SCENARIST: RESHAPING THE DRIVING DATA DISTRIBUTION THROUGH OPEN-ENDED FAILURE SCENARIOS, FROM LANGUAGE TO POLICY

Anonymous authors
Paper under double-blind review

ABSTRACT

Rare and safety-critical interactions are underrepresented in naturalistic driving logs and often cannot be collected deliberately, while the scenarios needed for improvement evolve with a driving policy's failures. We present Scenarist, a language-to-policy benchmark for configuring long-tail driving-data distributions through open-ended failure scenarios. Scenarist compiles natural-language failure-scenario intents into rule programs composed of reusable map, signal, and interaction primitives, with deterministic checks ensuring semantic and physical validity. It connects responsive traffic generation, safety-constrained response exploration, and trajectory-conditioned RGB synthesis to produce action-aligned supervision for vision-language-action (VLA) policies. Safe responses from Group Relative Policy Optimization provide positive training targets, while unsuccessful responses provide negative feedback. The benchmark contains 16,397 interactions across eight vehicle and pedestrian failure-scenario intents, including Scenarist-2K, a realistically rendered subset of 2,144 interactions. Our response policy reduces collision rate from 36.70% to 6.23% while substantially increasing safe-response diversity. VLA post-training with positive-sample SFT and negative-sample RL reduces collision rate from 24.96% to 1.42% and increases safe rate from 74.76% to 98.58% on Scenarist-2K validation, while maintaining general-driving performance on the NAVSIM benchmark. Thus, configurable synthetic supervision can improve long-tail safety while largely preserving routine-driving performance.

1 INTRODUCTION

Collecting more miles multiplies the distribution the fleet sees; it does not reshape it. Expanding the fleet tenfold increases highway cruising and sudden cut-ins tenfold, preserving their proportions. The natural distribution supplies orders of magnitude too few of the kinds that decide outcomes (Feng et al., 2021; Rempe et al., 2022). Some situations are worse than rare: nobody can deliberately drive into a red-light runner, so their probability in any log one would willingly produce is zero.

The situations a policy needs, however, depend on that policy rather than driving. A model clearing two hundred logged four-way stops can still fail when a cyclist enters the crosswalk mid-turn: an ordinary configuration no catalogue would have anticipated. The deficit is discovered, not anticipated, and it moves as the policy improves. A distribution for a policy therefore cannot be fixed before that policy exists—which already rules out choosing it from a list.

Synthesis is the only data source whose distribution is chosen rather than sampled: one specifies instance counts per situation, including kinds with zero probability in any collectable log. But synthesis remains bounded by its specification: enumerating situations supplies anticipated cases rather than addressing discovered deficits. We compile natural-language failure-scenario intents into rule programs built from primitives, including the controlled agent, intended behavior, and trigger conditions. An LLM generates failure scenarios from these primitives, while deterministic rule-based checks on maps, traffic signals, and motion dynamics verify consistency with the failure-scenario intent. Only these primitives must be predefined; scenarios are composed from them. As long-tail coverage requirements and policy needs evolve, the generation distribution can be adjusted to target open-ended failure scenarios, connecting language-driven scenario construction to policy learning.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

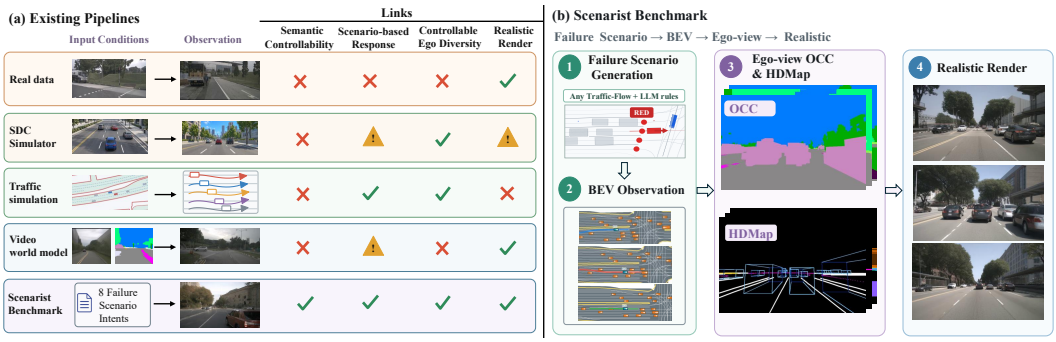


Figure 1: **The chain from language to policy, and where existing systems stop.** (a) Each prior system covers a prefix or suffix of the chain, never the whole: open-ended specification, controllable scene realization, real-domain observation, a response set with negatives, and a measured gain on real driving (check: available; triangle: partial; cross: absent). (b) Scenarist runs the full chain, so a failure described in a sentence becomes training data whose effect on a real-world policy is measurable.

Every link in this chain has been built. Simulators provide controllable scenes (Li et al., 2022; NVIDIA et al., 2025), language-based systems translate descriptions into executable scenarios (Zhang et al., 2024), traffic models generate plausible multi-agent responses (Feng et al., 2022; Peng et al., 2026), and video world models render prescribed futures photorealistically using inputs including RGB images and occupancy representations (Hu et al., 2023; Basant et al., 2026; Zhan et al., 2026). Yet chaining outputs is insufficient. Precise behavioral control can restrict natural surrounding-agent responses. Realistic videos may not follow ego actions, misaligning observations and action labels. Preserving intended conditions while exploring diverse safe ego responses adds another constraint. Success at one link does not validate the chain, which existing systems typically cover partially (Figure 1). Generated data has improved action models on real benchmarks, especially in difficult interactions (Wang et al., 2026b), but aggregate gains do not establish whether specific policy deficits are resolved; this requires evaluating corresponding failure scenarios.

Scenarist runs the chain end-to-end and brings these four links together, resolving those tensions in turn (Figure 2). A described failure becomes a structured rule program enforced by deterministic semantic checks. A responsive traffic model completes the surrounding agents, separating rule-based semantic precision from learned traffic plausibility. Group Relative Policy Optimization (GRPO) (Shao et al., 2024) then searches for feasible responses by comparing candidates sampled from the same event. An occupancy- and RGB-conditioned video model renders accepted responses from their ego frames, using action conditioning to produce distinct visual futures. Responses passing the safety filter supply positive targets; failed candidates supply negatives unavailable in logs. The same language interface supports a failure-directed loop: real-log failures can be specified, regenerated for post-training, and evaluated separately on the original cases. This protocol remains distinct from the generated-scenario validation reported throughout the current experiments.

Experimental results show that post-training a VLA policy on Scenarist data improves safety in long-tail scenarios while maintaining comparable general driving performance on the NAVSIM benchmark. Given the links' interdependence, we evaluate each link separately. Our main contributions across data generation and policy learning are summarized as follows:

- **Open-ended control of the driving data distribution.** We construct failure scenarios from language by composing reusable map, signal, and interaction primitives, adapting coverage and sampling to policy needs without relying on a fixed scenario catalogue.
- **Long-tail safety with routine-driving retention.** SFT and negative-sample RL on Scenarist-2K reduce validation collision rate by 94.3% and raise safe rate by 23.82 percentage points while preserving general driving performance on NAVSIM.
- **A generation-to-policy benchmark.** Scenarist provides 16,397 interactions across eight vehicle and pedestrian failure-scenario intents, including 2,144 in its RGB subset Scenarist-2K. Per-stage validation metadata supports four-link and downstream policy evaluation.

2 Related Work

2.1 Interactive simulation and traffic-flow generation

Driving simulators provide controllable environments for developing and evaluating driving policies. MetaDrive (Li et al., 2022) and AlpaSim (NVIDIA et al., 2025) provide configurable scenes and closed-loop observations. Learned traffic generators model traffic participants' joint motion: TrafficGen (Feng et al., 2022) synthesizes traffic scenes, UniGen (Mahjourian et al., 2024) unifies scene and motion generation, and SceneStreamer (Peng et al., 2026) supports responsive rollout with agents entering and leaving the scene. We adopt SceneStreamer as our structured traffic baseline and surrounding-agent completion model. Our focus is not long-horizon rollout itself, but control over which dangerous interactions populate the training distribution and how surrounding traffic responds to them throughout extended multi-agent traffic interactions.

2.2 Safety-critical scenario generation

Safety-critical generation increases exposure to events naturalistic logs rarely contain. NADE (Feng et al., 2021) accelerates evaluation through importance sampling, SceGAN (Yang et al., 2023) models highway cut-ins, and STRIVE (Rempe et al., 2022) seeks challenging yet solvable interactions. ChatScene (Zhang et al., 2024) specifies dangerous situations through language and executable scenario programs, while HSPG (Wang et al., 2026a) connects adversarial trajectories to multi-camera video generation. These approaches establish tools for constructing and exposing failures.

A generated failure scenario, however, is not by itself a target for learning how to respond. Scenarist connects scenario construction to response exploration and visual supervision. Language controls the interaction and its sampling, while safety-constrained GRPO (Shao et al., 2024) searches for feasible alternatives to the logged response. This separates the distribution of failure scenarios from the distribution of responses: one determines what the policy encounters, and the other determines which behaviors it can learn from each encounter. Diversity is evaluated among accepted safe responses rather than inferred from geometric variation alone.

2.3 Video world models and policy refinement

Video world models generate driving observations conditioned on scene context and motion, with inputs such as RGB images and occupancy representations. GAIA-1 (Hu et al., 2023) and OmniDreams (Basant et al., 2026) provide generative interfaces for driving video, including structured and occupancy-based control. CompoSIA (Zhan et al., 2026) separates scene structure, object identity, and camera motion for controllable visual generation. Such models provide visual observations needed by camera-based policies, but rendering a prescribed future does not establish that its action is a safe response to the underlying interaction.

BehaviorWorldGen (Wang et al., 2026b) also connects controllable traffic generation, world simulation, and action-model refinement. Its behavior-aware rollouts and rendered observations demonstrate the value of interaction-aware synthetic supervision. Scenarist emphasizes complementary control over the data distribution: language descriptions guide the construction of failure scenarios from reusable primitives, response exploration supplies alternative training targets, and rendering makes those targets available to visual policies. The benchmark separates scenario control, traffic response, ego diversity, and visual consistency, and evaluates their combined value through policy learning.

3 The Scenarist Benchmark

3.1 Overview and Data Formulation

To address distributional gaps in VLA training data, the Scenarist benchmark spans controllable construction of failure scenarios, exploration of diverse safe responses, and rendering of realistic ego-view RGB observations. Pairing these observations with their corresponding trajectories provides action-aligned supervision for VLA post-training. Scenario content and sampling frequencies can be adapted to coverage gaps or training needs, reducing reliance on the long-tailed distribution of naturalistic driving logs. Figure 2 illustrates how scenario construction and response exploration

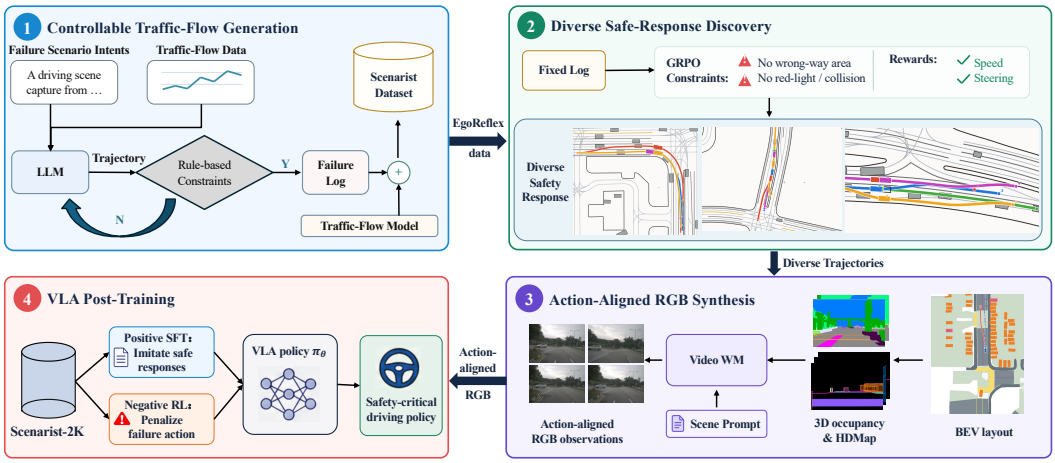


Figure 2: **Overview of the Scenarist benchmark.** (1) LLM-guided scenario rules and responsive traffic-flow generation construct failure scenarios for the benchmark. (2) Safety-constrained GRPO discovers diverse responses to a fixed event. (3) BEV rollouts condition an occupancy- and HD-map-guided video world model to produce action-aligned RGB observations. (4) Scenarist-2K supports positive-sample SFT and negative-sample RL for safety-critical VLA post-training.

connect to visual policy learning. For a failure scenario e constructed from a natural-language failure-scenario intent ℓ, the resulting trajectory–observation pairs are

$$\mathcal{D}(e) = \{(\tau, \mathcal{W}(t, \phi(\tau))) \mid \tau \in \mathcal{R}(e)\}. \tag{1}$$

Here, $\mathcal{R}$ denotes RL-based response exploration followed by safety and feasibility filtering. Using the scene context, ϕ converts each accepted BEV trajectory τ into video-generation conditions for the corresponding ego viewpoint. The video world model $\mathcal{W}$ combines these conditions with the scene text t associated with e to generate the corresponding ego-view RGB sequence. Each response is paired with observations generated along its own motion, yielding trajectory-specific video sequences. The following four links describe how these supervision pairs are constructed.

3.2 Connecting Language to Policy

Link 1 — Semantic Controllability. Given a natural-language failure-scenario intent ℓ, we construct a scenario e from a Waymo Open Motion Dataset scene (WOMD) (Ettinger et al., 2021). Figure 2, part 1, shows that ℓ specifies the controlled participant, target behavior, interaction region, trigger conditions, and temporal ordering. An LLM compiles these reusable elements into a rule program and generates the specified participant behavior. Rule-based constraints combine deterministic checks with traffic-signal context to assess map compatibility, temporal consistency, and motion feasibility. Violations are evaluated against the specified failure-scenario intent. When a candidate exhibits unintended wrong-way motion, intent mismatch, collisions, or road departures, we return each deviation's cause and severity to the LLM for revision. Generation and validation are repeated, and candidates are accepted only after passing applicable rule-based checks. Composing these elements and adjusting their sampling frequencies changes the scenario distribution without scripting each complete scenario separately in advance.

Link 2 — Scenario-based Response. Link 1 enables controllable failure-scenario construction, but editing a participant's trajectory in a BEV scene does not automatically update surrounding-agent behavior. If neighboring agents follow their original trajectories, the edited scene may contain inconsistent interactions or unintended collisions. We therefore use SceneStreamer (Peng et al., 2026) as the traffic-flow baseline to generate surrounding-agent motion conditioned on the validated intervention. The controlled participant behavior is preserved, while other participants respond to the scene, the response vehicle, and one another. The rollouts must realize the intended interaction and pass applicable feasibility checks. These multi-agent rollouts provide the interactive environment for response exploration in $\mathcal{R}$ and the scene context for constructing video conditions through ϕ.

Link 3 — Controllable Ego Diversity. A sudden dangerous interaction can affect multiple nearby vehicles rather than a single ego vehicle. We therefore consider all potentially affected surrounding vehicles as response agents and explore diverse trajectories for each, providing VLA supervision from their respective ego viewpoints. As illustrated in part 2 of Figure 2, we use GRPO (Shao et al., 2024) to optimize the response policy and explore different speed adjustments and steering choices while keeping the prescribed participant behavior fixed. The closed-loop reward encourages safe, diverse driving and strongly penalizes unsafe outcomes:

$$r = r_{\mathrm{saf}} + \lambda r_{\mathrm{div}} - r_{\mathrm{pen}}, \tag{2}$$

where r_{saf} captures safety-related scores, including vehicle clearance and time-to-collision (TTC). The diversity term r_{div} encourages different speed profiles among candidates passing the full-rollout safety checks, with weight λ. The penalty r_{pen} accounts for risks and violations, including collisions, road departures, wrong-way motion, and abrupt speed changes.

For each response vehicle, candidate rollouts undergo strict safety and feasibility filtering, followed by selection based on pairwise trajectory ADE to promote geometric diversity among retained responses. The retained trajectories form $\mathcal{R}(e)$ and are rendered from the corresponding vehicle's ego viewpoint to produce trajectory–observation pairs for VLA supervision.

Link 4 — Realistic Rendering. As illustrated in part 3 of Figure 2, for each accepted trajectory $\tau \in \mathcal{R}(e)$, ϕ uses the corresponding multi-agent BEV rollout to construct video-generation conditions for the response vehicle's ego viewpoint. These conditions include temporal 4D occupancy, coordinates, and 3D bounding boxes with HD maps. HD-map construction follows OmniDreams (Basant et al., 2026), while the video world model $\mathcal{W}$ is adapted from AnyScene (Zhou et al., 2026). Guided by the scene text t, $\mathcal{W}$ uses these visual conditions for image-to-video (I2V) generation, producing the corresponding ego-view RGB sequence. Each generated sequence is paired with the trajectory that defines its motion, forming the supervision pairs in $\mathcal{D}(e)$. Different responses thus yield observations along their respective ego motions rather than sharing a fixed log video. As shown in part 4 of Figure 2, rendering selected events produces the benchmark's RGB subset, Scenarist-2K. These trajectory–observation pairs support two-stage post-training of the VLA-based driving system. **Positive-sample SFT** updates the VLA using safe responses and their corresponding visual observations. **Negative-sample RL** freezes the SFT-tuned VLA and optimizes a separate routing policy for candidate-trajectory selection. The routing policy is trained with PPO using collision and off-road penalties from simulated rollouts, with failed VLA responses providing negative feedback. Thus, positive SFT trains trajectory generation, whereas negative RL trains trajectory selection without updating the frozen VLA generator parameters during routing optimization.

3.3 Discussion

The four links constrain one another. Actions specified by challenging or out-of-distribution failure scenarios can restrict freedom to generate plausible traffic responses. Dynamic responses from surrounding vehicles introduce uncertainty into the exploration of diverse ego trajectories, while realistic videos need not match observations expected along those trajectories. Consequently, neither ADE/FDE against recorded BEV trajectories nor video-similarity measures including SSIM and PSNR alone can establish the effectiveness of visual supervision. Teaching VLA policies to handle long-tail dangerous interactions requires diverse safe trajectories paired with observations consistent with their motion. This joint requirement is not fully captured by benchmark metrics.

The central question is whether generated supervision enables safe, flexible VLA responses to out-of-distribution failure scenarios. A recorded trajectory and its video capture only one outcome; alternative safe responses require corresponding observations. Scenarist creates usable supervision for unobserved alternatives by connecting the four links. Optimizing video fidelity or trajectory agreement with ground truth in isolation can improve local metrics without addressing this missing supervision. We therefore prioritize response safety and diversity, trajectory–video motion consistency, and downstream VLA learning. Established benchmark results provide complementary evidence of traffic realism and routine-driving retention, rather than the primary criterion for evaluating our benchmark.

Table 1: Failure-scenario distribution in the Scenarist benchmark dataset. Counts refer to distinct interactions, not response branches or video frames.

Failure scenario	**Count**	**Failure scenario**	**Count**
Hard brake	2,530	Multi-lane crossing	2,400
Sudden cut-in	2,317	Abort merge	2,252
Straight red-light run	1,898	Left-turn red-light run	1,840
Illegal left turn from a straight lane	1,710	Pedestrian red-light crossing	1,450

Table 2: Response evaluation on 2,977 controllable ego-role vehicles in Scenarist-2K. Completion and rates are fractions; Safe ADE averages pairwise ADE among five safe trajectories for the same vehicle and event. Bold marks the best result in each column.

Method	Completion ↑	Collision ↓	Off-road ↓	Success ↑	Min. clearance (m) ↑	Min. TTC (s) ↑	Safe ADE (m) ↑
Log-Replay	**0.7817**	0.1290	0.0836	0.5176	0.9124	1.8864	0.0000
SceneStreamer-base	0.7050	0.3670	0.1375	0.4486	1.1835	2.4411	1.6501
SceneStreamer-full	0.6811	0.4552	0.1821	0.3702	1.0961	2.2907	1.9406
Ours	0.7316	**0.0623**	**0.0677**	**0.5482**	**1.6144**	**2.9720**	**7.8251**

4 BENCHMARK DATA

The benchmark contains 16,397 constructed interactions from WOMD scene contexts. Its eight vehicle and pedestrian failure-scenario intents cover longitudinal, lateral, intersection, and signal-controlled interactions. Table 1 reports the number of interactions for each intent category. This distribution reflects our emphasis on balanced sampling across failure-scenario intent categories to mitigate the long-tail imbalance inherited from naturalistic driving data. Each event e records its specification, map and traffic context, responsive rollout, and validation metadata. Accepted response trajectories are represented by $\mathcal{R}(e)$ and provide structured supervision for learning alternative behaviors. The appendix will detail the dataset construction methods and procedures.

Scenarist-2K is the benchmark's realistically rendered RGB subset, containing 2,144 representative interactions with action-aligned ego-view observations. Figure 3 illustrates its out-of-distribution failure scenarios. For each selected event, accepted trajectories and their rendered observations provide visual supervision for VLA post-training. The subset includes the 288-sample validation set used in the dangerous-interaction evaluation. The subset count refers to underlying interactions, not the number of rendered frames or response branches.

5 EXPERIMENTS

Our evaluation asks whether dangerous interactions can supply diverse safe responses and useful visual supervision for VLA policies. We examine response safety and diversity alongside downstream performance on dangerous interactions, and use driving benchmarks as complementary checks of realism and routine-driving retention. We examine the visual prerequisites for trajectory–observation consistency through quantitative diagnostics in Table 6 and ego-view sequences in Figure 3, assessing semantic fidelity, video realism, and the evolution of dangerous interactions.

5.1 EXPERIMENTAL SETUP

Evaluation. We evaluate response safety, completion, and diversity for 2,977 controllable ego-role vehicles involved in dangerous interactions in Scenarist-2K rollouts (Table 2). Safety metrics are computed before acceptance filtering; Safe ADE is the mean pairwise ADE among five safe response trajectories for the same vehicle and event. The Scenarist-2K subset is excluded from GRPO training of the traffic-response policy, which uses the remaining benchmark data. On WOMD validation, all-agent evaluation reports collision and off-road rates alongside WOSAC realism metrics and

Table 3: All-agent safety and traffic realism on WOMD validation. Collision and off-road rates complement WOSAC metrics; – denotes unavailable results.

Method	Collision ↓	Off-road ↓	RMM ↑	Kin. ↑	Inter. ↑	Map ↑	minADE ↓
CAT-K (CAT-40) (Zhang et al., 2025)	–	–	0.761	0.456	0.811	0.874	**1.299**
SMART-R1 (Pei et al., 2026)	–	–	0.774	–	–	–	1.419
DecompGAIL (Guo et al., 2026)	–	–	**0.788**	**0.514**	**0.828**	0.895	–
baseline (Peng et al., 2026)	0.042	0.122	0.773	0.456	0.808	0.909	1.379
Ours	**0.037**	**0.118**	0.775	0.457	0.812	**0.909**	1.404

Table 4: Post-training our language-augmented RAP-DINO baseline, evaluated on the Scenarist-2K validation set. Bold marks the best result in each column.

Setting	Safe ↑ (%)	Collision ↓ (%)	Off-road ↓ (%)	ADE ↓ (m)	FDE ↓ (m)
RAP-DINO (Feng et al., 2026)	74.76	24.96	0.26	6.37	13.71
Positive SFT only	94.96	5.00	0.04	**1.23**	**2.03**
Negative RL only	93.38	6.46	0.15	5.12	10.35
Ours (Positive SFT + negative RL)	**98.58**	**1.42**	**0.00**	4.03	8.71

minADE. For VLA learning, our language-augmented RAP-DINO (Feng et al., 2026) is the baseline, and 288 samples from Scenarist-2K form the dangerous-interaction validation set. The NAVSIM v1 benchmark compares the baseline and combined SFT–RL policy for routine-driving retention. Video-conditioning ablations on nuScenes report mAP@50, mIoU, and FVD, with real video as a reference for generation quality and spatial consistency.

Implementation. We use Qwen3-4B for data construction, the SceneStreamer base model for responsive traffic generation, AnyScene's 14B image-to-video (I2V) model for RGB synthesis, and RAP-DINO with an added language-input module for VLA post-training. All experiments are conducted on 16 NVIDIA H200 GPUs. Detailed experimental settings will be provided in appendix.

5.2 Results

Response safety and diversity. Table 2 compares log replay, SceneStreamer-base, SceneStreamer-full, and our GRPO-trained response policy on Scenarist-2K. Compared with SceneStreamer-base, our policy reduces collision and off-road rates by 30.47 and 6.98 percentage points, respectively, while increasing completion and success by 2.66 and 9.96 points. Minimum clearance and TTC reach 1.6144 m and 2.9720 s, while Safe ADE increases from 1.6501 to 7.8251 m, indicating greater geometric diversity among safe responses. Our policy achieves the best collision, off-road, success, and Safe ADE scores across the four settings, demonstrating safer and more diverse responses.

All-agent safety in general driving. Table 3 evaluates our response policy on WOMD validation. Against the SceneStreamer baseline, collision and off-road rates decrease by 0.5 and 0.4 percentage points, respectively. Together with Table 2, these results show that the safety improvements extend beyond constructed failure scenarios to general driving, reducing collisions and road departures.

Learning to handle dangerous interactions. Table 4 evaluates the contributions of positive-sample SFT and negative-sample RL to the VLA-based driving system. Positive SFT adapts the VLA generator using safe demonstrations, whereas negative RL trains a separate routing policy using failure-based feedback. With the SFT-tuned VLA frozen, the combined system achieves a collision rate of 1.42% and a safe rate of 98.58%, compared with 5.00% and 94.96% for SFT alone. These additional gains are obtained through the routing stage rather than further updates to the VLA generator, which remains frozen throughout routing optimization.

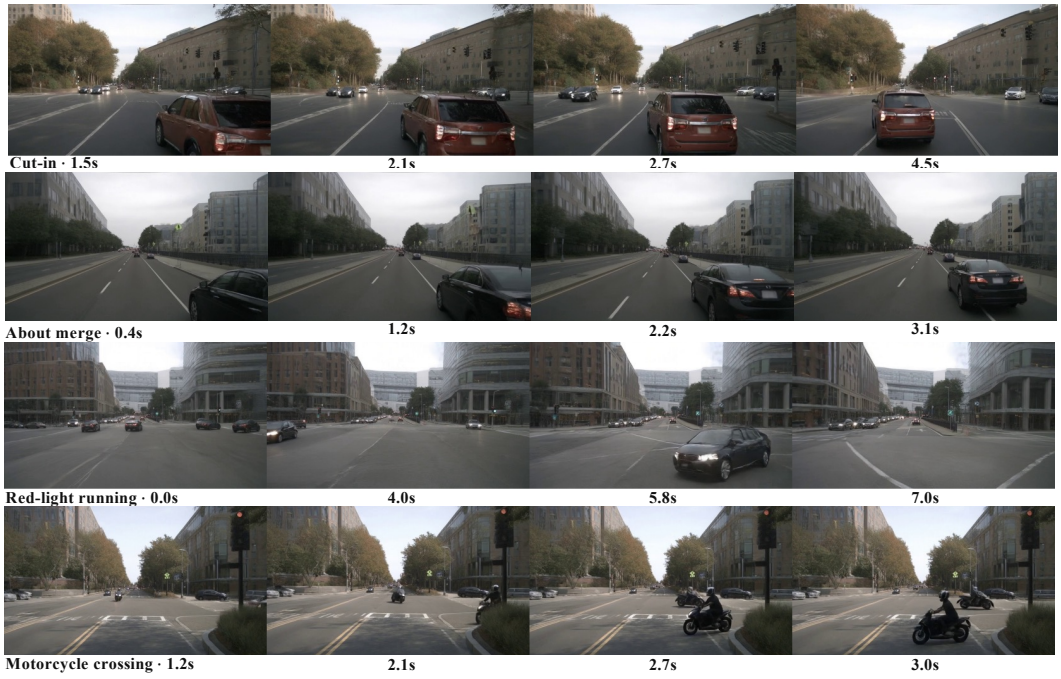


Figure 3: **Ego-view visualization of Scenarist-2K.** Each strip depicts a different out-of-distribution failure scenario, with sampled video frames ordered from left to right. The rendered observations remain visually plausible as the traffic scene evolves from the ego vehicle's viewpoint.

Table 5: NAVSIM comparison. RAP-DINO denotes our language-augmented baseline; ours adds positive-sample SFT and negative-sample RL. Bold marks column-best results.

Model	NC ↑	DAC ↑	TTC ↑	Comfort ↑	EP ↑	PDMS ↑
DrivingGPT (Chen et al., 2025)	98.9	90.7	94.9	95.6	79.7	82.4
DRAMA (Yuan et al., 2024)	98.0	93.1	94.8	100.0	80.1	85.5
ReCogDrive* (Xiong et al., 2026)	98.1	94.7	94.2	100.0	80.9	86.5
DiffusionDrive* (Liao et al., 2025)	98.1	96.1	94.2	100.0	82.1	87.7
DriveVLA-W0 (Li et al., 2026)	98.6	**99.1**	95.3	99.3	83.3	90.2
BehaviorWorldGen (Wang et al., 2026b)	98.9	98.7	96.2	100.0	89.2	93.3
RAP-DINO (Feng et al., 2026)	99.1	98.8	96.7	100.0	90.3	93.7
Ours	**99.1**	98.9	**96.7**	**100.0**	**90.3**	**93.8**

Using the same post-trained weights as in Table 4, our model preserves ego progress and comfort on NAVSIM while maintaining comparable PDMS, supporting long-tail safety improvements without sacrificing measured routine-driving performance in standard scenarios.

5.3 Ablation

This ablation tests how OCC, HDMap, and their fusion affect trajectory–observation consistency while retaining coordinate conditioning (Table 6). Joint conditioning achieves the best mAP@50 (13.30%) and FVD (708.73), whereas OCC alone gives the best mIoU (33.80% versus 33.57%). HDMap alone performs worst, and overlay fusion improves FVD over OCC but reduces mAP@50 and mIoU. Thus, occupancy information is essential and its fusion with HD maps matters.

Real video provides reference scores of 40.34% mAP@50 and 45.28% mIoU; joint conditioning reaches 33.57% mIoU despite the I2V domain gap. Figure 3 further shows motion-consistent observations across cut-ins, aborted merges, and red-light running. Together with the downstream

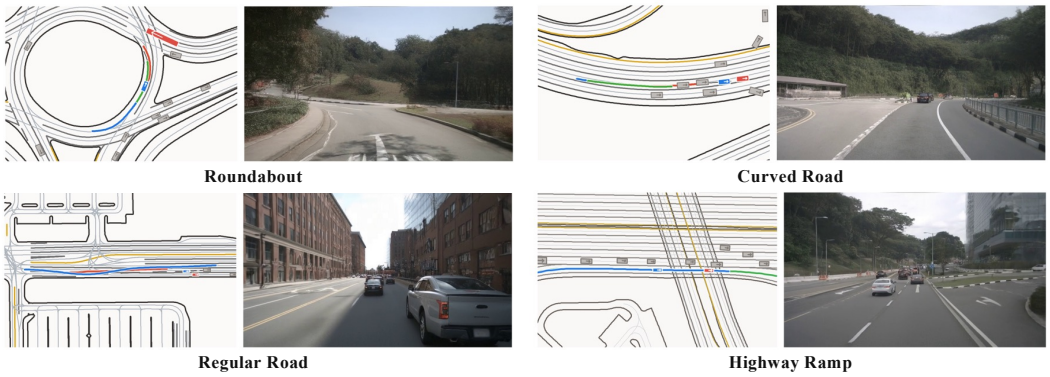


Figure 4: **Trajectory-conditioned rendering across road layouts.** Each example pairs a BEV representation of lane geometry and trajectories with its corresponding generated ego-view image. The four examples cover a roundabout, a curved road, a regular road, and a highway ramp.

Table 6: Video-conditioning ablation on nuScenes. Real video is a reference; bold marks the best generated result for each metric.

Input condition	mAP@50 ↑	mIoU ↑	FVD ↓
GT real video	40.34%	45.28%	–
OCC	13.02%	**33.80%**	731.44
HDMap	1.50%	16.06%	1278.03
Overlay	11.57%	30.20%	718.73
OCC + HDMap	**13.30%**	33.57%	**708.73**

gains in Table 4, these results support the value of the generated trajectory–observation pairs for VLA post-training across diverse failure-scenario interactions and road geometries.

5.4 Visualization

Figures 3 and 4 visualize temporal interaction evolution and BEV-to-RGB alignment across out-of-distribution failure scenarios and varied road layouts. The examples cover cut-ins, aborted merges, and red-light running, alongside road geometries ranging from roundabouts and curved roads to regular roads and highway ramps. They complement Table 6 with qualitative evidence of trajectory–observation consistency for VLA supervision.

6 Conclusion

We presented Scenarist, a language-to-policy benchmark that reshapes the driving data distribution through open-ended failure scenarios. It compiles language into reusable rule programs, makes scenario content and sampling configurable, and connects responsive traffic generation, diverse safe-response exploration, and trajectory-conditioned RGB synthesis. By controlling scenario coverage and sampling frequencies, Scenarist mitigates the long-tail imbalance inherited from naturalistic driving logs. The resulting action-aligned trajectory–observation pairs expose VLA policies to multiple feasible responses rarely covered by naturalistic logs. Positive-sample SFT learns safe trajectory generation, while negative-sample RL improves candidate selection through a separate routing policy, keeping the SFT-tuned VLA generator frozen. Scenarist provides 16,397 interactions across eight vehicle and pedestrian failure-scenario intents, including the 2,144-interaction RGB subset Scenarist-2K. Experiments demonstrate safer, more diverse traffic responses, substantially improved VLA safety on Scenarist-2K, and comparable general-driving performance on NAVSIM. Overall, closing the generation-to-policy chain rather than optimizing individual stages in isolation offers a practical route to improving policy behavior beyond recorded demonstrations.

AI USE STATEMENT

As described in Section 3, we use Qwen3-4B to generate rule programs and participant behaviors. LLM-based assistants also supported research and evaluation discussions, literature review, translation, manuscript drafting and polishing, result interpretation, and LaTeX editing. The authors remain responsible for the final content.

ETHICS STATEMENT

This work aims to support safer driving research in accordance with the ICLR Code of Ethics. The generated scenarios are intended for controlled evaluation, not deliberate reproduction on public roads. Simulation results do not certify deployment safety, and data use must respect applicable licenses and privacy restrictions.

REPRODUCIBILITY STATEMENT

Section 3 describes the four-link construction process, response objective, and filtering procedure; Section 4 summarizes the benchmark data; and Section 5 specifies the evaluated models, metrics, and comparisons. Appendix A.1 provides model initialization and training settings, while Appendix A.2 records evaluation datasets, training-data exclusions, and safety-metric definitions. Together, these sections document the reported experimental setup and distinguish traffic-response optimization, VLA post-training, and video generation. Code and datasets will be released upon acceptance.

REFERENCES

Aarti Basant, Amlan Kar, Despoina Paschalidou, Fangyin Wei, Francesco Ferroni, Guillermo Garcia Cobo, Haithem Turki, Huan Ling, Jaewoo Seo, James Lucas, et al. Nvidia omnidreams: Real-time generative world model for closed-loop autonomous vehicle simulation. *arXiv preprint arXiv:2606.03159*, 2026.

Yuntao Chen, Yuqi Wang, and Zhaoxiang Zhang. Drivinggpt: Unifying driving world modeling and planning with multi-modal autoregressive transformers. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 26890–26900. IEEE, 2025.

Scott Ettinger, Shuyang Cheng, Benjamin Caine, Chenxi Liu, Hang Zhao, Sabeek Pradhan, Yuning Chai, Ben Sapp, Charles R Qi, Yin Zhou, et al. Large scale interactive motion forecasting for autonomous driving: The waymo open motion dataset. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 9710–9719, 2021.

Lan Feng, Quanyi Li, Zhenghao Peng, Shuhan Tan, and Bolei Zhou. Trafficgen: Learning to generate diverse and realistic traffic scenarios. *arXiv preprint arXiv:2210.06609*, 2022.

Lan Feng, Yang Gao, Eloi Zablocki, Quanyi Li, Wuyang Li, Sichao Liu, Matthieu Cord, and Alexandre Alahi. Rap: 3d rasterization augmented end-to-end planning. In *International Conference on Learning Representations*, volume 2026, pp. 37722–37737, 2026.

Shuo Feng, Xintao Yan, Haowei Sun, Yiheng Feng, and Henry X Liu. Intelligent driving intelligence test for autonomous vehicles with naturalistic and adversarial environment. *Nature communications*, 12(1):748, 2021.

Ke Guo, Haochen Liu, Xiaojun Wu, and Chen Lv. Decompgail: Learning realistic traffic behaviors with decomposed multi-agent generative adversarial imitation learning. In *International Conference on Learning Representations*, volume 2026, pp. 83791–83808, 2026.

Anthony Hu, Lloyd Russell, Hudson Yeo, Zak Murez, George Fedoseev, Alex Kendall, Jamie Shotton, and Gianluca Corrado. Gaia-1: A generative world model for autonomous driving, 2023. URL https://arxiv.org/abs/2309.17080.

Quanyi Li, Zhenghao Peng, Lan Feng, Qihang Zhang, Zhenghai Xue, and Bolei Zhou. Metadrive: Composing diverse driving scenarios for generalizable reinforcement learning. *IEEE transactions on pattern analysis and machine intelligence*, 45(3):3461–3475, 2022.

Yingyan Li, Shuyao Shang, Weisong Liu, Bing Zhan, Haochen Wang, Yuqi Wang, Yuntao Chen, Xiaoman Wang, Yasong An, Chufeng Tang, et al. Drivevla-w0: World models amplify data scaling law in autonomous driving. In *International Conference on Learning Representations*, volume 2026, pp. 7890–7911, 2026.

Bencheng Liao, Shaoyu Chen, Haoran Yin, Bo Jiang, Cheng Wang, Sixu Yan, Xinbang Zhang, Xiangyu Li, Ying Zhang, Qian Zhang, et al. Diffusiondrive: Truncated diffusion model for end-to-end autonomous driving. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12037–12047. IEEE, 2025.

Reza Mahjourian, Rongbing Mu, Valerii Likhosherstov, Paul Mougin, Xiukun Huang, Joao Messias, and Shimon Whiteson. Unigen: Unified modeling of initial agent states and trajectories for generating autonomous driving scenarios. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 16367–16373. IEEE, 2024.

NVIDIA, Yulong Cao, Riccardo de Lutio, Sanja Fidler, Guillermo Garcia Cobo, Zan Gojcic, Maximilian Igl, Boris Ivanovic, Peter Karkus, Janick Martinez Esturo, Marco Pavone, Aaron Smith, Ellie Tanimura, Michal Tyszkiewicz, Michael Watson, Qi Wu, and Le Zhang. Alpasim: A modular, lightweight, and data-driven research simulator for autonomous driving, October 2025. URL `https://github.com/NVlabs/alpasim`.

Muleilan Pei, Shaoshuai Shi, and Shaojie Shen. Advancing multi-agent traffic simulation via r1-style reinforcement fine-tuning. In *International Conference on Learning Representations*, volume 2026, pp. 130791–130807, 2026.

Zhenghao Peng, Yuxin Liu, and Bolei Zhou. Scenestreamer: Continuous scenario generation as next token group prediction. In *International Conference on Learning Representations*, volume 2026, pp. 86038–86068, 2026.

Davis Rempe, Jonah Philion, Leonidas J Guibas, Sanja Fidler, and Or Litany. Generating useful accident-prone driving scenarios via a learned traffic prior. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17284–17294. IEEE, 2022.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Yang Wu, et al. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024.

Cheng Wang, Qiang Liu, Wenbo Fang, and Chen Xiong. Hspg: An open-loop testing framework for autonomous driving based on proactive generation of hazardous scenario. *Accident Analysis & Prevention*, 229:108449, 2026a.

Jiaqi Wang, Zhuo Zhang, Haining Guan, Tingguang Zhou, Haowen Cui, Zhongyang Zhu, Yulong Zheng, ChuanYe Wang, Xuefeng Chen, Zhen Yang, et al. Behaviorworldgen: Closing the loop between action models and world simulators via controllable behavior-aware structured world generation. *arXiv preprint arXiv:2608.22187*, 2026b.

Kaixin Xiong, Xiangyu Guo, Fang Li, Sixu Yan, Gangwei Xu, Lijun Zhou, Long Chen, Haiyang Sun, Bing Wang, Kun Ma, et al. Recogdrive: A reinforced cognitive framework for end-to-end autonomous driving. In *International Conference on Learning Representations*, volume 2026, pp. 157518–157556, 2026.

Lan Yang, Jiaqi Yuan, Xiangmo Zhao, Shan Fang, Zeyu He, Jiahao Zhan, Zhiqiang Hu, and Xia Li. Scegan: A method for generating autonomous vehicle cut-in scenarios on highways based on deep learning. *Journal of Intelligent and Connected Vehicles*, 6(4):264–274, 2023.

Chengran Yuan, Zhanqi Zhang, Jiawei Sun, Shuo Sun, Zefan Huang, Christina Dao Wen Lee, Dongen Li, Yuhang Han, Anthony Wong, Keng Peng Tee, et al. Drama: An efficient end-to-end motion planner for autonomous driving with mamba. *arXiv preprint arXiv:2408.03601*, 2024.

Yifan Zhan, Zhengqing Chen, Qingjie Wang, Zhuo He, Muyao Niu, Xiaoyang Guo, Wei Yin, Weiqiang Ren, Qian Zhang, and Yinqiang Zheng. Composing driving worlds through disentangled control for adversarial scenario generation. In *European Conference on Computer Vision*, pp. 467–487. Springer, 2026.

Jiawei Zhang, Chejian Xu, and Bo Li. Chatscene: Knowledge-enabled safety-critical scenario generation for autonomous vehicles. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15459–15469. IEEE, 2024.

Zhejun Zhang, Peter Karkus, Maximilian Igl, Wenhao Ding, Yuxiao Chen, Boris Ivanovic, and Marco Pavone. Closed-loop supervised fine-tuning of tokenized traffic models. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5422–5432. IEEE, 2025.

Junfei Zhou, Haiming Zhang, Feng Jiang, Jingzhong Li, Hao Zhang, Zhenglong Guo, Penglin Dai, Jifeng Dai, Yan Xie, and Benjin Zhu. Anyscene: Towards highly controllable driving scene generation at anywhere and beyond, 2026. URL https://arxiv.org/abs/2605.26113.

A Experimental Protocols and Additional Visualizations

A.1 Training Protocol

Traffic response and VLA. We fine-tune the SceneStreamer response policy with safety-constrained GRPO on WOMD training scenes and Scenarist interactions outside Scenarist-2K. Excluding Scenarist-2K's 2,144 interactions from 16,397 leaves 14,253 constructed training interactions. GRPO explores alternative responses while preserving the specified failure scenario, rewarding safety and diversity and penalizing collisions, road departures, and wrong-way motion. Only responses passing final safety and feasibility checks provide visual supervision. We initialize the VLA policy from our language-augmented RAP-DINO (Feng et al., 2026). Positive-sample SFT uses the 1,856-sample Scenarist-2K training split and its accepted trajectory–observation pairs; the remaining 288 source scenarios are held out for validation. In the combined SFT–RL setting, we freeze the SFT-tuned VLA and retain its failed responses as negative feedback. A routing policy selects candidate trajectories for simulation and is optimized with PPO using collision and off-road penalties, while the VLA remains frozen. Tables 2 and 4 provide complementary evidence for discovering diverse safe responses and exploiting alternative candidates through learned routing, respectively, with negative-sample RL reducing collision rate from 5.00% to 1.42% while keeping the SFT-tuned VLA generator frozen.

Video generation. We start from the pretrained AnyScene 14B image-to-video model. For the conditioning ablation in Table 6, each input configuration is fine-tuned for 1,000 iterations on nuScenes training scenes and evaluated by predicting the first 121 frames of each validation video. Training and evaluation use a fixed sampling rate of 10 Hz. This comparison isolates the effect of OCC, HDMap, and their fusion on trajectory–observation consistency.

A.2 Evaluation Protocol and Data Separation

Response evaluation. Table 2 evaluates 2,977 candidate ego-role vehicles in Scenarist-2K, with Safe ADE defined as the mean pairwise ADE among five safe responses for the same vehicle and event. Table 3 follows prior work and evaluates all traffic participants in 880 WOMD validation scenes.

VLA evaluation. We use two protocols. NAVSIM v1 (Table 5) combines open-loop prediction with simulator-based scoring in a pseudo-simulation setting. For each of 12,146 test windows, the model predicts one future trajectory from the logged observation. The official PDM scorer simulates it against cached scene context, surrounding agents, and map geometry to compute NC, DAC, TTC, Comfort, EP, and PDMS; the model receives no updated observation and does not replan. Scenarist-2K (Table 4) is evaluated separately on 2,600 windows derived from 288 held-out source scenarios. Each source sequence spans 12.1 s at 10 Hz, with predictions evaluated over 4 s. We compute ADE, FDE, and heading error against the target response, and compare predictions with logged participants and road geometry for window-level Collision, Off-road, and Safe rates. Surrounding agents follow fixed logged trajectories and do not react online. Until observation updates and replanning are verified, we describe Table 4 as trajectory-rollout evaluation rather than ego closed-loop.

Joint interpretation of safety and routine-driving performance. Tables 4 and 5 evaluate the same post-trained system with identical VLA-generator and routing-policy weights; NAVSIM uses no separate checkpoint. On Scenarist-2K, post-training reduces collision rate from 24.96% to 1.42% and increases safe rate from 74.76% to 98.58%. On NAVSIM, the system preserves ego progress (EP: 90.3) and comfort (100.0), with comparable PDMS (93.7 versus 93.8). Together, these results support improved long-tail safety without degrading routine-driving progress or comfort. ADE and FDE in Table 4 measure reference-response agreement, not driving progress; their increase over SFT-only does not establish lost driving capability.

Data separation and safety metrics. Traffic-response fine-tuning uses WOMD training scenes and their Scenarist interactions, excluding Scenarist-2K; the 880 WOMD validation scenes are held out to prevent overlap across training and evaluation data. Scenarist-2K derives from WOMD, keeping VLA post-training separate from NAVSIM evaluation. Within Scenarist-2K, 1,856 samples train the model and 288 source scenarios validate it, with no constructed scenario shared across partitions. A collision occurs when the ego's oriented bounding box overlaps that of any valid non-ego participant during the 4-s horizon, irrespective of fault. Clear off-road occurs when the ego center lies more than 0.8 m outside the approximated drivable region or when more than half of nine footprint points fall outside it. A window is Safe only when neither event occurs.

A.3 Additional Visualizations and Evaluation Scope

Figures 5 and 6 show random interactions from eight failure-scenario intents; each row follows one interaction across four timestamps. Blue denotes the ego, red the controlled failure-scenario participant, and gray surrounding traffic. They cover highway and intersection settings with varied spatial layouts and temporal evolution. The eight intents define the evaluation coverage of the current benchmark, not the configuration space of the generation interface: the language interface can target any participant and condition its trajectory on a requested behavior, allowing Scenarist to construct scenarios beyond the evaluated categories as coverage gaps and policy needs evolve.

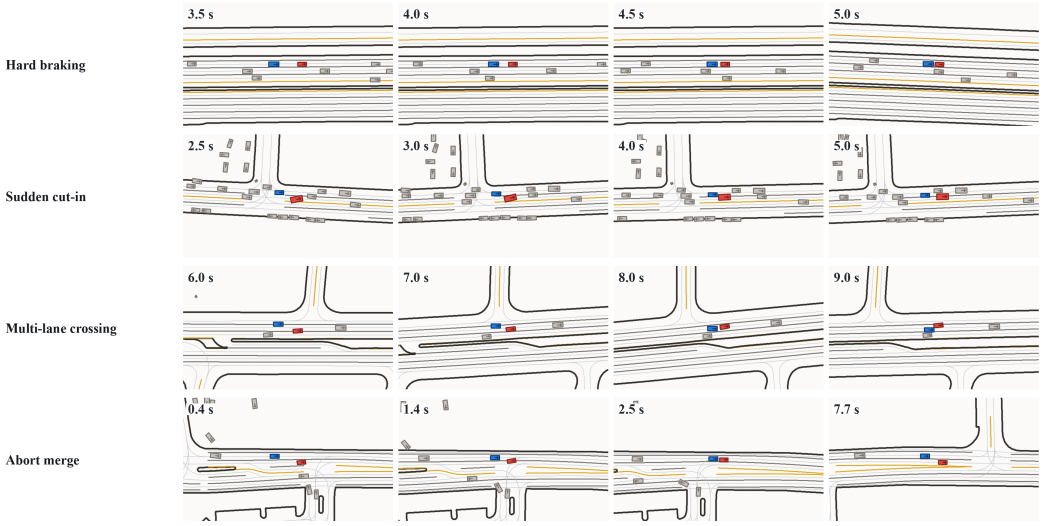


Figure 5: Randomly sampled interactions from four vehicle-centric failure-scenario intents: hard braking, sudden cut-in, multi-lane crossing, and aborted merge. Each row shows four timestamps; blue marks the ego vehicle and red marks the failure-scenario agent.

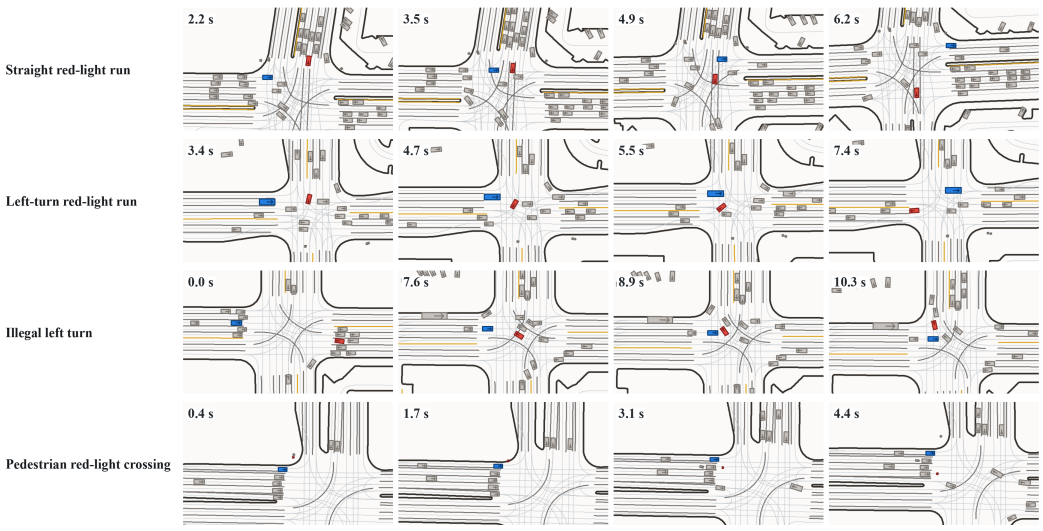


Figure 6: Randomly sampled interactions from four signal- and rule-related failure-scenario intents: straight red-light run, left-turn red-light run, illegal left turn, and pedestrian red-light crossing. The color convention follows Figure 5.