Y. Fang, J. Hong, and H. Peng, "Mcity data collection for automated vehicles study," *arXiv preprint arXiv:1912.06258*, 2019.
- [64] J. Jacobson, P. Janevik, and P. Wallin, "Challenges in creating astazero, the active safety test area," in *Transport Research Arena (TRA) 5th Conference: Transport Solutions from Research to Deployment*. European Commission/Conference of European Directors of Roads (CEDR) European Road Transport Research Advisory Council (ERTRAC), 2014.
- [65] Y. Ma, C. Sun, J. Chen, D. Cao, and L. Xiong, "Verification and validation methods for decision-making and planning of automated vehicles: A review," *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 3, pp. 480–498, 2022.
- [66] R. Mariani, "An overview of autonomous vehicles safety," in *2018 IEEE International Reliability Physics Symposium (IRPS)*. IEEE, 2018, pp. 6A–1.
- [67] International Organization for Standardization, "ISO 34501:2022 - Road vehicles — Test scenarios for automated driving systems — Vocabulary," 2022.
- [68] F. Batsch, S. Kanarachos, M. Cheah, R. Ponticelli, and M. Blundell, "A taxonomy of validation strategies to ensure the safe operation of highly automated vehicles," *Journal of Intelligent Transportation Systems*, vol. 26, no. 1, pp. 14–33, 2021.
- [69] C. Wang, K. Storms, N. Zhang, and H. Winner, "Runtime unknown unsafe scenarios identification for sofif of autonomous vehicles," *Accident Analysis & Prevention*, vol. 195, p. 107410, 2024.
- [70] M. Klischat and M. Althoff, "Generating critical test scenarios for automated vehicles with evolutionary algorithms," in *IEEE Intelligent Vehicles Symposium (IV)*, 2019.
- [71] M. Althoff and S. Lutz, "Automatic generation of safety-critical test scenarios for collision avoidance of road vehicles," in *IEEE Intelligent Vehicles Symposium (IV)*, 2018.
- [72] S. Feng, Y. Feng, H. Sun, Y. Zhang, and H. X. Liu, "Testing scenario library generation for connected and automated vehicles: An adaptive framework," *IEEE Transactions on Intelligent Transportation Systems*, 2020.
- [73] F. Gao, J. Duan, Y. He, and Z. Wang, "A test scenario automatic generation strategy for intelligent driving systems," *Mathematical Problems in Engineering*, 2019.
- [74] J. Zhang, C. Xu, and B. Li, "Chatscene: Knowledge-enabled safety-critical scenario generation for autonomous vehicles," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.
- [75] J. Zhang, Q. Huang, and Y. Huang, "Dp-trajgan: A privacy-aware trajectory generation model with differential privacy," *Future Generation Computer Systems*, 2023.
- [76] R. Krajewski, T. Moers, and D. Nerger, "Data-driven maneuver modeling using generative adversarial networks and variational autoencoders for safety validation of highly automated vehicles," in *21st International Conference on Intelligent Transportation Systems (ITSC)*, 2018.
- [77] J. Wang, A. Pun, and J. Tu, "Advsim: Generating safety-critical scenarios for self-driving vehicles," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.
- [78] L. Westhofen, C. Neurohr, and T. Koopmann, "Criticality metrics for automated driving: A review and suitability analysis of the state of the art," *Archives of Computational Methods in Engineering*, 2023.
- [79] M. Kang, J. Seo, and K. Hwang, "Critical voxel learning with vision transformer and derivation of logical av safety assessment scenarios," *Accident Analysis & Prevention*, 2024.
- [80] R. Arvin, M. Kamrani, and A. J. Khattak, "The role of pre-crash driving instability in contributing to crash intensity using naturalistic driving data," *Accident Analysis & Prevention*, 2019.
- [81] J. A. Bolte, A. Bar, and D. Lipinski, "Towards corner case detection for autonomous driving," in *IEEE Intelligent Vehicles Symposium*, 2019.
- [82] H. Muslim, S. Endo, and H. Imanaga, "Cut-out scenario generation with reasonability foreseeable parameter range from real highway dataset for autonomous vehicle assessment," *IEEE Access*, 2023.
- [83] P. Huang, W. Ding, and J. Francis, "Cadre: Controllable and diverse generation of safety-critical driving scenarios using real-world trajectories," *arXiv preprint arXiv:2401.XXXX*, 2024.
- [84] R. Zhou, Z. Lin, G. Zhang, H. Huang, H. Zhou, and J. Chen, "Evaluating autonomous vehicle safety performance through analysis of pre-crash trajectories of powered two-wheelers," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 10, pp. 13 560–13 572, 2024.
- [85] R. Zhou, W. Gui, H. Huang, X. Liu, Z. Wei, and J. Bian, "Diffcrash: Leveraging denoising diffusion probabilistic models to expand high-risk testing scenarios using in-depth crash data," *Expert Systems with Applications*, p. 128140, 2025.
- [86] G. Zhang, H. Huang, R. Zhou, S. Li, and J. Bian, "High-risk trajectories generation for safety testing of autonomous vehicles based on in-depth crash data," *IEEE Transactions on Intelligent Transportation Systems*, 2025.
- [87] B. B. Oliveira, M. A. Carravilla, and J. F. Oliveira, "A diversity-based genetic algorithm for scenario generation," *European journal of operational research*, vol. 299, no. 3, pp. 1128–1141, 2022.
- [88] F. Batsch, S. Kanarachos, M. Cheah, R. Ponticelli, and M. Blundell, "A taxonomy of validation strategies to ensure the safe operation of highly automated vehicles," *Journal of Intelligent Transportation Systems*, vol. 26, pp. 14 – 33, 2020.
- [89] Q. Chu, Y. Yue, D. Yao, and H. Pei, "Dicritest: Testing scenario generation for decision-making agents considering diversity and criticality," *arXiv preprint arXiv:2508.11514*, 2025.
- [90] W. Ding, C. Xu, M. Arief, H. Lin, B. Li, and D. Zhao, "A survey on safety-critical driving scenario generation—a methodological perspective," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 7, pp. 6971–6988, 2023.
- [91] R. Zhou, H. Huang, G. Zhang, H. Zhou, and J. Bian, "Crash-based safety testing of autonomous vehicles: Insights from generating safety-critical scenarios based on in-depth crash data," *IEEE Transactions on Intelligent Transportation Systems*, 2025.

- [92] R. Zhou, Z. Lin, X. Huang, J. Peng, and H. Huang, "Testing scenarios construction for connected and automated vehicles based on dynamic trajectory clustering method," in *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2022, pp. 3304–3308.
- [93] S. Li, R. Zhou, and H. Huang, "Multidimensional evaluation of autonomous driving test scenarios based on ahp-ewm-topsis models," *Automotive Innovation*, pp. 1–15, 2025.
- [94] Z. Wei, H. Huang, G. Zhang, R. Zhou, X. Luo, S. Li, and H. Zhou, "Interactive critical scenario generation for autonomous vehicles testing based on in-depth crash data using reinforcement learning," *IEEE Transactions on Intelligent Vehicles*, 2024.
- [95] X. Luo, Z. Wei, G. Zhang, H. Huang, and R. Zhou, "High-risk powered two-wheelers scenarios generation for autonomous vehicle testing using wgan," *Traffic Injury Prevention*, vol. 26, no. 2, pp. 243–251, 2025.
- [96] Z. Wei, H. Zhou, and R. Zhou, "Risk and complexity assessment of autonomous vehicle testing scenarios," *Applied Sciences*, vol. 14, no. 21, p. 9866, 2024.
- [97] Z. Wei, J. Bian, H. Huang, R. Zhou, and H. Zhou, "Generating risky and realistic scenarios for autonomous vehicle tests involving powered two-wheelers: A novel reinforcement learning framework," *Accident Analysis & Prevention*, vol. 218, p. 108038, 2025.
- [98] S. Tang, Z. Zhang, Y. Zhang, J. Zhou, Y. ling Guo, S. Liu, S. Guo, Y. Li, L. Ma, Y. Xue, and Y. Liu, "A survey on automated driving system testing: Landscapes and trends," *ACM Transactions on Software Engineering and Methodology*, vol. 32, pp. 1 – 62, 2022.