Proc. of the European Safety and Reliability Conference (ESREL 2026)
Edited by José C. Matos, Paulo B. Lourenço, Michael Beer, Daniel Oliveira,
Edoardo Patelli, Hélder S. Sousa, Rui A. Silva and Monica Santamaria
 Published by Research Publishing, Singapore.
doi: 10.3850/ESREL2026061419_esrel26-p26154-cd

BAAS: Benchmarking Adversarial Agent Strategies - A Comparative Study of Gradient-based, Reinforcement, and Evolutionary Paradigms for Safety-Critical Scenario Generation

Carmen Mei-Ling Frischknecht-Gruber

School of Engineering & Department of Informatics, Zurich University of Applied Sciences & University of Fribourg, Switzerland. E-mail: frsh@zhaw.ch

Prof. Dr. Monika Reif

School of Engineering, Zurich University of Applied Sciences, Switzerland. E-mail: reif@zhaw.ch

Prof. Dr. Andreas Fischer

Department of Informatics, University of Fribourg, Switzerland. E-mail: andreas.fischer@unifr.ch

The verification and validation of autonomous driving systems demand systematic exposure to safety-critical interactions that reveal weaknesses in control policies. Traditional verification methods rely on parameter sweeps or fixed test suites, which offer limited coverage of rare but high-risk situations. Recent adversarial approaches increase scenario criticality, i.e., generate more challenging and risky conditions, but often prioritize collision occurrence over behavioral plausibility or diversity. This work investigates the generation of adversarial agents as a reproducible and scalable approach to create diverse yet coherent safety-critical driving interactions. Five paradigms are compared: (1) parameter sweeps as baseline, (2) gradient-based optimization, (3) reinforcement-learning (RL) adversaries, (4) quality-diversity (QD) search via MAP-Elites and QD-RL. A pre-trained driving policy serves as the ego under test in a highway-driving environment. Each method is evaluated on three dimensions: effectiveness, diversity, and behavioral plausibility and feasibility. Additionally, we analyze the impact of single-adversary setups and discuss extensions toward multi-adversary configurations and hybrid curricula combining MAP-Elites seeding with adversarial RL. The study quantifies trade-offs between adaptiveness, diversity and feasibility in the generation of safety-critical scenarios and identifies mechanisms for exposing weaknesses in robust control policies. The framework provides a structured comparison of learning-based, evolutionary and gradient-based paradigms, supporting systematic robustness evaluation of autonomous driving functions and the creation of reusable adversarial scenario databases for future studies.

Keywords: Scenario generation, autonomous systems, testing, reinforcement learning, machine learning

1. Introduction

Autonomous driving systems require rigorous verification and validation (V&V) to ensure reliable behavior under diverse and safety-critical operational conditions (International Organization for Standardization (2022, 2018)). Although large fleets now accumulate data of millions of real-world kilometers driven (California Department of Motor Vehicles (2024); Waymo LLC (2020)), the frequency of genuinely hazardous corner cases remains extremely low. As a result, existing datasets, despite their scale, do not sufficiently capture rare but safety-relevant interactions (Kalra and Paddock (2016)). Simulation, therefore, plays a central role in exposing autonomous agents to controlled, repeatable, and systematically varied safety-critical situations. However, constructing such scenarios in a principled manner remains a challenge. A central limitation is that standard scenario catalogs are inherently incomplete. By design, catalogs enumerate only known classes of events and rely strongly on expert judgment, which introduces bias, restricts variability, and prevents systematic coverage of unforeseen failure modes (Ding et al. (2023)). This incompleteness creates a validation gap that directly affects system robustness and limits the ability to assess autonomous controllers under realistic uncer-

tainty. Recent generative approaches synthesize scenes via architectures like GANs or diffusion models (Yan et al. (2023); Xu et al. (2025)), but often lack diversity or generate physically inconsistent scenarios (Ding et al. (2023); Ramakrishna et al. (2022)). Adversarial approaches explicitly search for agent behaviors that expose weaknesses in ego policies by optimizing interactions within a simulation environment. Yet many existing adversarial methods optimize narrowly for collision occurrence, resulting in implausible trajectories, limited diversity of failure modes, and a lack of reproducibility across repeated evaluations (Doreste (2025)). To the best of our knowledge, no unified comparative analysis has yet examined gradient-based reinforcement learning (RL) and evolutionary or Quality Diversity (QD)-based adversarial agent paradigms within a common experimental setting. This raises a central research question: *How do different adversarial-agent paradigms differ in their ability to induce failures in autonomous-driving control policies, and what trade-offs emerge between adaptiveness, diversity, and reproducibility?* Adversarial agents offer a promising approach to systematic stress testing. These agents act as controlled counterparts to the ego vehicle, actively probing its decision-making to reveal hidden flaws, create targeted variations of safety-critical scenarios, and uncover classes of failures unlikely to arise in random or expert-defined testing. Their systematic integration into V&V pipelines requires a principled evaluation protocol that ensures comparability across methods and supports reproducible scenario libraries. The framework discourages lowest-hanging-fruit exploitation by penalizing trivial early collisions and nuisance behaviors, and by evaluating adversaries over scenario families rather than single worst-case rollouts. To address these gaps, we introduce a unified experimental framework for benchmarking adversarial-agent generation strategies in highway-driving scenarios. We evaluate five paradigms under identical conditions: a classical parameter sweep (baseline), a gradient-based adversary, an RL adversary, and an evolutionary adversary based on QD search. To ensure reproducible evaluation, all adversaries are tested against a pre-trained, high-performance driving policy. Their performance is assessed along the following key dimensions: effectiveness in inducing failures, behavioral diversity, behavioral plausibility and feasibility, and reproducibility across repeated runs. In addition, we analyze single-agent adversarial setups and outline extensions toward multi-agent adversarial configurations. This comparative analysis characterizes the strengths, limitations, and trade-offs of different adversarial paradigms. The framework supports constructing reusable adversarial scenario libraries.

2. Related Work

Recent surveys underline the need for automated, systematic generation of safety-critical scenarios, emphasizing scalability, controllability, and reproducibility as persistent limitations of current practice (Wäschle et al. (2022)).

Adversarial approaches directly target failure-inducing interactions by optimizing environment configurations, agent behaviors, or trajectory perturbations. Gradient-based methods, such as KING (Hanselmann et al. (2022)), exploit differentiable proxy dynamics to adjust interacting agents efficiently and produce plausible collision scenarios that degrade imitation-learning policies under tightly controlled conditions. Extensions to real-world datasets, such as ReGentS, demonstrate that trajectory optimization combined with differentiable simulation can maintain realism while identifying meaningful adversarial behaviors in logged urban traffic (Yin et al. (2024)). Other frameworks, including ANTI-CARLA by Ramakrishna et al. (2022), focus on perturbing environmental and sensor conditions to identify brittle operating regions in complex pipelines. Evolutionary and fuzzing-based strategies, such as those underlying AV-Fuzzer-type approaches, mutate scenario parameters to expose failures, while recent diffusion-based generators steer sampling toward constraint-compliant yet safety-critical situations (Xu et al. (2025)). Although these methods reveal vulnerabilities that remain hidden in conventional testing, they typically focus on a single paradigm, rely on heterogeneous evalu-

ation metrics, and provide limited insight into behavioral diversity or reproducibility across repeated runs. Generative and text-driven scenario construction frameworks provide complementary capabilities. Systems such as SLEDGE by Chitta et al. (2024) use diffusion-based models to synthesize large, structurally coherent driving scenes from data, supporting controllable route- or lane-conditioned testing. TARGET, Txt2Sce, and OmniTester (Deng et al. (2023); Ji et al. (2025); Lu et al. (2025)) convert natural language rules, textual descriptions, or accident narratives into executable simulation scenarios, thereby reducing manual authoring effort and linking high-level requirements to simulation configurations. While these approaches are valuable in terms of coverage and traceability, they do not construct adversarial agents and rarely evaluate the ability of a scenario to induce policy failures in closed-loop settings. Artificial Life (ALife) approaches, such as QD, demonstrate how structured exploration of behavioral or environmental spaces can uncover challenging yet solvable tasks. The utilization of Multi-dimensional Archive of Phenotypic Elites (MAP) based approaches, latent-space evolution, and domain-specific examples has been shown to be beneficial for maintaining diverse environments, with behavior descriptors incorporated to guide the search (Mouret and Clune (2015); Pugh et al. (2016)). ALife-inspired environment shaping, such as evolving spatial viability landscapes, further shows how adaptive pressure can be embedded into the environment itself to produce progressively more demanding conditions Gaylinn et al. (2025). Despite their relevance for discovering broad ranges of failure modes, these methods have seen limited application in safety-critical systems and have not been benchmarked against gradient-based or RL adversaries under common conditions. Existing work provides strong individual paradigms: gradient-based optimization, RL-driven adversaries, evolutionary and fuzzing strategies, generative simulators, and text-driven scenario construction. These are typically evaluated in isolation, use incompatible metrics or simulators, and do not assess trade-offs between effectiveness, behavioral diversity, and reproducibility.

3. Methodology

All experiments are conducted in the `highway-env` simulator [a], configured with fixed lane geometry, traffic density, simulation frequency, and deterministic seeds. Each episode has a horizon H, and the ego vehicle interacts with one adversarial vehicle and fixed traffic under identical conditions.

Ego policy The ego vehicle uses a deep Q-network (DQN) with a CNN encoder trained on stacked grayscale renderings of the highway environment. Once trained, this policy is frozen and reused across all experiments, ensuring that the evaluation of adversarial paradigms is not confounded by changes in the ego controller.

Adversarial interface Each adversarial agent generates an action sequence

$$\mathbf{a}_{1:H} = (a_1, a_2, \ldots, a_H), \tag{1}$$

where $a_t \in \mathcal{A}$ is a discrete meta-action executed in `highway-env`. We use the default discrete meta-action set of `highway-env`, consisting of lane-change and longitudinal speed actions.

State transitions follow

$$s_{t+1} = f\big(s_t, \pi_{\text{ego}}(s_t), a_t\big), \tag{2}$$

with termination at collision, off-road events, or timeout. Time-to-critical-incident (TTCI) is triggered by terminal ego collisions, off-road events, or predefined proximity-based risk thresholds.

Constraints and plausibility boundaries Adversarial actions are bounded by speed, steering, and acceleration limits defined by the simulator. Vehicles must remain on-road until termination, and non-physical maneuvers are not permitted. These constraints ensure comparability between paradigms and define the admissible action envelope, while behavioral plausibility is assessed post hoc via trace diagnostics.

Reproducibility All experiments use fixed seeds, logged initial configurations, and replayable trajectories. Each adversarial run can be evaluated deterministically under identical conditions. For parameter-sweep and MAP-Elites–based adversaries, evaluation is performed

[a] https://highway-env.farama.org/

at the family level using a fixed ego policy, aggregating outcomes across systematic perturbations encoded in shared rollout specifications that are used consistently across all paradigms. We refer to a scenario family as a set of rollouts generated from a common base seed with systematic perturbations of adversarial initial conditions or actions.

3.1. *Paradigms*

All paradigms share the executed action interface in Eq. (1) with $a_t \in \mathcal{A}$ and the simulator dynamics in Eq. (2).

Parameter sweeps A deterministic, non-learning baseline that probes predefined maneuver families by varying parameters such as lateral offset, braking intensity, or timing. The parameter sweep does not adapt online to the ego policy and serves as a reproducible non-learning baseline.

Gradient-based optimization Following KING (Hanselmann et al. (2022)), gradient-based adversaries optimize a proxy control sequence $\mathbf{u}_{1:H}$ by minimizing a distance-based adversarial objective:

$$\mathcal{L}(\mathbf{u}) = \sum_{t=1}^{H} w_t \cdot \ell\left(d_{\text{ego,adv}}(t)\right) + \mathcal{R}(\mathbf{u}) \quad (3)$$

where ℓ is a distance-based potential, w_t represents temporal weighting to prioritize interactions, and $\mathcal{R}(\mathbf{u})$ regularizes against degenerate solutions (e.g., trivial early collisions or stalled ego progress).The gradient $\nabla_{\mathbf{u}}\mathcal{L}$ is computed through a differentiable proxy of the vehicle dynamics, enabling efficient local search. This paradigm is highly adaptive but can converge to relatively narrow classes of solutions due to local minima. Neither the objective nor the proxy directly optimizes for terminal collisions, these arise implicitly from sustained close interactions.

Reinforcement-learning adversaries RL-based adversaries treat the problem as a sequential decision-making task. In our Proximal Policy Optimization (PPO) setup, the adversary receives dense proximity shaping, a small per-step time cost, a terminal reward for ego collisions (down-weighted if the collision occurs too early), and a penalty for adversary-only collisions:

$$\begin{aligned} r_t \;&= w_p\, \psi\big(d_{\text{ego,adv}}(t)\big) - c_{\text{time}} \\ &\quad + \mathbb{I}[\text{ego\_col}(t)]\, r_{\text{term}}(t) \\ &\quad - \mathbb{I}[\text{adv\_col}(t) \wedge \neg\text{ego\_col}(t)]\, c_{\text{nuis}}\big(d_{\text{ego,adv}}(t)\big) \end{aligned} \quad (4)$$

where: $\psi(\cdot)$ is a distance-based proximity shaping term (implemented as $\psi(d) = 1/(1+d)$), encouraging sustained close interaction between the ego vehicle and the adversary. $r_{\text{term}}(t)$ is a terminal ego-collision reward. If the collision occurs before a minimum time-to-failure threshold $t < t_{\min}$, we down-weight it via an additional early-collision penalty, otherwise, a larger terminal reward is granted. In our implementation, the PPO adversary uses fixed proximity shaping, time penalties, terminal collision rewards, and nuisance penalties. All coefficients are held constant across runs and reported in the released configuration.

Quality-Diversity Reinforcement Learning (QD-RL) extends the RL adversary by embedding policy learning within a QD framework. Instead of maintaining a single adversarial policy, MAP-Elites is used to construct an archive in which each cell stores a distinct trained adversary policy. Policies are trained under descriptor-conditioned objectives or initial conditions, encouraging both adversarial effectiveness and behavioral diversity using the descriptors defined below. Unlike fixed action-sequence search, QD-RL adversaries adapt through learning during training, while diversity is enforced at the population level via behavioral descriptors.

Quality Diversity (MAP-Elites) We implement a QD adversarial generator based on MAP-Elites (Mouret and Clune, 2015; Pugh et al., 2016). The adversary is not a trained policy but a parameterized open-loop controller: a fixed-length discrete action sequence $\mathbf{a}_{1:H} \in \mathcal{A}^H$. To reduce dimensionality, we compress the sequence into B blocks of length L (with $H = B \cdot L$), such that the genome is $\mathbf{g} = (g_1, \ldots, g_B)$ and

$$a_t = g_b \quad \text{for } t \in \{(b-1)L+1, \ldots, bL\}. \quad (5)$$

MAP-Elites maintains an archive of elites indexed by behavioral descriptors $\phi(\tau)$ computed from the executed trajectory τ, thereby constructing a diverse repertoire of high-performing adversarial behaviors rather than converging to a single strategy. Each candidate genome is evaluated on

a fixed set of rollout specifications (shared across all methods), and its fitness is computed from safety-relevant outcomes while penalizing degenerate self-collisions of the adversary. Concretely, we maximize

$$J(\tau) = \alpha\, \mathbb{I}[\mathrm{CI}] + \beta\, \mathbb{I}[\mathrm{coll}] - \delta\, \mathbb{I}[\mathrm{nuis}], \quad (6)$$

where CI denotes a critical incident (ego collision or proximity-gated risk triggers), coll is an ego collision, and nuis denotes an adversary nuisance collision without ego safety impact. As behavioral descriptors, we use

$$\phi(\tau) = \big(\min_t d_{\mathrm{ego,adv}}(t),\ \mathrm{TTCI}(\tau)\big), \quad (7)$$

where $\min_t d_{\mathrm{ego,adv}}(t)$ denotes the minimum ego–adversary distance and $\mathrm{TTCI}(\tau)$ the time to first critical incident.

3.2. *Metrics*

Effectiveness measures the adversary's ability to induce unsafe outcomes, quantified by the collision rate p_{coll} and the time-to-critical-incident (TTCI), defined as the first occurrence of an ego collision, ego off-road, or a predefined proximity-based risk threshold. We report TTCI in policy steps at which any critical-incident condition is triggered, including near-collision risk thresholds, even if the episode does not terminate in a collision. Time-to-failure is reported only when a terminal collision occurs and is interpreted as a scenario-difficulty indicator rather than a general safety metric. Higher p_{coll} and lower TTCI indicate stronger adversarial pressure. For non-learning baselines such as parameter sweeps and fixed action-sequence search, TTCI is computed post hoc and is not optimized during scenario generation.

Behavioral Diversity is quantified independently of the adversarial generation paradigm. For each generated scenario, we extract a behavioral descriptor $\phi(\tau)$ (7). The descriptor space is discretized into a fixed two-dimensional grid shared across all methods. Coverage is defined as the fraction of grid cells occupied by at least one generated scenario. This metric captures how broadly a method explores qualitatively different critical interaction profiles.

Behavioral plausibility and feasibility are assessed via replay-based diagnostics rather than separate physical feasibility metrics, as vehicle motion and hard constraints are enforced by the simulator dynamics. We report maximum acceleration and jerk, lane-departure or off-road events, and related rule-consistency proxies. Family-level feasibility is defined as the fraction of rollouts within a scenario family that avoid terminal ego failure (ego collision or ego off-road) under fixed perturbations, distinguishing challenging but solvable edge cases from impossible scenarios.

Reproducibility is evaluated differently for learning and non-learning adversarial paradigms. For non-learning methods (parameter sweep, gradient-based optimization with fixed solutions, and MAP-Elites open-loop action sequences), reproducibility is assessed by re-evaluating the same generated scenario assets on identical rollout specifications. Under fixed seeds and deterministic simulation settings, outcome metrics are expected to match up to numerical determinism. For RL-based adversaries, reproducibility is assessed at two levels. First, evaluation determinism is measured by replaying a fixed trained checkpoint across identical rollout specifications. Second, training variability is quantified by comparing outcomes across independently trained checkpoints obtained from different training seeds. This separation ensures fair comparison without assuming deterministic policy learning.

4. Experiments

This section describes the experimental methodology used to compare the adversarial paradigms introduced in Section 3.1.

4.1. *Environment and ego policy*

We adopt a standard multi-lane highway configuration with moderate traffic density. The ego vehicle is controlled by a pre-trained deep Q-network (DQN) and remains fixed throughout all experiments. The policy is selected to exhibit stable, high-performance driving in the nominal environment, ensuring that discovered failures reflect genuine weaknesses of the ego policy rather than training noise. Background traffic is generated using `highway-env`'s stochastic traffic model and

logged for reproducibility.

4.2. *Adversarial configurations*

One background vehicle is designated as adversarial, and optimization targets its full action sequence or adversarial-policy parameters. Initial positions are selected such that interaction with the ego vehicle can occur within the episode horizon.

4.3. *Evaluation protocol*

For each combination of paradigm, adversarial configuration, and constraint regime, we generate a fixed number of scenarios from independent environment seeds. All methods receive an equal optimization budget, while evaluation uses equal-n resampling and macro-averaging to avoid data-volume bias across runs and paradigms. The resulting scenarios are evaluated by rolling out the fixed ego policy and computing effectiveness, diversity, behavioral plausibility and feasibility, and reproducibility metrics as defined in Section 3.2. Seeds, environment parameters, and policy versions are logged to ensure full replicability of results.

5. Results

This section reports comparative results for all adversarial paradigms under the unified evaluation protocol. Table 1 reports collision rate p_{coll} and mean TTCI. The parameter-sweep baseline yields the highest collision rate and the lowest TTCI due to worst-case initial configurations. The gradient-based adversary shifts outcomes toward earlier and more severe critical incidents. RL and QD-based adversaries show longer TTCI with non-zero collision rates, indicating sustained near-boundary interaction rather than trivial early collisions, with MAP-Elites in between. Collision rate alone is insufficient to characterize scenario quality. MAP-Elites searches open-loop action sequences, whereas QD-RL learns closed-loop policies, which aligns with QD-RL's lower collision rate and longer interaction horizons.

Table 2 reports behavioral diversity measured in the descriptor space ϕ based on evaluated trajectories. Coverage denotes the fraction of occupied ϕ-cells. Parameter sweep is omitted because it is a non-learning baseline and does not produce a comparable adversarial behavior representation (policy or optimized action sequence) for diversity analysis under the shared protocol. Gradient-based optimization shows the highest observed coverage, followed by RL adversaries and MAP-Elites, while QD-RL exhibits lower realized diversity under evaluation. The reported metrics reflect diversity observed in evaluation rollouts rather than full archive diversity.

Table 1. Effectiveness of adversarial paradigms measured by collision rate p_{coll} and mean TTCI in simulation steps.

Method	p_{coll}	Mean TTCI
Parameter Sweep	0.833	3.2
Gradient-Based	0.570	35.3
RL Adversary	0.433	118.0
QD-RL	0.233	117.1
MAP-Elites	0.300	31.0

Table 2. behavioral diversity of adversarial scenarios in descriptor space ϕ.

Method	Coverage (ϕ)	Diversity Score
Gradient-Based	0.031	0.222
RL Adversary	0.028	0.196
MAP-Elites	0.026	0.204
QD-RL	0.023	0.155

The high coverage of gradient-based methods reflects a large number of discovered failures within a narrow, aggressive regime, whereas QD-methods focus on qualitatively distinct interaction profiles that require more extensive exploration to fully populate the archive. Archive-level diversity is analyzed separately to characterize the exploratory capacity of QD methods. At the archive level (Table 3), MAP-Elites exhibits higher coverage and entropy in the descriptor space ϕ, while QD-RL shows lower archive coverage in the analyzed configuration.

Table 4 summarizes behavioral plausibility and family-level feasibility diagnostics computed from replay traces. Feasibility is defined as the fraction of rollouts within a scenario family in

Table 3. Archive metrics apply only to QD-based paradigms and are not applicable to standard RL or open-loop baselines.

Method	Coverage (ϕ)	Entropy (ϕ)	Diversity Score
MAP-Elites	0.306	0.801	0.554
QD-RL	0.031	0.429	0.230

which the ego vehicle avoids terminal failure under fixed perturbations. Parameter sweep shows low feasibility, indicating trivially unsolvable scenarios. Gradient-based optimization improves feasibility but exhibits highly aggressive dynamics. RL and QD methods achieve higher feasibility, with QD-RL attaining the highest values, while hard physical constraints are enforced by the simulator.

Table 4. Feasibility and behavioral plausibility diagnostics computed from replay traces.

Method	Max Acc.	Max Jerk	Feas.
Parameter Sweep	–	–	0.170
Gradient-Based	125.42	1676.45	0.430
RL Adversary	17.70	305.22	0.567
MAP-Elites	73.49	961.15	0.700
QD-RL	8.23	13.71	0.767

Reproducibility is assessed by re-evaluating fixed adversarial artifacts under identical rollout specifications. For gradient-based optimization, variance across independent runs is reported (Table 5). Low variance in collision rate and moderate variance in TTCI indicate consistent failure induction.

6. Discussion and Outlook

This study provides a unified comparison of adversarial-agent paradigms under a reproducible evaluation protocol and shows that adversarial effectiveness, behavioral diversity, and feasibility are not aligned objectives. While parameter sweeps are most effective at inducing collisions, QD-RL identifies more plausible and feasible interactions. The current QD results are preliminary; extended optimization budgets are expected to further improve archive density and diversity. Gradient-based and RL adversaries reliably induce early or severe failures but concentrate on narrow behavioral regimes, whereas QD-based methods emphasize sustained near-boundary interaction and diversity at the population level. While the parameter sweep achieves the highest collision rate, it primarily uncovers trivial failure modes with near-instantaneous onset. In contrast, evolutionary and RL paradigms induce failures through sustained interaction, probing the ego policy's decision-making boundary. Explicit feasibility diagnostics are essential to distinguish challenging yet solvable scenarios from trivially unsolvable ones and reveal that learning- and QD-based methods maintain adversarial pressure within the solvable envelope of the ego policy. Archive-based diversity is only partially reflected in fixed evaluation rollouts. From a V&V perspective, the results support adversarial scenario generation as structured exploration of behavioral space rather than worst-case optimization. QD and proxy-guided search are particularly suited for constructing reusable scenario repertoires that characterize policy weaknesses beyond isolated failures. A potential concern is that adversarial failures may exploit kinematic brittleness of the simulator or the discretized meta-action interface rather than semantically meaningful driving errors. We mitigate this by penalizing nuisance collisions, reporting feasibility under fixed perturbations, and analyzing stabilization, ensuring that observed failures correspond to solvable yet challenging interactions rather than artifacts of unphysical control or simulator instability. Future work will investigate longer-running QD optimization, as these paradigms are expected to benefit most from increased computational budgets to discover more complex corner cases. Additionally, we will explore coordinated multi-adversary settings and surrogate-based rollouts.

Table 5. Reproducibility variance is reported here specifically for the gradient-based proxy to highlight its sensitivity to initial conditions.

Method	Var(p_{coll})	Var(TTCI)
Gradient-Based	0.015	34.8

References

California Department of Motor Vehicles (2024). Autonomous vehicle disengagement reports. Accessed 2025-01-15.

Chitta, K., D. Dauner, and A. Geiger (2024). Sledge: Synthesizing driving environments with generative models and rule-based traffic. In *European Conference on Computer Vision*, pp. 57–74. Springer.

Deng, Y., J. Yao, Z. Tu, X. Zheng, M. Zhang, and T. Zhang (2023). Target: Automated scenario generation from traffic rules for testing autonomous vehicles. *arXiv preprint arXiv:2305.06018*.

Ding, W., C. Xu, M. Arief, H. Lin, B. Li, and D. Zhao (2023). A survey on safety-critical driving scenario generation—a methodological perspective. *IEEE Transactions on Intelligent Transportation Systems 24*(7), 6971–6988.

Doreste, A. (2025). Adversarial testing with reinforcement learning. In *2025 IEEE Conference on Software Testing, Verification and Validation (ICST)*, pp. 771–773. IEEE.

Gaylinn, N., A. Rees, N. Cheney, and J. Bongard (2025). Ramps and ratchets: Evolving spatial viability landscapes. In *Artificial Life Conference Proceedings 37*, Volume 2025, pp. 65. MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info ....

Hanselmann, N., K. Renz, K. Chitta, A. Bhattacharyya, and A. Geiger (2022). King: Generating safety-critical driving scenarios for robust imitation via kinematics gradients. In *European Conference on Computer Vision*, pp. 335–352. Springer.

International Organization for Standardization (2018). Iso 26262: Road vehicles — functional safety. Standard.

International Organization for Standardization (2022). Iso 21448: Road vehicles — safety of the intended functionality (sotif). Standard.

Ji, P., Y. Feng, Z. Li, X. Zhou, J. Liu, J. Sun, and Z. Zhao (2025). Txt2sce: Scenario generation for autonomous driving system testing based on textual reports. *arXiv preprint arXiv:2509.02150*.

Kalra, N. and S. M. Paddock (2016). Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability? *Transportation research part A: policy and practice 94*, 182–193.

Lu, Q., X. Wang, Y. Jiang, G. Zhao, M. Ma, and S. Feng (2025). Omnitester: Multimodal large language model driven scenario testing for autonomous vehicles. *Automotive Innovation*, 1–15.

Mouret, J.-B. and J. Clune (2015). Illuminating search spaces by mapping elites. corr abs/1504.04909 (2015). *arXiv preprint arXiv:1504.04909*.

Pugh, J. K., L. B. Soros, and K. O. Stanley (2016). Quality diversity: A new frontier for evolutionary computation. *Frontiers in Robotics and AI 3*, 40.

Ramakrishna, S., B. Luo, C. B. Kuhn, G. Karsai, and A. Dubey (2022). Anti-carla: An adversarial testing framework for autonomous vehicles in carla. In *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 2620–2627. IEEE.

Wäschle, M., F. Thaler, A. Berres, F. Pölzlbauer, and A. Albers (2022). A review on ai safety in highly automated driving. *Frontiers in artificial intelligence 5*, 952773.

Waymo LLC (2020, October). Waymo safety methodologies and safety readiness determinations. Accessed 2025-01-15.

Xu, C., A. Petiushko, D. Zhao, and B. Li (2025). Diffscene: Diffusion-based safety-critical scenario generation for autonomous vehicles. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Volume 39, pp. 8797–8805.

Yan, X., Z. Zou, S. Feng, H. Zhu, H. Sun, and H. X. Liu (2023). Learning naturalistic driving environment with statistical realism. *Nature communications 14*(1), 2037.

Yin, Y., P. Khayatan, É. Zablocki, A. Boulch, and M. Cord (2024). Regents: Real-world safety-critical driving scenario generation made stable. In *European Conference on Computer Vision*, pp. 262–276. Springer.