Frontiers in Computer Science and Artificial Intelligence
ISSN: 2798-8048
DOI: 10.32996/fcsai
Journal Homepage: www.al-kindipublisher.com/index.php/fcsai

Evaluating Learning from Failure in End-to-End Autonomous Driving: A Framework for Safety, Generalization, and Training Cost

Krunal Panchal ✉[1], **Hemang Upadhyay**[2], **Nixalkumar Patel**[3]

[1] *Independent Scholar, United States*

[2] *Senior Product Manager, LG Electronics USA, New Jersey, United States*

[3] *Independent Researcher, New York, United States*

Corresponding Author: Krunal Panchal**, E-mail:** Krunal.cp@live.com

ABSTRACT

End-to-end autonomous driving systems can work well in most scenarios but fail in unique scenarios and under less-than-ideal conditions. What went wrong can enhance safety. Still, if one could learn to perform better after more training, it is not proven that the learning is more effective through failure-oriented training than through a more normal form of continued training. Critical-scenario generation and failure feedback are already employed in existing research, as is counterfactual repair, while reinforcement learning is also already employed. The difficulty of the evaluation challenge will be to distinguish useful learning from further computation, further exposure to similar scenarios, and reduced performance in normal driving. This paper outlines a method for the evaluation of such a comparison. It differentiates failures from near misses, rare events and out-of-distribution events. It splits the distribution that is used to find failures, from the distribution that is used to test the revised policy. There are two complementary comparisons in the framework: one with matched simulator interactions and one with matched computational resources. They both cover failure discovery, scenario generation, verification, and policy adaptation. Generalization is assessed with independent roots of scenarios, unseen families of interactions, and pre-defined combinations of distribution shifts. Assessment of safety outcomes is done in addition to route completion, driving comfort and nominal-performance retention. The paper also outlines appropriate baselines, statistical reporting requirements, and circumstances that would serve as a test to the main hypothesis. No new experiments of training are reported. The framework offers a practical approach to assess the potential for transferable safety gains from learning from failure at reduced adaptation expense.

KEYWORDS

Autonomous driving; end-to-end learning; reinforcement learning; failure analysis; scenario generation; distribution shift; safety evaluation.

ARTICLE INFORMATION

ACCEPTED: 30 September 2026 **PUBLISHED:** 31 October 2026 **DOI:** 10.32996/fcsai.2026.5.11.1

1. Introduction

Reliable perception and effective decision-making are key to autonomous driving. The vehicle needs to sense the presence of nearby objects, learn their movement, obey the traffic laws and react properly to changes in the environment. Object detection is a critical component of good performance yet not an automatic guarantee of safe driving.

In our last article, "Adverse Weather Sensor Fusion Using Machine Learning for Robust Object Detection in Self-Driving Cars," [1] we looked at how to use camera, LiDAR and radar information to enhance perception under adverse weather conditions. It was a multi-level fusion approach, which fused complementary information from sensors. That work was along the lines of object detection in the presence of an unreliable individual sensor. The question considered in the present paper is a related but different one: How to assess the impact of learning from observed driving errors on subsequent decision-making and interactions?

Even if objects are detected correctly, a driving policy can be a failure. It can, for instance, fail to brake in time for a cut-in, hesitate at a junction or take an unsafe route around a pedestrian. These failures may also be caused by perception errors. The behavior of the entire driving system must, therefore, be considered when evaluating.

Copyright: © 2026 the Author(s). This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) 4.0 license (https://creativecommons.org/licenses/by/4.0/). Published by Al-Kindi Centre for Research and Development, London, United Kingdom.

Closed-loop evaluation is particularly important. The policy acts, changes its state in the environment, and gets new observations. A mistake made early would influence later decisions, and would allow for conditions that were not included in the original data.

The study of learning from difficult/fail scenarios is a well-trodden research path. Adversarial traffic is created for closed-loop training by CAT [2] for policy-specific adversarial testing. The generated critical scenarios and expert solutions with KING are used to enhance imitation policies [3]. STRIVE is designed to find hard but solvable scenarios, where a learned traffic prior is used [4].

Recent techniques take this line of approach. In the dense learning approach, informative failures and successful avoidance cases are taken into account [5]. In failed tasks, Counterplay revisits previous states, and learns from known alternative behaviors within a given additional interaction budget [6]. ELF-VLA uses explicit failure feedback to create corrective training samples [7].

These studies lead to the practical question of:

In terms of overall adaptation cost, will failure-based adaptation offer better transfer of safety benefits than standard continuation training?

This paper suggests a model of how to answer that question. It does not present a new driving policy or that failure-conditioned learning is a new phenomenon. It instead condenses what is required for fair comparison: clear failure definitions, independent evaluation scenarios, full cost accounting, solid baselines, and measurement of normal driving performance.

2. Related Work

A. Learning from Critical Scenarios

Critical scenarios are scenarios that reveal vulnerabilities within a driving system. This can include emergencies such as rapid stops, poor cutting in, pedestrian-to-vehicle interaction or challenging intersections.

CAT is a combination of: policy training by iterative learning and adversarial scenario generation [2]. The KING is used to generate critical traffic behavior and to enhance an imitation policy [3]. STRIVE looks for scenarios that are a stretch for a planner that still have a viable solution [4].

These approaches indicate that there is a need for a greater methodological contribution than merely collecting difficult situations and training on them. In order to be a new study, it must point out differences from previous generation, selection, correction or evaluation methods.

SEAL is a solution that employs adversarial skills to produce challenging closed-loop problems and documents gains in various test scenarios [13]. The generalization evidence is also relevant in the context of previous methods assessing only known scenarios, as demonstrated by this paper.

B. Failure Feedback and Counterfactual Learning

Failure label signifies that an undesirable result has taken place. Does not necessarily provide the reasons for the outcome or for the change of the policy.

The aim of Counterplay is to roll back to previous states, to try out other driving behaviors, to ensure that recoveries are performed correctly, and to move corrective behavior into the deployment policy [6]. ELF-VLA offers a structured failure feedback to help the development of corrected samples [7].

CRAFT is a ground-based closed-loop policy adaptation process that includes counterfactual feedback [14]. The results of these methods show that there is already a solid ground of failure correction and counterfactual post-training.

If there is a comparison between the original and the alternate rollout, then there should be a description of the differences. This can be an actor changing their behaviour in a controlled fashion, or a different policy action, or an intervention time. The term random augmentation should not be considered as proof of causal thinking because the intervention is not specified.

C. Scenario Selection and Safety Objectives

Prioritized level replay is a method of choosing the training environments based on the estimated learning potential [8]. Thus it is not enough to give greater weight to challenging scenarios to create novelty.

There are also foundations for safety objectives. Constrained policy optimization is the approach that distinguishes between performance goals and explicit constraints [9]. CVaR, or conditional value at risk, is a measure that emphasizes the bad outcomes and offers a foundation for taking decisions while considering risk.

These approaches can be used to assist in the research of autonomous-driving. A constraint or curriculum, or a tail-risk term does not tell you a new method by itself. The paper should provide an explanation of the target distribution, the assumptions and the specific contribution.

D. Evaluation Environments

Bench2Drive allows for closed-loop, interactive evaluation of the whole driving process in several scenarios and weather conditions [11].

The original NAVSIM protocol is based on short horizon non-reactive simulation with real observations [12]. It serves as a useful planning tool but does not replicate the interaction process that can take place in a reactive closed loop simulator.

These may not be used as substitutes for one another. A study of interactive failure recovery must have an initial evaluation that has the impact of the policy on the observations and interactions.

3. Problem Definition

A. Driving Policy and Scenario

Let e be a driving situation. It contains road layout, initial conditions, surrounding actors, ambient conditions and the configuration of the sensors.

Let π_θ be the driving policy with parameters θ. Observations are made to the policy and an action or a trajectory of the planned direction follows.

Let τ be the resulting rollout of the sequence of observations, actions, and states.

If the study is conducted in an end-to-end manner using sensors, the sensor observations must be specified in the policy deployed. The use of simulator ground truth can be for training and/or evaluation, and it must be reported.

B. Failure Definition

Define the failure indicator as:

$$F(\tau, e) = \begin{cases} 1, & \text{if a predefined failure occurs,} \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

Failing should be correlated with the actual occurrences such as collisions, off road driving, or specific traffic rule violations.

Task failures (timeouts and routes not completed) should be counted separately from physical safety failures. Adding together all outcomes may mask significant trade-offs.

C. Different types of scenarios

The following categories describe different properties.

Category	Meaning
Failure	A predefined undesirable event occurred during the rollout.
Near miss	A safety conflict without reaching to failure.
Low probability scenario	Scenario for which the probability is low given a stated reference distribution.
Out-of-Distribution (OOD) scenario	The scenario is different from a specified training distribution.
Difficult scenario	The scenario involves challenging decisions or interactions.
Safety-critical scenario	Decision(s) in the scenario may result in significant safety impact.

These categories can overlap but are not synonymous.

Frequent scenario generated by an adversarial system is not necessarily frequent in normal traffic. Likewise, a challenging situation isn't necessarily beyond distribution.

Near-misses can be identified using time to collision and minimum separation. Geometry, relative motion, visibility and available evasive action all influence their interpretation. Justification of thresholds for operating conditions should be done and set in advance of the main comparison.

D. Training and Evaluation Distributions

Let $q_\theta(e)$ describe the distribution used to collect training scenarios.

The observed failure set depends on both this collection distribution and the behavior of the current policy. Its frequencies should not be interpreted directly as real-world event probabilities.

Let $p_{\text{test}}(e)$ describe an independent evaluation distribution. The failure risk of interest is:

$$R_{\text{test}}(\theta) = \mathbb{E}_{e \sim p_{\text{test}},\, \tau \sim \pi_\theta}[F(\tau, e)]. \tag{2}$$

If the evaluation distribution is focused on the challenging scenarios, then the outcome is related to the performance under that stress-test distribution.

For any claim involving deployment risk, a defensible distribution of deployment and an appropriate estimator are needed. This framework does not make an assumption that such a distribution is available.

4. Proposed Evaluation Framework

A. Central Hypothesis

The main prediction is:

Failure-focused adaptation helps to mitigate failure risk on independent scenarios more efficiently (with less ordinary continued training) without reducing nominal driving performance.

This hypothesis needs to be proven, not assumed.

Let θ_T represent the targeted adaptation and θ_R represent the continued-training control. At a specified budget, define:

$$\Delta R = R_{\text{test}}(\theta_R) - R_{\text{test}}(\theta_T). \tag{3}$$

A positive value favors targeted adaptation.

Before assessing a study, the minimum practically meaningful improvement should be defined. It should also specify a nominal-performance margin that is acceptable. These thresholds are not to be chosen based on the results.

B. Failure Discovery

The original policy should be executed on a discovery set.

The logging system should be able to store:

- Identifiers and original roots of scenarios.
- Seeds for the simulator and policy versions.
- Sensor and environment settings.
- Actor trajectories and relevant policy actions.
- Failure Timestamps and Termination reasons.
- Use safety, task failure and near miss labels.

Failure representations can start with familiar features such as conflict timing, road geometry, types of interaction, visibility and relative movement.

Clustering can be beneficial if it leads to fewer duplicate cases and/or identifies gaps in coverage. It should not just be added for the sake of making the method more complicated.

Discovery should also remember some learning successes. A successful avoidance case will expose the action that averted the harm. There is an important precedent in the space of dense-learning research to consider both informative failures and successes [5].

C. Controlled Scenario Variants

A scenario variant is a change in a specified variable, whilst maintaining other variables of interest constant.

Possible parameters include:

- Cut-in timing.
- Pedestrian emergence time.
- Relative speed.
- Braking intensity.
- Visibility.
- Road friction.
- Noise or delay from the sensor.

Study should include valid ranges and validity checks for the parameters.

The physical feasibility, realism and avoid ability are distinct properties. A scenario can be physically feasible but not feasible. It could also be realistic and offer no feasible avoidance opportunity, given the observation constraints of the policy.

A successful answer to the question by the reference-controller shows that there is at least one solution that was tested. It doesn't mean that every policy contains the information for identifying that solution.

All scenarios that are variants of the same scenario should be kept in the same data partition.

D. Policy Adaptation

The first comparison should involve the following factors remaining constant:

- Policy support and action platform.
- Reinforcement-learning optimizer.
- Reward definition.
- Nominal-data mixture.

- Scenario-validity checks.

There are no changes to the scenarios except for the allocation. This helps remove the ambiguity surrounding the value of targeted selection.

A second full system comparison can be used to test a method involving different scenario allocation and different learning objectives.

Supervised correction should also be compared to reinforcement learning, if possible. Otherwise, improvements might be the result of better information from experts than of learning through interaction.

E. Nominal-Performance Preservation

A policy that's safer in some extreme scenarios might be worse when it's driving with normal cars.

To evaluate these, route completion, goal success, progress, frequency of timeout, comfort, and traffic violations should be included in a nominal evaluation.

There are three potential modes for preservation: nominal replay, teacher regularization and constrained adaptation. These mechanisms should be the same for all allocation controls, unless they are being ablated.

A measured nominal-performance requirement does not guarantee safety for all distribution shifts.

5. Cost Accounting and Baselines

A. Complete Adaptation Cost

The adaptation budget should include work performed specifically for the comparison.

$$N_{\text{total}} = N_{\text{discovery}} + N_{\text{search}} + N_{\text{verification}} + N_{\text{training}}. \tag{4}$$

Each simulator transition should be counted once.

When multiple agents are involved in simulation, researchers need to differentiate environment transition and agent transition.

Rejected variants, unsuccessful searches, alternative branches and verification rollouts should be included. Rendering and model queries can use up computation resources without adding to the amount of simulator transitions, and therefore need to be reported separately.

The following are common "pretrained assets" to be disclosed. Generators or experts of a specific method are not to be taken for granted.

B. Complementary Budget Comparisons

There are two comparisons recommended:

1. Interaction matched comparison: All methods are given the same allowance for simulator interaction.
2. Compute-matched comparison: Both methods are given the same amount of computational resources given a certain hardware setup.

Equal interactions do not imply equal computation. Equal computation does not equal equal training updates.

Simulate transitions and report accelerator-hours, processor-hours, expert queries, elapsed time. This is helpful for readers to know where a method is using its resources.

C. Required Baselines

Baseline	Scientific purpose
Original policy	Measures improvement relative to the starting point.
Random continued training	Tests whether additional ordinary training is sufficient.
Uniform scenario training	Tests whether broader coverage explains the improvement.
Failure replay	Tests whether replay alone provides the benefit.
Prioritized scenario selection	Compares against established allocation strategies.
Failure training with random perturbation	Tests whether nearby additional data is sufficient.
Supervised correction	Tests whether reinforcement learning is necessary.
Closest applicable published method	Tests the claimed distinction from prior work.

The selection of baseline should be based on the proposed mechanism.

The branch-repair method should include verified recovery (getting verified by Counterplay [6]). A good scenario-selection method should have a good prioritization control [8]. Imitation-based improvement, such as KING [3] should be tackled by a method based on generated expert corrections.

If you can't reproduce a method exactly as it is written, you should say so.

6. Generalization and Statistical Evaluation

A. Independent evaluation of partitions.

The framework suggests that there should be separate partitions for:

- Training failures.
- Familiar interaction families that are held out.
- Familiar roots that are independent.
- Unseen interaction families.
- Nominal driving.
- Predefined distribution shifts.

Partitioning should be done before the mining and augmentation.

If there are variants of the original scenario, and related trajectory segments, then they should all stay in the same partition.

The within-family transfer and cross-family transfer should be reported separately. Local transfer has been observed with a policy operating with a different pedestrian emergence time. It's a different and more challenging task to deal with the unseen cyclist interaction.

B. Compound Distribution Shifts

Compound shits are weather (rain) and dark elements (sensor delay), or heavy traffic and sensor delay.

All factors must be simulated and pertinent to the said conditions.

The study should document the presence of any of the following or combinations of the following: The study should document the following or combinations of the following: An out-of-distribution condition is not implied by a new random seed.

Changes in the environment and the sensor have inspired us to explore perceptions, as in our previous sensor-fusion research [1]. But there are different pieces of evidence for perception robustness and driving-policy robustness.

C. Safety and Driving Outcomes

The denominator should be defined in the collision risk.

Report failure prob at the episode level, with an appropriate measure of exposure (distance, time, etc.). A lesser-distance policy can have more collisions and be less capable.

Other safety measures are:

- Off-road events.
- Traffic-rule violations.
- Validated near-miss measures.
- Safety-relevant outcomes of other actors.

Driving measures include:

- Route completion.
- Goal success.
- Progress.
- Timeout frequency.
- Comfort and jerk.

Efficiency should be provided as a safety versus budget curve and not a final score.

Present absolute risk difference and budget required to achieve predefined targets. If the target isn't met, report the result directly.

D. Statistical Requirements

Training seeds and scenario roots are an example of different sources of variation.

Running the same scenario multiple times does not constitute evaluation over multiple independent scenarios. Report both counts.

Sample sizes should be based on the pilot variance estimates and the minimum effect the study must be able to detect. No one knows the number of episodes that would provide enough evidence.

Assess policies on matched scenario roots, if applicable. Clustering should be taken into account and this pairing should be preserved when constructing confidence intervals.

When the failures are rare, special care must be taken with regard to the rare-event analysis. A zero collision sample is not a zero risk sample. In cases of many zeros in outcomes, the bootstrap methods should also be examined.

Predict one main comparison and result. Other variables and changes should be considered as secondary or dealt with via an appropriate multiple comparison procedure.

7. Interpretation and Rejection Criteria

Only if the targeted method could be shown to have a practically meaningful safety benefit on independent scenarios, withstand the stated cost accounting, and satisfy the nominal-performance requirement is the hypothesis supported.

There are several results that would contradict the hypothesis:

- The same or improved safety-efficiency curve is realized with continued training.
- Gains are wiped out once discovered and generation costs are added.
- There is significant route-completion loss along with collision reduction.
- Improvements only happen on variants of familiar scenario roots.
- There are improvements in safety within families, but declines in safety elsewhere.

These findings are important and negative. They deem areas of targeted adaptation as inefficient, too specialized or too conservative.

It is important to note that "insufficient evidence" should not be confused with evidence against practical superiority. If the mean value lies outside the specified meaningful gain range, the study can conclude that this mean value is not a gain in the tested environment. If the interval is maintained, then there is no sign of an ending.

Ablated areas may help determine why improvement has occurred:

Ablation	Question answered
Remove scenario generation	Will failure replay do?
Switch to uniform selection instead of targeted selection	Is the allocation rule helpful?
Apply supervised correction to the same situations.	Should reinforcement learning be required?
Replace the mixture of nominal data.	What is the impact of focused training on normal driving?
Only assess independent root domains.	Is it still true when related scenarios are deleted?

One factor should be changed for each ablation, and the other conditions kept constant.

8. Limitations

This paper presents an evaluation framework and no new policy-training experiments. Provides no measure of safety improvement, sample efficiency or generalization for a specific driving policy.

The literature discussion is a selective rather than a systematic review. New preprints are referenced as preprints and can be subject to modification prior to publication.

The results of the simulator tests are also dependent upon traffic behavior, simulator of sensors, geometry of the roads and the operating conditions. The effects of simulated weather or sensor degradation may not be the same as in the real world.

Pretrained assets and expert information also impact cost comparisons. When reporting pretraining and adaptation costs separately, there is greater transparency, but all the sources of imbalance are not eliminated.

Lastly, an empirical nominal-performance requirement only supports the tested distribution. Does not provide pointwise safety for arbitrary shifts.

The framework allows for well-defined robustness statements to be made. Further evidence is needed for physical deployment, formal guarantees and operational safety assurance.

9. Conclusion

There is extensive research currently in progress on learning from failure; this is an important direction for research. It should be evaluated based on its own safety enhancements, its ability to generalize to other applications, the ability to keep normal driving, and the resources needed to gain those benefits.

In a previous research on sensor fusion, we investigated perception in adverse weather [1]. This paper adds to the research focus on the assessment of driver behavior and failure based adaptation. It does not preclude that better "seeing" will lead to better decisions as far as safety is concerned.

Strong controls for continued-training, replay, prioritization and supervised-correction continue to be needed in the proposed framework. It also demands a complete adaptation-cost accounting and a partitioning of scenarios before mining.

These requirements form a practical means for determining when targeted learning will be useful, when targeted learning primarily serves to add computation, and when the value of targeted learning does not transfer.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Benjamin Stoler, Ingrid Navarro, Jonathan Francis, and Jean Oh, "SEAL: Towards Safe Autonomous Driving via Skill-Enabled Adversary Learning for Closed-Loop Scenario Generation," IEEE Robotics and Automation Letters, 2025. Author version available: https://benstoler.com/pdfs/seal.pdf.

[2] Daniel Dauner, Marcel Hallgarten, Tianyu Li, Xinshuo Weng, Zhiyu Huang, Zetong Yang, Hongyang Li, Igor Gilitschenski, Boris Ivanovic, Marco Pavone, Andreas Geiger, and Kashyap Chitta, "NAVSIM: Data-Driven Non-Reactive Autonomous Vehicle Simulation and Benchmarking," Advances in Neural Information Processing Systems, Datasets and Benchmarks Track, 2024. Available: https://proceedings.neurips.cc/paper_files/paper/2024/hash/32768f7faf1995026ef9821c696f3404-Abstract-Datasets_and_Benchmarks_Track.html.

[3] Davis Rempe, Jonah Philion, Leonidas J. Guibas, Sanja Fidler, and Or Litany, "Generating Useful Accident-Prone Driving Scenarios via a Learned Traffic Prior," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022. Available: https://openaccess.thecvf.com/content/CVPR2022/html/Rempe_Generating_Useful_Accident-Prone_Driving_Scenarios_via_a_Learned_Traffic_Prior_CVPR_2022_paper.html.

[4] Jiarong Wei, Yin Wu, Runkai He, and Abhinav Valada, "*CounterPlay: Counterfactual Post-Training for Self-Play Driving Policies,*" arXiv preprint arXiv:2609.21617, 2026. Available: https://arxiv.org/abs/2609.21617.

[5] Joshua Achiam, David Held, Aviv Tamar, and Pieter Abbeel, "Constrained Policy Optimization," Proceedings of the 34th International Conference on Machine Learning, 2017. Available: https://proceedings.mlr.press/v70/achiam17a.html.

[6] Keyu Chen, Nanfei Ye, Yida Wang, Wenchao Sun, Danqi Zhao, Hao Cheng, and Sifa Zheng, "*CRAFT: Counterfactual-to-Interactive Reinforcement Fine-Tuning for Driving Policies,*" arXiv preprint arXiv:2605.04470, 2026. Available: https://arxiv.org/abs/2605.04470.

[7] Krunal Panchal, Arpan Shaileshbhai Korat, and Saurav Rajanikant Pathak, "Sensor Fusion Using Machine Learning for Robust Object Detection in Adverse Weather Conditions for Self-Driving Cars," International Journal of Computational and Experimental Science and Engineering, vol. 11, no. 3, 2025. DOI: https://doi.org/10.22399/ijcesen.3589.

[8] Linrui Zhang, Zhenghao Peng, Quanyi Li, and Bolei Zhou, "CAT: Closed-loop Adversarial Training for Safe End-to-End Driving," Proceedings of the 7th Conference on Robot Learning, PMLR, vol. 229, pp. 2357–2372, 2023. Available: https://proceedings.mlr.press/v229/zhang23g.html.

[9] Minqi Jiang, Edward Grefenstette, and Tim Rocktäschel, "Prioritized Level Replay," Proceedings of the 38th International Conference on Machine Learning, PMLR, vol. 139, pp. 4940–4950, 2021. Available: https://proceedings.mlr.press/v139/jiang21b.html.

[10] Niklas Hanselmann, Katrin Renz, Kashyap Chitta, Apratim Bhattacharyya, and Andreas Geiger, "KING: Generating Safety-Critical Driving Scenarios for Robust Imitation via Kinematics Gradients," European Conference on Computer Vision, 2022. DOI: https://doi.org/10.1007/978-3-031-19839-7_20.

[11] Shuo Feng, Haojie Zhu, Haowei Sun, Xintao Yan, Linxuan He, Jingxuan Yang, Guangzhen Su, Boqi Li, Shu Li, Ling Wang, Shengyin Shen, and Henry X. Liu, "Breaking through safety performance stagnation in autonomous vehicles with dense learning," Nature Communications, vol. 17, article 3163, 2026. DOI: https://doi.org/10.1038/s41467-026-69761-x.

[12] Xiaosong Jia, Zhenjie Yang, Qifeng Li, Zhiyuan Zhang, and Junchi Yan, "Bench2Drive: Towards Multi-Ability Benchmarking of Closed-Loop End-To-End Autonomous Driving," Advances in Neural Information Processing Systems, Datasets and Benchmarks Track, 2024. Available: https://proceedings.nips.cc/paper_files/paper/2024/hash/017761f94a1cd66d01c041aff85492c4-Abstract-Datasets_and_Benchmarks_Track.html.

[13] Yinlam Chow, Aviv Tamar, Shie Mannor, and Marco Pavone, "Risk-Sensitive and Robust Decision-Making: A CVaR Optimization Approach," Advances in Neural Information Processing Systems, 2015. Available: https://papers.nips.cc/paper/6014-risk-sensitive-and-robust-decision-making-a-cvar-optimization-approach.

[14] Yuechen Luo, Fang Li, Qimao Chen, Shaoqing Xu, Jiaxin Liu, Ziying Song, Zhi-xin Yang, and Fuxi Wen, "Unleashing VLA Potentials in Autonomous Driving via Explicit Learning from Failures," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 24833–24842, 2026. Available: https://openaccess.thecvf.com/content/CVPR2026/html/Luo_Unleashing_VLA_Potentials_in_Autonomous_Driving_via_Explicit_Learning_from_CVPR_2026_paper.html.