RISE: ENHANCING SAFETY FOR END-TO-END AUTONOMOUS DRIVING THROUGH RISK-AWARE ENERGY-GUIDED DIFFUSION PLANNER

Anonymous authors
Paper under double-blind review

ABSTRACT

In end-to-end autonomous driving, vision-language-action (VLA) planners still tend to generate the high-risk trajectories in some long-tail scenarios. Existing safety-oriented approaches rely mostly on the post-training, which requires a costly fine-tuning of the pre-trained backbone and often degrades its general driving capability, thus leading to an inevitable safety-efficiency trade-off. In this paper, we propose RISE, a **RIS**k-aware **E**nergy-guided planning framework that incorporates these safety constraints into the trajectory generation of a diffusion planner, without any additional fine-tuning of the pre-trained VLA backbone. Specifically, we design a mixed energy guidance consisting of a rule-based guidance, an expert-based guidance and a risk assessor. The rule-based guidance constructs differentiable energy fields from drivable-area compliance and lane-keeping constraints, while the expert-based guidance takes reference trajectories from a safety-specialized expert trained separately via a safe reinforcement learning variant. Since the expert is inherently conservative, a lightweight rule-based risk assessor is further designed to activate the expert guidance only when potential risk is detected, thus preserving driving efficiency in ordinary scenarios. Experiments show that RISE improves the ReCogDrive baseline by 1.1 PDMS on NAVSIM (91.5) and by 24.6 PDMS on its high-risk subset, and achieves a Driving Score of 85.52 with a 59.09% Success Rate on the Bench2Drive closed-loop benchmark, outperforming the baseline by 12.03 and 17.27 percentage point, respectively.

1 INTRODUCTION

End-to-end autonomous driving (AD) has recently shifted towards the vision-language-action (VLA) architectures, which incorporate the visual perception, language understanding, and action generation. Typical frameworks use a pre-trained vision-language model (VLM) (Liu et al., 2023) as the backbone and a diffusion planner (Zheng et al., 2025; Bjorck et al., 2025; Chi et al., 2025a) as the action head, where the VLM encodes the driving scene into planning tokens that condition the generation of continuous trajectories. Such a design benefits from the world knowledge and reasoning ability of the VLM, and has shown strong generalization and instruction-following capabilities.

Despite their strong performance in the common scenarios (Fu et al., 2025; Xu et al., 2025; Jiang et al., 2025c), VLA planners still tend to generate the high-risk trajectories in some long-tail scenarios due to the scarcity of such data (Ye et al., 2025; Chi et al., 2025b; Ghosh et al., 2025; Yurt et al., 2025). To mitigate this, recent works resort to the post-training, which injects safety constraints by fine-tuning the pre-trained model with reinforcement learning (RL) (Jiang et al., 2025a; Zou et al., 2025). However, the RL-based post-training is computationally expensive, and the safety alignment tends to shift the model away from its

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.

original behavior distribution, degrading the general driving capability. This has motivated a training-free energy guidance (Liu et al., 2025a; Bao et al., 2022; Yao et al., 2026; Pei et al., 2026; Jiang et al., 2025b), where an energy field adjusts the generation process without updating the backbone. Along this line, rule-based energy guidance (LeCun et al., 2006; Zhao et al., 2022; Ho & Salimans, 2022; Tang et al., 2025) serves as a final safeguard and reduces the high-risk trajectories in these long-tail cases.

However, the purely rule-based guidance has two drawbacks. First, defining comprehensive driving rules is labor-intensive and often conflicting, making the energy function difficult to design. Second, applying guidance uniformly without risk distinction sacrifices the performance and may cause overly conservative driving even in some simple scenarios (Zhang et al., 2024). Moreover, we empirically observe that guiding the planner with a safety-specialized expert in all the scenarios improves collision avoidance but noticeably reduces the ego progress, which exhibits a typical safety-efficiency trade-off.

To address these issues, we propose RISE, a risk-aware energy-guided planning framework that incorporates both the rule-based and expert-based guidance through a risk assessor. Unlike other post-training methods, RISE requires no fine-tuning of the pre-trained VLA backbone, making it a low-intrusion safety alignment solution. It further overcomes the two drawbacks of the purely rule-based guidance by incorporating both the rule-based and expert-based guidance through a risk assessor. Specifically, the rule-based guidance enforces deterministic safety constraints (e.g., drivable area compliance) with a high interpretability, while the expert-based guidance, trained via Safe-GRPO (safe reinforcement learning incorporating the Group Relative Policy Optimization with Lagrangian relaxation), captures the complex semantics in challenging scenarios. Since the expert is inherently conservative, we further design a heuristic rule-based risk assessor that estimates the risk of the predicted trajectory in the current scene, and activates the expert guidance only when the risk exceeds a threshold. Our main contributions can be summarized as follows.

- We present RISE, a risk-aware energy-guided planning framework for autonomous driving VLA models, which incorporates *rule-based* guidance with high determinism and interpretability and *expert-based* guidance with rich semantics and better generalizability in complex high-risk scenarios. Unlike other post-training methods, RISE operates purely at inference time without fine-tuning pre-trained VLA backbone.
- We develop a safety-specialized expert via Safe-GRPO, and design a risk assessor that gates the expert guidance according to the risk level of the scene, so that conservative corrections are applied only in risky scenarios while the efficiency in ordinary scenarios is preserved, alleviating the safety-efficiency trade-off.
- Extensive experiments on the NAVSIM and Bench2Drive benchmarks show that RISE improves the ReCogDrive baseline by 1.1 PDMS on NAVSIM (91.5) and by 24.6 PDMS on its high-risk subset, and achieves a driving score of 85.52 with a 59.09% success rate on Bench2Drive (+12.03 DS and +17.27 SR).

2 Related Work

Reinforcement Learning in Autonomous Driving. Reinforcement learning has demonstrated unique capabilities in fields such as LLMs (Chen et al., 2025), embodied artificial intelligence (Zhang et al., 2025), and autonomous driving. The application of online reinforcement learning in autonomous driving often involves challenges, such as social safety and substantial costs, leading to the extensive development and adoption of various offline reinforcement learning algorithms (Noguchi & Yamamoto, 2025). However, with the development of simulators based on open-loop or closed-loop systems, such as NAVSIM (Dauner et al., 2024), online reinforcement learning methods have advanced, leading to the application of algorithms like GRPO. These approaches have gradually become crucial methods for autonomous driving. Many works have evolved numerous variants and optimizations based on GRPO.

Energy Guidance in Diffusion Models. In end-to-end autonomous driving systems, traditional VLA models typically employ diffusion models as action heads. While previous work has thoroughly explored the training of diffusion models (Liao et al., 2025), their prediction and generation capabilities have not re-

ceived a sufficient research attention. Recent work on energy guidance has led to the development of ideas for generating standardized diffusion models (Yu et al., 2023). Building upon this foundation, methods like Guideflow have also extended the concept of energy guidance into the field of autonomous driving. Energy-guided methods typically regulate the generation of diffusion models at a lower cost, such as training-free guidance approaches, yielding significant results.

3 PRELIMINARIES

3.1 DIFFUSION POLICY

In autonomous driving, VLA models are essential to extract rich semantic representations and comprehend high-level driving intents from the complex scenarios. However, directly decoding these latent representations into continuous trajectories remains challenging. A common solution is to formulate the trajectory generation process as a conditional diffusion model. Specifically, to ensure a real-time efficiency without sacrificing generation quality, we adopt the denoising diffusion implicit model (DDIM) (Song et al., 2022), a more efficient class of iterative implicit probabilistic models with the same training procedure as denoising diffusion probabilistic models (DDPMs) (Ho et al., 2020; Nichol & Dhariwal, 2021). DDIM no longer restricts the diffusion process to be a Markov chain, enabling to accelerate the generation process by using fewer sampling steps. Given a clean trajectory x_0, the Non-Markovian forward process is defined as $q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{x}_0) = \frac{q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{x}_0) q(\boldsymbol{x}_t|\boldsymbol{x}_0)}{q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_0)}$, where $q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t,\ \boldsymbol{x}_0) = \mathcal{N}\left(\sqrt{\alpha_{t-1}}\boldsymbol{x}_0 + \sqrt{1-\alpha_{t-1}-\sigma_t^2}\cdot \frac{\boldsymbol{x}_t - \sqrt{\alpha_t}\boldsymbol{x}_0}{\sqrt{1-\alpha_t}},\ \sigma_t^2 \boldsymbol{I}\right)$, and $q(\boldsymbol{x}_t|\boldsymbol{x}_0) = \mathcal{N}(\sqrt{\alpha_t}\boldsymbol{x}_0,\ (1-\alpha_t)\boldsymbol{I})$. Diffusion model trains a denoising network ϵ_θ to predict noises added to the clean trajectory:

$$\epsilon_\theta(x_t,\ t) \to \epsilon = \frac{x_t - \sqrt{\alpha_t} x_0}{\sqrt{1-\alpha_t}}. \tag{1}$$

At inference, trajectories are generated by initializing $x_T \sim N(0,\ I)$ and iteratively denoising via:

$$\boldsymbol{x}_{t-1} = \sqrt{\alpha_{t-1}} \underbrace{\left(\frac{\boldsymbol{x}_t - \sqrt{1-\alpha_t}\epsilon_\theta(\boldsymbol{x}_t)}{\sqrt{\alpha_t}}\right)}_{\text{“predicted } \boldsymbol{x}_0\text{”}} + \underbrace{\sqrt{1-\alpha_{t-1}-\sigma_t^2}\cdot \epsilon_\theta(\boldsymbol{x}_t)}_{\text{“direction pointing to } \boldsymbol{x}_t\text{”}} + \underbrace{\sigma_t \epsilon_t}_{\text{random noise}}, \tag{2}$$

where $\epsilon_t \sim N(0, I)$ is a standard Gaussian noise independent of x_t.

3.2 ENERGY GUIDANCE

The core idea of energy guidance is to define an energy function $\mathcal{E}(\cdot)$ that assigns low values to desirable (*e.g.*, safe) trajectories, and to steer the sampling process toward low-energy regions without retraining the denoiser. At each denoising step, the clean trajectory can be estimated as $\hat{x}_0 = \big(x_t - \sqrt{1-\bar{\alpha}_t}\,\epsilon_\theta(x_t, t)\big)/\sqrt{\bar{\alpha}_t}$. Although $\hat{x}_0$ depends on x_t nonlinearly through ϵ_θ, it is an affine function of the noise prediction for a fixed x_t. Hence, modifying the noise prediction directly controls the clean estimate and, in turn, the subsequent DDIM update. A common choice is forward energy guidance, which perturbs the noise prediction as $\hat{\epsilon}_\theta(x_t, t) = \epsilon_\theta(x_t, t) + \sqrt{1-\bar{\alpha}_t}\,\nabla_{x_t}\mathcal{E}(\hat{x}_0)$. However, the exact gradient requires back-propagating through the denoiser, and the Jacobian $\partial\epsilon_\theta/\partial x_t$ is usually dropped in practice, introducing an approximation error. Enlarging the guidance scale to compensate often causes instability, pushing the trajectory off the data manifold (Bansal et al., 2023). We therefore adopt backward guidance, which operates directly in the clean-trajectory space. Specifically, we compute a correction Δx_0 by approximately solving $\min_\Delta \mathcal{E}(\hat{x}_0 + \Delta)$ with m gradient-descent steps, $\Delta \leftarrow \Delta - \eta\nabla_\Delta \mathcal{E}(\hat{x}_0 + \Delta)$, initialized at $\Delta = 0$ with a small step size η, so that the energy is reduced gradually rather than by a single large update. Using the affine relation above, the noise prediction whose induced clean estimate equals $\hat{x}_0 + \Delta x_0$ is:

$$\tilde{\epsilon}(x_t, t) = \epsilon_\theta(x_t, t) - \sqrt{\bar{\alpha}_t/(1-\bar{\alpha}_t)}\,\Delta x_0. \tag{3}$$

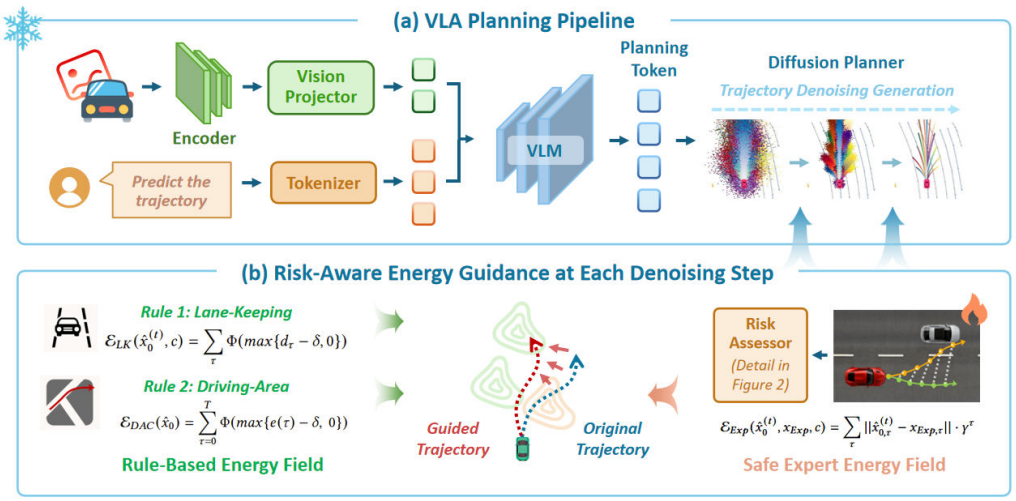

Figure 1: Overview of RISE. *(a) Basic VLA-based planning pipeline.* A VLM encodes multi-view images, ego state, and navigation commands into planning tokens, conditioning the diffusion planner to generate a raw trajectory via iterative denoising. *(b) Risk-aware energy guidance.* At each denoising step, the partially denoised trajectory is refined by two complementary energy fields: *i)* rule-based energy (e.g., lane-keeping and drivable-area compliance) enforcing strict safety constraints, and *ii)* expert-based energy derived from a safety-specialized expert and gated by a risk assessor (Fig. 3). The final trajectory is obtained after full denoising without fine-tuning the baseline model.

Replacing ϵ_θ with $\tilde{\epsilon}$ in Eq. (2) yields the guided DDIM update.

4 Proposed Method

4.1 Safety-Aware Energy-Guided VLA Planning

Fig. 1 illustrates the overall framework of RISE. It follows a basic VLA-based planning pipeline, where we adopt InternVL3-8B (Zhu et al., 2025) (InternViT visual encoder (Chen et al., 2024)) as the vision–language backbone and a 30M-parameter diffusion planner (Xiong et al., 2026) as the action head. The VLM encodes multimodal inputs (ego state S_{ego}, navigation commands L_{nav}, camera images I_{cam}) into planning tokens, which condition the diffusion planner to iteratively denoise a random trajectory $\mathbf{x}_t$ into a smooth and feasible trajectory.

To steer generation towards safer outcomes, we introduce risk-aware energy guidance into each denoising step. The guidance adjusts the predicted noise using the gradient of an energy function, as derived in Eq. (3). This operates solely on the action head and requires no fine-tuning of the VLA model. A safety-specialized expert is trained to provide reference trajectories only when the risk assessor detects potential danger. We consider two complementary energy fields: ***i) Rule-based energy field*** enforces hard safety constraints (e.g., lane-keeping and drivable area compliance); ***ii) Expert-based energy field*** is derived from a safety-specialized expert policy and controlled by a risk assessor. Together, they guide the diffusion process toward trajectories that are both safe and semantically plausible.

4.2 Rule-Based Guidance

We design rule-based energy guidance to provide explicit and explainable guidance signals from different physical perspectives, consisting of drivable area compliance (DAC) and lane-keeping (LK). These rules

construct energy fields that continuously guide the diffusion-based planner to generate trajectories strictly staying within the drivable area and adhering to the lane centerline.

Drivable Area Compliance (DAC) Guidance. To prevent the generated trajectories from violating road boundaries, we first extract information about the surrounding drivable areas and construct a binary energy grid. Following (Cheng et al., 2024), the initial discrete energy grid is defined as $M_{i,j} = \mathbf{1}\{(X_{\text{grid}}, Y_{\text{grid}}) \notin \bigcup_k P_k\}$, where P_k is the set of drivable areas (consisting of roadblocks, roadblock connectors, intersections, and carpark areas), and $M_{i,j} \in \{0, 1\}$ represents the penalty state of a specific grid cell (i, j). Since diffusion models require continuous and differentiable gradients to perform classifier-guided sampling, a hard binary boundary is sub-optimal. To provide a smoother gradient and avoid excessive trajectory correction, we apply a Gaussian blur to the discrete grid, softening the edges and allowing for a safe margin around the boundaries: $\tilde{M} = M * G_\sigma$, where $G_\sigma(u, v) = \frac{1}{2\pi\sigma^2} \exp(-\frac{u^2+v^2}{2\sigma^2})$. Next, for a given predicted trajectory $\hat{x}_0$, we calculate the continuous penalty value $e(\hat{x}_{0,\tau})$ for the trajectory waypoint at timestep τ. We utilize bilinear interpolation to map the discrete grid space $\tilde{M}$ to the continuous trajectory space: $e(x) = (1 - \alpha)(1 - \beta)\tilde{M}_{i_x,j_x} + \alpha(1 - \beta)\tilde{M}_{i_x+1,j_x} + (1 - \alpha)\beta\tilde{M}_{i_x,j_x+1} + \alpha\beta\tilde{M}_{i_x+1,j_x+1}$, where x is a random point in the continuous trajectory space, (i_x, j_x) is the top-left grid index corresponding to the given position x, and α and β are the fractional distances from the position to the discrete grid cell edges. Finally, the DAC energy function is formulated by accumulating the penalties over the entire trajectory horizon T:

$$\mathcal{E}_{DAC}(\hat{x}_0) = \sum_{\tau=1}^{T} \Phi(\max\{e(\hat{x}_{0,\tau}) - \delta_{DAC},\ 0\}), \tag{4}$$

where δ_{DAC} is the allowable error margin, and $\Phi(z) = e^z - z$. This exponential function provides strong, continuous gradients that significantly increase the energy penalty as the vehicle deviates further from the drivable area.

Lane-Keeping (LK) Guidance. In addition to ensuring the vehicle stays in the lane, we encourage the trajectory to follow the centerline of the current lane. For each waypoint τ, we calculate its lateral distance $d(\hat{x}_{0,\tau})$ to the target centerline (Werling et al., 2010). By allowing for a certain margin of error δ_{LK}, the lane-keeping energy function is defined as:

$$\mathcal{E}_{LK}(\hat{x}_0) = \sum_{\tau=1}^{T} \Phi(\max\{d(\hat{x}_{0,\tau}) - \delta_{LK},\ 0\}). \tag{5}$$

Overall Rule-Based Energy. Ultimately, the total rule-based energy guidance is formulated as a weighted sum of the DAC and LK energy components:

$$\mathcal{E}_{Rule}(\hat{x}_0) = \lambda_1 \mathcal{E}_{DAC}(\hat{x}_0) + \lambda_2 \mathcal{E}_{LK}(\hat{x}_0), \tag{6}$$

where λ_1 and λ_2 are scaling coefficients used to balance the importance of drivable area constraints and centerline adherence.

4.3 Expert-Based Guidance

Though rule-based guidance is effective, it struggles to cover complex interactions and may cause overly conservative behavior when applied uniformly. To complement the limited coverage of handcrafted rules, we introduce a safety-specialized expert model that encodes richer interactive priors. To mitigate the conservativeness of the expert, we further introduce a risk assessor that gates the expert guidance, activating it only when potential risk is detected. Throughout, the pre-trained backbone remains frozen, making RISE a low-intrusion solution. This decouples the safety-performance trade-off into two sub-problems: *i)* training a safety-specialized expert, and *ii)* using its trajectory as an energy reference, with the risk threshold acting as a tunable knob that determines when the expert should intervene.

For the expert, we propose Safe-GRPO based on safe reinforcement learning. The pre-trained backbone explores in NAVSIM, and simulator metrics are converted into reward and cost signals. The reward is a weighted sum of TTC, Comf., and EP, while the cost is $cost_i = 1 - NC \cdot DAC$, where NC and DAC denote No At-Fault Collision and Drivable Area Compliance. Following GRPO (Shao et al., 2024), the fused advantage is $\hat{A}^i_{\text{fused}} = (\hat{A}^i_r - \lambda_{\text{safe}} \cdot \hat{A}^i_c)/(1 + \lambda_{\text{safe}})$. The Safe-GRPO loss is

$$\mathcal{L} = -\frac{1}{G}\sum_{i=1}^{G}\frac{1}{T}\sum_{t=1}^{T}\gamma^{t-1}\log \pi_\theta(x^i_{t-1}|x^i_t)\hat{A}^i_{\text{fused}} - \lambda\frac{1}{G}\sum_{i=1}^{G}\frac{1}{T}\sum_{t=1}^{T}\log \pi_\theta(\tilde{x}^i_{t-1}|\tilde{x}^i_t), \tag{7}$$

where γ discounts early denoising steps, λ weights the behavior cloning loss, and $\tilde{x}_{t-1}, \tilde{x}_t$ are sampled from π_{ref}. The Lagrange multiplier updates as $\lambda_{\text{safe}} = \lambda_{\text{safe}} + \beta_{\lambda_{\text{safe}}}(\overline{cost} - \delta)$, with $\beta_{\lambda_{\text{safe}}}$ as learning rate, $\overline{cost}$ as group-average cost, and δ as a cost threshold. The trained expert then guides generation via the energy field

$$\mathcal{E}_{Exp}(\hat{x}^{(t)}_0, x_{Exp}) = \sum_{\tau=1}^{T}\|\hat{x}^{(t)}_{0,\tau} - x_{Exp,\tau}\| \cdot \gamma^\tau, \tag{8}$$

where $\hat{x}^{(t)}_0$ is the predicted clean trajectory at denoising step, and γ decays distant deviations. Since most pre-trained VLA models handle common scenarios well, expert guidance is only valuable in high-risk cases. We therefore introduce a risk assessor (Leng et al., 2025; Yang et al., 2021; Keswani & Bhattacharyya, 2025) that decides whether to activate expert guidance.

4.4 Risk Assessor

The expert model prioritizes safety and tends to produce conservative trajectories, while energy guidance extracts semantic information from these trajectories and converts it into guiding signals. Blindly applying a one-size-fits-all guidance is a root cause of the safety-performance trade-off. To address this, we introduce a risk assessor that detects potential risks in the current scene and dynamically decides whether to activate the safety-specialized expert. As illustrated in Fig. 3 in appendix, it applies guidance precisely to risky trajectories while preserving the original model's performance in simple scenarios, transforming the trade-off into scenario-adaptive decision-making.

Specifically, the rule-based risk assessor takes the ego's predicted trajectory $\hat{x}^{(t)}_0$ and environment conditions c as input and outputs a scalar risk score $s \in [0, 1]$, computed as a weighted sum of five sub-scores: obstacle distance, dynamic environment, trajectory complexity, traffic density, and drivable area compliance. Each sub-score captures a different aspect of driving risk, and their detailed definitions, and weights are provided in Appendix C. This modular formulation captures geometric, kinematic, and semantic constraints and integrates expert-driven modeling with rule-based guidance. Through this risk assessor, we can seamlessly integrate the expert-driven modeling with the rule-based guidance.

4.5 Implementation

As shown in Fig. 3, we use the risk assessor to incorporate the rule-based guidance with expert-based guidance. When the risk assessor detects a potential risk, it activates the expert model guidance, thus utilizing the semantically rich trajectories provided by the expert model to compensate for the shortcomings of the rule-based guidance. Ultimately, we define the overall energy function as:

$$\mathcal{E}(\hat{x}^{(t)}_0, x_{safe}, c) = \mathbf{1}\{\mathcal{R}(c, \hat{x}^{(t)}_0) > \delta\} \cdot \mathcal{E}_{Expert} + \alpha \cdot \mathcal{E}_{Rule}, \tag{9}$$

where α determines the weight assigned to the rule-based guidance, and δ represents the threshold of risk assessor, quantifying the gradient of the energy field that shapes the trajectory towards safety and conservatism. This approach enables a quantifiable trade-off between the performance and safety. We provide the pseudo-code of the generation process in Appendix B.

5 EXPERIMENTS

5.1 EXPERIMENTAL SETTINGS

For energy guidance, we set for the backward guidance in the equation we discussed in Subsection 3.2 learning rate $\lambda = 0.01$, and backward steps $m = 3$. For safety-specialized model in expert-based guidance, we employ the Safe-GRPO algorithm finetuning method to train the diffusion model for 25 epochs with cosine annealing learning rate schedule. For risk assessor, we adopt rule-based risk assessor, and set the risk threshold $\delta = 0.4$, the weights of each sub-metrics are as follows: $\omega_{OD} = 0.25$, $\omega_{DE} = 0.3$, $\omega_{TC} = 0.05$, $\omega_{TD} = 0.1$, $\omega_{DAC} = 0.3$. For the rule-based guidance, the drivable area and lane information are obtained from ground-truth annotations in open-loop experiments, while in closed-loop experiments we use YOLOPv2 (Han et al., 2022) as the perception module to provide these inputs.

Dataset. NAVSIM (Dauner et al., 2024) is a planning-oriented dataset built on OpenScene (Contributors, 2023), a redistribution of nuPlan (Caesar et al., 2021). It provides eight 1920×1080 cameras, a fused LiDAR point cloud from five sensors over the current and three previous frames, and 2 Hz annotations including HD maps and object bounding boxes, which are also used for our rule-based energy guidance. It splits into navtrain (1,192 scenes) for expert training and navtest (136 scenes) for evaluation.

Bench2Drive (Jia et al., 2024) is the first closed-loop benchmark for multi-ability assessment of end-to-end autonomous driving. Built on CARLA v2, it contains 2M annotated frames from 10,000 clips, covering 44 interactive scenarios, 23 weather conditions, and 12 towns, collected by the expert model Think2Drive. Evaluation requires passing 44 scenarios under different locations and weathers, totaling 220 short routes of 150 m each.

Metric. On NAVSIM, we use the predictive driver model score (PDMS): PDMS $=$ NC $\times$ DAC $\times$ $(5\text{EP} + 5\text{TTC} + 2\text{Comf})/12$, where NC, DAC, TTC, Comf, and EP denote No At-Fault Collision, Drivable Area Compliance, Time-to-Collision, Comfort, and Ego Progress.

On Bench2Drive (Jia et al., 2024), we report Driving Score (DS), Success Rate (SR), and Efficiency. DS is $\text{DS} = \frac{1}{N}\sum_{i=1}^{N} \text{RC}_i \prod_j p_{i,j}$, where N is the number of routes, RC_i is the route-completion percentage of route i, and $p_{i,j}$ is the penalty for the j-th infraction on route i. SR is n_{success}/N, and Efficiency checks whether the ego speed is excessively low compared to nearby traffic.

5.2 MAIN RESULTS

Experiments on NAVSIM Benchmark. Table 1 compares RISE with existing end-to-end methods on the NAVSIM *navtest* subset. RISE achieves the best PDMS of **91.5**, improving over the ReCogDrive baseline (90.4) by **+1.1** without additional training or fine-tuning for the VLA model. The gain mainly comes from safety: DAC improves from 97.7 to **99.0**, while NC (97.8 vs. 97.7) and TTC (94.8 vs. 94.6) are maintained. Meanwhile, EP increases from 86.2 to **87.1**, showing that the safety gain does not sacrifice progress. This distinguishes RISE from the safe-oriented expert model, which achieves higher NC (98.7) and TTC (96.2) but a much lower EP (81.1) and PDMS (89.5) due to overly conservative behavior. By combining mixed guidance with a risk assessor, RISE inherits the expert's safety advantages while avoiding unnecessary conservatism, leading to the highest overall PDMS on NAVSIM.

Experiments on Bench2Drive Benchmark. For brevity, we compare a representative subset of recent methods here; the complete table is provided in Appendix D.1. Table 2 presents the closed-loop planning results on Bench2Drive, where the baseline is our reproduced ReCogDrive. RISE achieves a driving score of 85.52 and a success rate of 59.09%, outperforming the baseline by +12.03 and +17.27%, respectively, demonstrating the effectiveness of risk-aware energy guidance in closed-loop execution. In terms of efficiency, RISE reaches 128.08, improving over the baseline by +6.48 while maintaining a favorable balance

Table 1: Performance on NAVSIM *navtest* subset.

Method	NC↑	DAC↑	TTC↑	Comf. ↑	EP↑	PDMS↑
Constant Velocity	68.0	57.8	50.0	**100**	19.4	20.6
Ego Status MLP	93.0	77.3	83.6	**100**	62.8	65.6
UniAD (Hu et al., 2023)	97.8	91.9	92.9	**100**	78.8	83.4
DiffusionDrive (Zou et al., 2025)	98.2	96.2	94.7	**100**	82.2	88.1
WoTE (Li et al., 2025b)	98.5	96.8	94.9	99.9	81.9	88.3
DriveVLA-W0-7B (Li et al., 2025a)	**98.7**	**99.1**	94.9	99.3	83.3	90.2
AdaThinkDrive-8B (Luo et al., 2025)	98.4	97.8	95.2	**100**	84.4	90.3
ReCogDrive (Baseline) (Xiong et al., 2026)	97.7	97.7	94.6	**100**	86.2	90.4
Safety-Specialized Expert Model (Ours)	**98.7**	98.5	**96.2**	**100**	81.1	89.5
Mixed Guidance with Risk Assessor (Ours)	97.8↑	99.0↑	94.8↑	**100**	87.1↑	**91.5**↑

Table 2: **Closed-loop planning and multi-ability results on Bench2Drive (Jia et al., 2024) benchmark.** Bold indicates the best performance per column. Values in parentheses show improvement over Baseline.

Method	Closed-loop Planning			Multi-Ability (%)					
	DS↑	SR(%)↑	Eff↑	Mer↑	Ovt↑	EB↑	GW↑	TS↑	Mean↑
Raw2Drive (Yang et al., 2025)	71.36	50.24	**214.17**	**43.35**	51.11	60.00	50.00	62.26	53.34
DriveTrans (Jia et al., 2025)	63.46	35.01	100.64	17.57	35.00	48.36	40.00	52.10	38.60
ReasonPlan (Liu et al., 2025b)	64.01	34.55	180.64	37.50	26.67	33.30	40.00	45.79	36.66
ReCogDrive (Xiong et al., 2026)	71.36	45.45	138.18	29.73	20.00	69.09	20.00	**71.34**	42.03
ORION (Fu et al., 2025)	77.74	54.62	151.48	25.00	71.11	**78.33**	30.00	69.15	54.72
ReCogDrive (Baseline)	73.49	41.82	121.60	20.00	75.56	51.67	**70.00**	21.05	47.65
RISE (Ours)	**85.52 (12.03↑)**	**59.09**↑	128.08↑	41.25↑	**97.78**↑	68.33↑	60.00	30.53↑	**59.58 (11.93↑)**

Table 3: Ablation study on the proposed components of RISE. Results are reported on the full NAVSIM benchmark (Average) and on a subset of high-risk scenarios (High-Risk). Values in parentheses denote the PDMS drop relative to the full model (ID 4).

ID	Rule-Based	Expert-Based	Risk Assessor	Average						High-Risk Subset					
				NC↑	DAC↑	TTC↑	Comf.↑	EP↑	PDMS↑	NC↑	DAC↑	TTC↑	Comf.↑	EP↑	PDMS↑
1	✗	✗	✗	97.7	97.7	94.6	**100**	86.2	90.4 (1.1↓)	77.4	76.8	74.6	**100**	50.3	52.2 (24.6↓)
2	✓	✗	✗	97.7	**98.9**	94.5	**100**	**87.1**	91.3 (0.2↓)	80.1	**93.2**	75.6	**100**	65.8	68.0 (8.8↓)
3	✗	✓	✗	**98.0**	97.9	**95.1**	**100**	86.0	90.7 (0.8↓)	91.0	88.3	84.5	99.7	67.8	72.3 (4.5↓)
4	✓	✓	✓	97.8	99.0	94.8	100	87.1	**91.5**	**92.0**	92.7	**85.3**	99.1	**72.4**	**76.8**

between safety and progress. Although some methods report higher efficiency, they often sacrifice driving score and success rate. RISE instead delivers the best overall closed-loop performance, validating its generalization to the Bench2Drive benchmark.

Computational Cost. RISE introduces additional inference latency compared to the baseline ReCogDrive (372.66 ms → 874.58 ms). This overhead arises from two sources: (1) backward guidance performs multiple gradient-descent steps per denoising step, following the universal-guidance paradigm; and (2) rule-based energy requires frequent map queries, trajectory-to-map matching, and CPU I/O. These are implementation limitations rather than fundamental ones. Future improvements include applying backward guidance only at selected denoising steps instead of every step, and precomputing map-based rule energies into GPU-resident differentiable cost maps to avoid repeated CPU I/O.

Table 4: Backward guidance weights.

Weight	NC↑	DAC↑	TTC↑	Comf.↑	EP↑	PDMS↑
0.002	**97.8**	97.6	**94.8**	**100.0**	86.1	90.3
0.01	**97.8**	**98.9**	**94.8**	**100.0**	**87.1**	**91.5**
0.05	97.3	98.9	93.9	98.8	86.5	90.6

Table 5: Risk-assessor threshold.

Thresh.	NC↑	DAC↑	TTC↑	Comf.↑	EP↑	PDMS↑
0.2	**97.9**	98.8	94.7	**100.0**	**87.1**	91.4
0.4	97.8	**99.0**	**94.8**	**100.0**	**87.1**	**91.5**
0.8	97.7	98.8	94.6	**100.0**	87.0	91.3

5.3 Ablation Study

Table 3 ablates the contribution of each component. Rule-based guidance alone (ID 2) improves PDMS from 90.4 to 91.3 on average and from 52.2 to 68.0 on the high-risk subset, serving as a deterministic safety backup. Using expert guidance without the risk assessor (ID 3) yields only a marginal gain on Average (90.7) and even underperforms rule-only, because the conservative expert is applied indiscriminately and harms efficiency in nominal scenarios. This confirms that expert guidance alone is not beneficial when applied uniformly. In contrast, the full RISE with the risk assessor (ID 4) achieves 91.5 PDMS on Average and 76.8 on the high-risk subset. The Average gain over rule-only is small (+0.2), which is expected: most scenarios are low-risk, so the expert is rarely activated and its effect is diluted. However, on the high-risk subset, the benefit becomes pronounced: PDMS improves by +24.6 over the no-guidance baseline and by +8.8 over rule-only, and even exceeds the ungated expert (ID 3) by +4.5. This demonstrates that *the risk assessor is essential*: it activates the expert only when needed, preserving nominal performance while delivering substantial safety gains in high-risk scenarios. Additional ablation studies on rule-based weights λ_1/λ_2 and expert weight α are provided in Appendix D.2.

Impact of Backward Guidance Weights. Table 4 shows the effect of the backward guidance weight. Too small a weight (e.g., 0.002) makes guidance negligible, while too large a weight (e.g., 0.05) overrides the diffusion model and distorts trajectories. A moderate value (0.01) corrects unsafe elements and achieves the best PDMS of 91.5.

Impact of Risk-Assessor Threshold. Table 5 studies the risk threshold δ. A low threshold (0.2) triggers expert guidance frequently, yielding conservative driving; a high threshold (0.8) activates it only in genuine danger, yielding more efficient driving. A balanced value (0.4) achieves the best PDMS of 91.5, allowing continuous tuning between conservativeness and aggressiveness.

Due to space limitations, additional results are provided in the appendix, including the training details of the safety-specialized expert (Appendix A), the comparison between rule-based and model-based risk assessors (Appendix C), the complete Bench2Drive results (Appendix D.1), sensitivity analyses of hyperparameters (Appendix D.2), and qualitative visualizations with failure cases (Appendix E).

6 Conclusion

In this work, we proposed RISE, a risk-aware energy-guided planning framework for autonomous driving. It enabled an efficient and controllable safety alignment, allowing customizable driving styles without compromising driving performance. Specifically, we employed a dual guidance mechanism: (i) rule-based guidance to provide deterministic safety constraints with high interpretability, and (ii) expert-based guidance, trained via safe reinforcement learning with Lagrangian relaxation, to capture complex semantics in challenging scenarios. A risk assessor fused these two signals to monitor the driving context. When potential risk was detected, expert-based energy guidance would be activated for safer guidance. We have evaluated RISE on both the NAVSIM and Bench2Drive. Across these closed-loop benchmarks, RISE consistently outperformed the baseline and other representative methods, demonstrating that risk-aware energy guidance provided an effective and low-intrusion solution for safe autonomous driving.

AI USE STATEMENT

In this work, we used generative AI tools to aid and polish the writing. We have not used generative AI tools for any of the required disclosure tasks, and the rest of the required disclosure tasks are not applicable to this work. Additionally, we used generative AI tools for editing the research paper to improve readability. We have reviewed all AI-assisted work. For example, we checked LLM-generated text for potential plagiarism through a manual literature survey and verified that all technical content is accurate and consistent with our original research. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

We provide the information necessary to reproduce our experiments in the main paper and appendices, including dataset and evaluation protocols, model architectures, training objectives, reward and cost definitions, hyperparameters, and hardware settings. The main paper describes the VLA backbone, the diffusion planner, the rule-based and expert-based energy formulations, and the risk assessor. The appendix provides detailed risk sub-score definitions, the Safe-GRPO training procedure, the Lagrange multiplier update rule, and hyperparameter sensitivity analyses. All datasets (NAVSIM and Bench2Drive) are publicly available and used in accordance with their intended purposes.

REFERENCES

Arpit Bansal, Hong-Min Chu, Avi Schwarzschild, Soumyadip Sengupta, Micah Goldblum, Jonas Geiping, and Tom Goldstein. Universal guidance for diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 843–852, 2023.

Fan Bao, Min Zhao, Zhongkai Hao, Peiyao Li, Chongxuan Li, and Jun Zhu. Equivariant energy-guided sde for inverse molecular design. *arXiv preprint arXiv:2209.15408*, 2022.

Johan Bjorck, Fernando Castañeda, Nikita Cherniadev, Xingye Da, Runyu Ding, Linxi Fan, Yu Fang, Dieter Fox, Fengyuan Hu, Spencer Huang, et al. Gr00t n1: An open foundation model for generalist humanoid robots. *arXiv preprint arXiv:2503.14734*, 2025.

Holger Caesar, Juraj Kabzan, Kok Seang Tan, Whye Kit Fong, Eric Wolff, Alex Lang, Luke Fletcher, Oscar Beijbom, and Sammy Omari. nuplan: A closed-loop ml-based planning benchmark for autonomous vehicles. *arXiv preprint arXiv:2106.11810*, 2021.

Peter Chen, Xiaopeng Li, Ziniu Li, Xi Chen, and Tianyi Lin. Spectral policy optimization: Coloring your incorrect reasoning in GRPO. In *2nd AI for Math Workshop @ ICML 2025*, 2025.

Zhe Chen, Jiannan Wu, Wenhai Wang, Weijie Su, Guo Chen, Sen Xing, Muyan Zhong, Qinglong Zhang, Xizhou Zhu, Lewei Lu, et al. Internvl: Scaling up vision foundation models and aligning for generic visual-linguistic tasks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 24185–24198, 2024.

Jie Cheng, Yingbing Chen, and Qifeng Chen. Pluto: Pushing the limit of imitation learning-based planning for autonomous driving. *arXiv preprint arXiv:2404.14327*, 2024.

Cheng Chi, Zhenjia Xu, Siyuan Feng, Eric Cousineau, Yilun Du, Benjamin Burchfiel, Russ Tedrake, and Shuran Song. Diffusion policy: Visuomotor policy learning via action diffusion. *The International Journal of Robotics Research*, 44(10-11):1684–1704, 2025a.

Haohan Chi, Huan-ang Gao, Ziming Liu, Jianing Liu, Chenyu Liu, Jinwei Li, Kaisen Yang, Yangcheng Yu, Zeda Wang, Wenyi Li, et al. Impromptu vla: Open weights and open data for driving vision-language-action models. *arXiv preprint arXiv:2505.23757*, 2025b.

OpenScene Contributors. Openscene: The largest up-to-date 3d occupancy prediction benchmark in autonomous driving. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Vancouver, Canada*, pp. 18–22, 2023.

Daniel Dauner, Marcel Hallgarten, Tianyu Li, Xinshuo Weng, Zhiyu Huang, Zetong Yang, Hongyang Li, Igor Gilitschenski, Boris Ivanovic, Marco Pavone, et al. Navsim: Data-driven non-reactive autonomous vehicle simulation and benchmarking. *Advances in Neural Information Processing Systems*, 37:28706–28719, 2024.

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.

Haoyu Fu, Diankun Zhang, Zongchuang Zhao, Jianfeng Cui, Dingkang Liang, Chong Zhang, Dingyuan Zhang, Hongwei Xie, Bing Wang, and Xiang Bai. Orion: A holistic end-to-end autonomous driving framework by vision-language instructed action generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 24823–24834, 2025.

Anurag Ghosh, Shen Zheng, Robert Tamburo, Khiem Vuong, Juan Alvarez-Padilla, Hailiang Zhu, Michael Cardei, Nicholas Dunn, Christoph Mertz, and Srinivasa G Narasimhan. Roadwork: A dataset and benchmark for learning to recognize, observe, analyze and drive through work zones. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 6132–6142, 2025.

Cheng Han, Qichao Zhao, Shuyi Zhang, Yinzi Chen, Zhenlin Zhang, and Jinwei Yuan. Yolopv2: Better, faster, stronger for panoptic driving perception, 2022. URL https://arxiv.org/abs/2208.11434.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Yihan Hu, Jiazhi Yang, Li Chen, Keyu Li, Chonghao Sima, Xizhou Zhu, Siqi Chai, Senyao Du, Tianwei Lin, Wenhai Wang, Lewei Lu, Xiaosong Jia, Qiang Liu, Jifeng Dai, Yu Qiao, and Hongyang Li. Planning-oriented autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17853–17862, June 2023.

Xiaosong Jia, Yulu Gao, Li Chen, Junchi Yan, Patrick Langechuan Liu, and Hongyang Li. Driveadapter: Breaking the coupling barrier of perception and planning in end-to-end autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 7953–7963, 2023a.

Xiaosong Jia, Penghao Wu, Li Chen, Jiangwei Xie, Conghui He, Junchi Yan, and Hongyang Li. Think twice before driving: Towards scalable decoders for end-to-end autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 21983–21994, 2023b.

Xiaosong Jia, Zhenjie Yang, Qifeng Li, Zhiyuan Zhang, and Junchi Yan. Bench2drive: Towards multi-ability benchmarking of closed-loop end-to-end autonomous driving. In *Advances in Neural Information Processing Systems, Datasets and Benchmarks Track*, 2024.

Xiaosong Jia, Junqi You, Zhiyuan Zhang, and Junchi Yan. Drivetransformer: Unified transformer for scalable end-to-end autonomous driving. In *International Conference on Learning Representations*, volume 2025, pp. 67227–67243, 2025.

Bo Jiang, Shaoyu Chen, Qing Xu, Bencheng Liao, Jiajie Chen, Helong Zhou, Qian Zhang, Wenyu Liu, Chang Huang, and Xinggang Wang. Vad: Vectorized scene representation for efficient autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8340–8350, 2023.

Bo Jiang, Shaoyu Chen, Qian Zhang, Wenyu Liu, and Xinggang Wang. Alphadrive: Unleashing the power of vlms in autonomous driving via reinforcement learning and reasoning. *arXiv preprint arXiv:2503.07608*, 2025a.

Hao Jiang, Zhipeng Zhang, Yu Gao, Zhigang Sun, Yiru Wang, Yuwen Heng, Shuo Wang, Jinhao Chai, Zhuo Chen, Hao Zhao, et al. Flowdrive: Energy flow field for end-to-end autonomous driving. *arXiv preprint arXiv:2509.14303*, 2025b.

Xuefeng Jiang, Yuan Ma, Pengxiang Li, Leimeng Xu, Xin Wen, Kun Zhan, Zhongpu Xia, Peng Jia, Xianpeng Lang, and Sheng Sun. Transdiffuser: Diverse trajectory generation with decorrelated multi-modal representation for end-to-end autonomous driving, 2025c.

Mahesh Keswani and Raunak Bhattacharyya. Learning safe autonomous driving policies using predictive safety representations. *arXiv preprint arXiv:2512.17586*, 2025.

Yann LeCun, Sumit Chopra, Raia Hadsell, M Ranzato, Fujie Huang, et al. A tutorial on energy-based learning. *Predicting structured data*, 1(0), 2006.

Bo Leng, Ran Yu, Wei Han, Lu Xiong, Zhuoren Li, and Hailong Huang. Risk-aware reinforcement learning for autonomous driving: Improving safety when driving through intersection, 2025.

Yingyan Li, Shuyao Shang, Weisong Liu, Bing Zhan, Haochen Wang, Yuqi Wang, Yuntao Chen, Xiaoman Wang, Yasong An, Chufeng Tang, Lu Hou, Lue Fan, and Zhaoxiang Zhang. Drivevla-w0: World models amplify data scaling law in autonomous driving, 2025a.

Yingyan Li, Yuqi Wang, Yang Liu, Jiawei He, Lue Fan, and Zhaoxiang Zhang. End-to-end driving with online trajectory evaluation via bev world model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 27137–27146, October 2025b.

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 *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 12037–12047, 2025.

Haochen Liu, Tianyu Li, Haohan Yang, Li Chen, Caojun Wang, Ke Guo, Haochen Tian, Hongchen Li, Hongyang Li, and Chen Lv. Reinforced refinement with self-aware expansion for end-to-end autonomous driving. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2026.

Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. Visual instruction tuning. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 34892–34916, 2023.

Lin Liu, Caiyan Jia, Guanyi Yu, Ziying Song, JunQiao Li, Feiyang Jia, Peiliang Wu, Xiaoshuai Hao, and Yandan Luo. Guideflow: Constraint-guided flow matching for planning in end-to-end autonomous driving. *arXiv preprint arXiv:2511.18729*, 2025a.

Xueyi Liu, Zuodong Zhong, Yuxin Guo, Yun-Fu Liu, Zhiguo Su, Qichao Zhang, Junli Wang, Yinfeng Gao, Yupeng Zheng, Qiao Lin, et al. Reasonplan: Unified scene prediction and decision reasoning for closed-loop autonomous driving. *arXiv preprint arXiv:2505.20024*, 2025b.

Yuhang Lu, Jiadong Tu, Yuexin Ma, and Xinge Zhu. Real-ad: Towards human-like reasoning in end-to-end autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 27783–27793, 2025.

Yuechen Luo, Fang Li, Shaoqing Xu, Zhiyi Lai, Lei Yang, Qimao Chen, Ziang Luo, Zixun Xie, Shengyin Jiang, Jiaxin Liu, Long Chen, Bing Wang, and Zhi xin Yang. Adathinkdrive: Adaptive thinking via reinforcement learning for autonomous driving, 2025.

Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International conference on machine learning*, pp. 8162–8171. PMLR, 2021.

Chihiro Noguchi and Takaki Yamamoto. Offline reinforcement learning for end-to-end autonomous driving, 2025.

Shuo Pei, Yong Wang, Yuanchen Zhu, Chen Sun, Qin Li, Yanan Zhao, and Huachun Tan. Safe and stylized trajectory planning for autonomous driving via diffusion model. *arXiv preprint arXiv:2602.04329*, 2026.

Katrin Renz, Long Chen, Elahe Arani, and Oleg Sinavski. Simlingo: Vision-only closed-loop autonomous driving with language-action alignment. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 11993–12003, 2025.

Shuyao Shang, Yuntao Chen, Yuqi Wang, Yingyan Li, and ZHAO-XIANG ZHANG. Drivedpo: Policy learning via safety dpo for end-to-end autonomous driving. *Advances in Neural Information Processing Systems*, 38:81565–81585, 2026.

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.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models, 2022.

Zhicong Tang, Jianmin Bao, Dong Chen, and Baining Guo. Diffusion models without classifier-free guidance. *arXiv preprint arXiv:2502.12154*, 2025.

Moritz Werling, Julius Ziegler, Sören Kammel, and Sebastian Thrun. Optimal trajectory generation for dynamic street scenarios in a frenet frame. In *2010 IEEE international conference on robotics and automation*, pp. 987–993, 2010.

Penghao Wu, Xiaosong Jia, Li Chen, Junchi Yan, Hongyang Li, and Yu Qiao. Trajectory-guided control prediction for end-to-end autonomous driving: A simple yet strong baseline. *Advances in Neural Information Processing Systems*, 35:6119–6132, 2022.

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.

Chengkai Xu, Jiaqi Liu, Yicheng Guo, Peng Hang, and Jian Sun. A knowledge-driven diffusion policy for end-to-end autonomous driving based on expert routing, 2025.

Qisong Yang, Thiago D Simão, Simon H Tindemans, and Matthijs TJ Spaan. Wcsac: Worst-case soft actor critic for safety-constrained reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 10639–10646, 2021.

Zhenjie Yang et al. Raw2drive: Reinforcement learning with aligned world models for end-to-end autonomous driving. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.

Xuemei Yao, Xiao Yang, Jianbin Sun, Liuwei Xie, Xuebin Shao, Xiyu Fang, Hang Su, and Kewei Yang. Reflexdiffusion: Reflection-enhanced trajectory planning for high-lateral-acceleration scenarios in autonomous driving. *arXiv preprint arXiv:2601.09377*, 2026.

Haotian Ye, Kaiwen Zheng, Jiashu Xu, Puheng Li, Huayu Chen, Jiaqi Han, Sheng Liu, Qinsheng Zhang, Hanzi Mao, Zekun Hao, et al. Data-regularized reinforcement learning for diffusion models at scale. *arXiv preprint arXiv:2512.04332*, 2025.

Jiwen Yu, Yinhuai Wang, Chen Zhao, Bernard Ghanem, and Jian Zhang. Freedom: Training-free energy-guided conditional diffusion model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 23174–23184, 2023.

Mahmut Yurt, Xin Ye, Yunsheng Ma, Jingru Luo, Abhirup Mallik, John Pauly, Burhaneddin Yaman, and Liu Ren. Ltda-drive: Llms-guided generative models based long-tail data augmentation for autonomous driving. *arXiv preprint arXiv:2505.18198*, 2025.

Jiang-Tian Zhai, Ze Feng, Jinhao Du, Yongqiang Mao, Jiang-Jiang Liu, Zichang Tan, Yifu Zhang, Xiaoqing Ye, and Jingdong Wang. Rethinking the open-loop evaluation of end-to-end autonomous driving in nuscenes. *arXiv preprint arXiv:2305.10430*, 2023.

Borong Zhang, Yuhao Zhang, Jiaming Ji, Yingshan Lei, Josef Dai, Yuanpei Chen, and Yaodong Yang. Safevla: Towards safety alignment of vision-language-action model via constrained learning, 2025.

Zheng Zhang, Qi Liu, Yanjie Li, Ke Lin, and Linyu Li. Safe reinforcement learning in autonomous driving with epistemic uncertainty estimation. *IEEE Transactions on Intelligent Transportation Systems*, 25(10): 13653–13666, 2024.

Min Zhao, Fan Bao, Chongxuan Li, and Jun Zhu. Egsde: Unpaired image-to-image translation via energy-guided stochastic differential equations. *Advances in Neural Information Processing Systems*, 35:3609–3623, 2022.

Wenzhao Zheng, Ruiqi Song, Xianda Guo, Chenming Zhang, and Long Chen. Genad: Generative end-to-end autonomous driving. In *European Conference on Computer Vision*, pp. 87–104. Springer, 2024.

Yinan Zheng, Ruiming Liang, Kexin Zheng, Jinliang Zheng, Liyuan Mao, Jianxiong Li, Weihao Gu, Rui Ai, Shengbo Eben Li, Xianyuan Zhan, et al. Diffusion-based planning for autonomous driving with flexible guidance. *arXiv preprint arXiv:2501.15564*, 2025.

Zewei Zhou, Tianhui Cai, Seth Zhao, Yun Zhang, Zhiyu Huang, Bolei Zhou, and Jiaqi Ma. Autovla: A vision-language-action model for end-to-end autonomous driving with adaptive reasoning and reinforcement fine-tuning. *Advances in Neural Information Processing Systems*, 38:27920–27956, 2026.

Jinguo Zhu, Weiyun Wang, Zhe Chen, Zhaoyang Liu, Shenglong Ye, Lixin Gu, Hao Tian, Yuchen Duan, Weijie Su, Jie Shao, et al. Internvl3: Exploring advanced training and test-time recipes for open-source multimodal models. *arXiv preprint arXiv:2504.10479*, 2025.

Jialv Zou, Shaoyu Chen, Bencheng Liao, Zhiyu Zheng, Yuehao Song, Lefei Zhang, Qian Zhang, Wenyu Liu, and Xinggang Wang. Diffusiondrivev2: Reinforcement learning-constrained truncated diffusion modeling in end-to-end autonomous driving, 2025.

A Training Process of Expert Model

We conduct our Safe-GRPO fine-tuning based on a pretrained IL checkpoint (ReCogDrive DiT planner) for 25 epochs with a batch size of 8, using the AdamW optimizer and a cosine learning rate schedule that decays from 1×10^{-4} to 1×10^{-7}. The cost limit is set to 0.05, and the Lagrange multiplier learning rate is 0.05. A behavior cloning coefficient of 0.1 is applied to regularize the policy. The group size for GRPO is 8, and gradients are clipped at a maximum norm of 1.0. To stabilize diffusion-based policy learning, we clip the denoised output to ± 1.0, set the minimum denoising standard deviation during training to 0.04, and clamp the standard deviation used for log-probability computation to at least 0.1. A discount factor $\gamma = 0.9$ is applied to downweight early denoising steps, and advantages are clipped at the upper quantile of 1.0.

Table 6: Core training parameters of Safe Expert Model.

Parameter	**Value**
Batch size	8
Learning rate	1×10^{-4}
Learning rate scheduler	Cosine
Max epochs	25
Cost limit	0.05
Lambda learning rate	0.05
BC coefficient	0.1
Group size	8
Gradient clip value	1.0
Denoised output clip value	1.0
Min denoising std (training)	0.04
Min logprob denoising std	0.1
Discount factor for denoising (γ)	0.9
Advantage upper quantile	1.0

The detailed training process is presented in Algorithm 1. Figure 2 presents the training process of the expert model over 25 epochs, reporting cost (to be minimized) and reward (to be maximized), and the dynamic process of Lagrangian multiplier λ. We observe a common pattern: during the first few epochs, the cost increases slightly while the reward decreases modestly. This initial performance drop is not indicative of algorithmic failure but rather reflects the exploration phase of the safe reinforcement learning process. The agent actively probes the environment to discover high-reward, low-cost regions, which inevitably leads to occasional suboptimal decisions. As training progresses (beyond epoch 5), the cost consistently decreases and reward rises, demonstrating effective convergence. Through online reinforcement learning, our safety-specialized expert network tends to output more conservative and safer trajectories. The energy field required for the core model's energy guidance is characterized.

B Pseudo-code of the RISE Generation Process

Algorithm 2 summarizes the generation process of RISE. Given the state condition c and the pre-trained expert model π_{exp}, RISE first obtains an expert trajectory x_{Exp} as a safety reference. Starting from Gaussian noise x_T, the algorithm iteratively denoises over T steps. At each predefined guidance timestep, it estimates the clean trajectory $\hat{x}_0^{(t)}$ and computes the accumulated shift Δx_0 by minimizing the energy function via gradient descent. The rule-based energy $\mathcal{E}_{Rule}$ is then evaluated using the DAC and LK terms (Eqs. equation 4 and equation 5), and the noise prediction is refined to $\tilde{\epsilon}$ following Eq. equation 3. After the DDIM

update, the risk assessor evaluates the current scenario and the guided trajectory. If potential risk is detected, the expert energy $\mathcal{E}_{Exp}$ is computed and an additional energy guidance step is applied. The final denoised trajectory x_0 is returned after all steps.

Algorithm 1 Training Method of Safe-GRPO

Require: state condition c, initial noisy sample x_T, group size G, denoiser model π_θ, reward function $R(\cdot)$, cost function $C(\cdot)$

procedure SAFE_GRPO_STEP(c, x_T)
  for each replica $g = 1 \dots G$ **do**
    Initialize noisy trajectories: $x_T^{(g)} \leftarrow x_T$
    Get denoising chain by Eq. equation 2
    Evaluate terminal reward and cost by functions mentioned above
  end for
  Compute standardized advantages
  Aggregate per-step log-probabilities into a single scalar per chain to obtain the total loss Eq. equation 7
  Update Lagrange multiplier
  return $\{\mathcal{L}, \lambda_{safe}\}$
end procedure

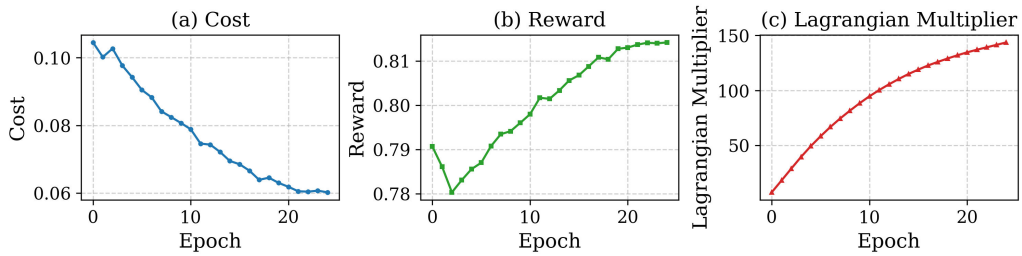

Figure 2: Training process of the expert model.

C Detailed Risk Assessor Design

As shown in Fig. 3, the risk assessor takes the ego's predicted clean trajectory $\hat{x}_0^{(t)}$ and environment conditions c as input, and outputs a scalar risk score $\mathcal{R} \in [0, 1]$.

The risk assessor takes the ego's predicted clean trajectory $\hat{x}_0^{(t)}$ and environment conditions c as input, and outputs a scalar risk score $s \in [0, 1]$. The score is a weighted sum of five sub-scores, each capturing a different aspect of driving risk. Below we describe each sub-score following the definitions in Sec. 4.4.

C.1 Obstacle Distance Score

We predict the future positions of surrounding dynamic obstacles over a horizon T (up to 2.0 s) using a constant-acceleration model derived from their recent and historical states. The minimum distance between the ego trajectory and all obstacles over the horizon is computed. The obstacle distance score S_{OD} is defined via a step-wise mapping function that heavily penalizes close encounters.

Algorithm 2 Generation process of RISE.

Require: state condition c, pre-defined parameters β_t, α_t, guidance weight α, ρ, pretrained expert model π_{exp}

procedure RISE_STEP(c, α)
  get expert trajectory $x_{Exp} = \pi_{exp}(c)$
  $x_T \sim \mathcal{N}(0, I)$
  for t = T, ..., 1 **do**
    if t in guidance timestep list **then**
      compute Δx_0 by minimizing equation discussed in Section 3.2 with gradient descent
      compute $\mathcal{E}_{Rule}(\hat{x}_0^{(t)})$ by Eq. equation 4 and Eq. equation 5
      improve $\tilde{\epsilon}$ by Eq. equation 3
      $x_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\hat{x}_0^{(t)} + \sqrt{1 - \bar{\alpha}_{t-1} - \sigma_t^2} \cdot \tilde{\epsilon} + \sigma_t \mathcal{N}(0, I)$
      detect potential risk within guided $\tilde{x}_0$ and current scenario using risk assessor
      if Potential Risk Detected **then**
        compute $\mathcal{E}_{Exp}(\hat{x}_0^{(t)},\ x_{Exp})$ by Eq. equation 8
        apply energy guidance using the method described above
      end if
    end if
  end for
  return denoised trajectory x_0
end procedure

C.2 Dynamic Environment Score

High-speed environments increase collision severity and reduce reaction time. We extract the maximum speed $v_{\max}$ and mean speed v_{mean} of surrounding vehicles. The dynamic environment score is a linear combination of the peak and average speed penalties:

$$S_{\text{DE}} = \alpha \cdot f(v_{\max}) + (1 - \alpha) \cdot g(v_{\text{mean}}),$$

where the penalty functions $f(\cdot)$ and $g(\cdot)$ map speeds to risk values in $[0, 1]$.

C.3 Trajectory Complexity Score

Complex and aggressive trajectories increase the likelihood of losing control or unexpected interactions. We evaluate trajectory complexity using the mean absolute steering angle $\bar{\theta}$ and the mean absolute angular velocity $\bar{\omega}$. The trajectory complexity score is defined as the maximum of the two corresponding penalty mappings:

$$S_{\text{TC}} = \max\left(\mathcal{M}_{\text{steer}}(\bar{\theta}), \mathcal{M}_{\text{ang_vel}}(\bar{\omega})\right),$$

where both mappings assign higher scores to more aggressive turning behaviors.

C.4 Traffic Density Score

Crowded scenes restrict drivable space and increase the probability of multi-agent interactions. The traffic density score S_{TD} is a function of the total number of dynamic road users.

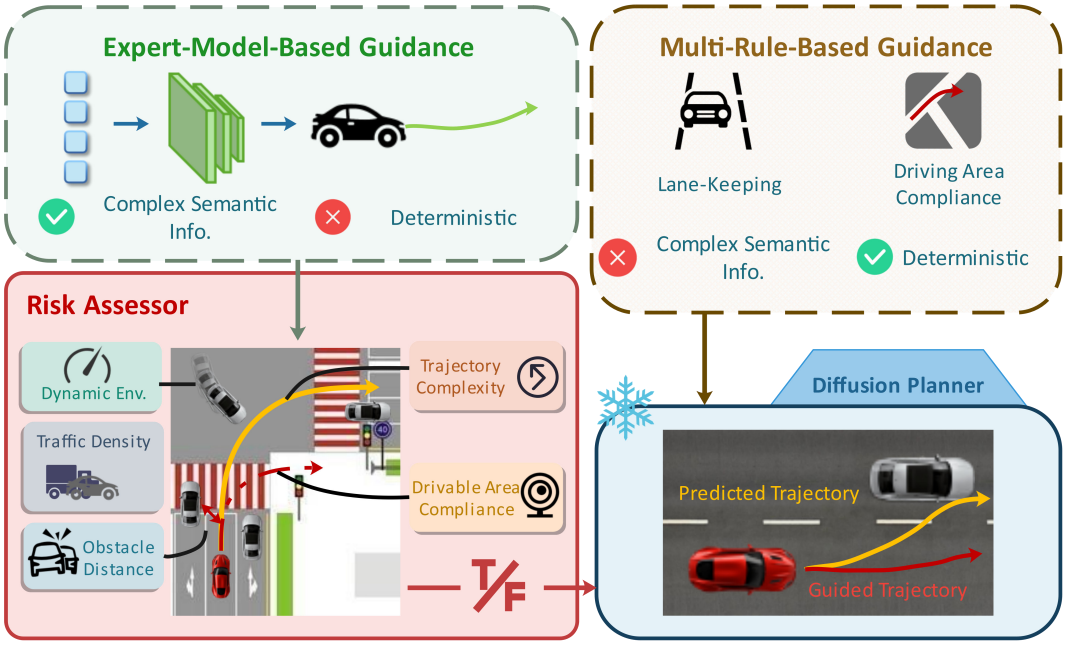

Figure 3: Proposed energy guidance method. When it comes to safety-related rule-based guidance, we generally aim for zero violations. The part of rule-based guidance is therefore always enabled. However, the complex semantics of expert models are only valuable in complex or high-risk scenarios. Therefore, whether expert model guidance is enabled is controlled by the risk assessor.

C.5 Drivable Area Compliance Score

To ensure the vehicle remains within safe boundaries, we evaluate the trajectory against the HD map's drivable polygons (e.g., drivable areas, intersections). If any of the four corners of the vehicle footprint falls outside the drivable polygons at any predicted time step within the horizon, a strict violation penalty is applied:

$$S_{\text{DAC}} = \begin{cases} 1.0, & \text{if the trajectory exits the drivable area,} \\ 0.0, & \text{otherwise.} \end{cases}$$

C.6 Aggregated Risk Score

The final aggregated risk score is computed as a weighted sum of the five sub-scores, clipped to the interval $[0, 1]$:

$$\mathcal{R}(c, \hat{x}_0^{(t)}) = \text{clip}\left\{ \sum_{i \in \{\text{OD,DE,TC,TD,DAC}\}} \omega_i \cdot S_i,\ 0.0,\ 1.0 \right\},$$

where ω_i are normalized tunable weights for each sub-score. This modular formulation captures geometric, kinematic, and semantic constraints, and integrates expert-driven modeling with rule-based guidance.

C.7 Comparison between Different Risk Assessors

We compare two alternative implementations of the risk assessor: a model-based predictor and our proposed rule-based assessor. The model-based assessor is trained via supervised learning to regress a risk score labeled by submetrics of PDMS. However, it suffers from several critical drawbacks. First, it is notoriously difficult to train due to the inherent ambiguity of risk labels and the high-dimensional nature of driving scenes. Second, a non-negligible distribution shift exists between the training and test sets, which is further exacerbated by the generative nature of the diffusion planner, causing the learned assessor to produce unreliable scores. Third, the network requires a complete forward pass at each denoising step, incurring substantial computational overhead and slowing down the overall planning frequency. In contrast, our rule-based assessor computes risk analytically from handcrafted metrics without any trainable parameters. It is lightweight, deterministic, and generalizes robustly across scenes. As shown in Table 7, the rule-based assessor achieves comparable or better scores on several submetrics (especially DAC and EP) while requiring no training and minimal runtime. Therefore, we adopt the rule-based risk assessor in RISE to balance effectiveness and real-time performance.

Table 7: Comparison between model-based and rule-based risk assessors on the NAVSIM v1 navtest subset.

Type	NC	DAC	EP	TTC	Comf.	PDMS
Model-based	**98.3**	97.9	86.7	**95.9**	**100.0**	91.3
Rule-based	97.8	**98.9**	**87.1**	94.8	**100.0**	**91.5**

D More Experiments

D.1 Full Results on Bench2Drive

Table 8 presents the complete comparison on the Bench2Drive benchmark. Methods marked with † are trained on the PDM-Lite expert, which provides higher-quality demonstrations than the official Bench2Drive expert (Think2Drive); these are shown in gray for reference only and are not directly comparable. RISE is trained under the standard Think2Drive setting and achieves the best overall performance among comparable methods.

D.2 Sensitivity to Hyperparameters

We analyze the sensitivity of RISE to two more key hyperparameters: the rule-based weight ratio λ_1/λ_2 and the expert guidance weight α. Tables 9 and 10 report the results on the NAVSIM *navtest* subset.

Sensitivity to rule-based weights λ_1/λ_2. Table 9 shows that performance is largely insensitive to the choice of λ_1/λ_2. Across the tested values (5×10^{-3}, 5×10^{-2}, and 5×10^{-1}), PDMS remains within 91.3–91.5. NC and DAC vary by at most 0.1, TTC stays in the range 94.7–94.8, comfort is consistently 100.0, and EP ranges from 86.9 to 87.1. The best performance is achieved at $\lambda_1/\lambda_2 = 5\times10^{-2}$, where NC, DAC, TTC, EP, and PDMS all reach their highest values. These results confirm that the rule-based guidance is robust to the weight ratio, as the two energy terms (drivable area compliance and lane-keeping) provide complementary yet non-conflicting constraints.

Sensitivity to expert weight α. Table 10 shows similarly stable behavior for α. PDMS remains between 91.3 and 91.5, with the best result at $\alpha = 5 \times 10^{-2}$. TTC varies only slightly (94.5–94.8), EP stays within 86.9–87.1, and comfort is constantly 100.0. This indicates that the expert guidance weight is not a sensitive parameter in the tested range. The reason is that the expert energy is already gated by the risk assessor: it

Table 8: **Complete closed-loop planning and multi-ability results on the Bench2Drive benchmark (Jia et al., 2024).** Bold indicates the best performance among methods trained on Think2Drive. Methods marked with † are trained on PDM-Lite and are shown in gray for reference only.

Method	Closed-loop Planning			Multi-Ability (%)					
	DS↑	SR↑	Eff↑	Mer↑	Ovt↑	EB↑	GW↑	TS↑	Mean↑
Trained on PDM-Lite expert (reference only)									
AutoVLA† (Zhou et al., 2026)	78.84	57.73	146.93	–	–	–	–	–	–
SimLingo† (Renz et al., 2025)	85.94	66.82	244.18	54.01	57.04	88.33	53.33	82.45	67.03
R2SE† (Liu et al., 2026)	86.28	69.54	243.89	53.33	61.25	90.00	50.00	84.21	67.76
Trained on Think2Drive expert (partial multi-ability)									
GenAD (Zheng et al., 2024)	44.81	15.90	–	–	–	–	–	–	–
DriveDPO (Shang et al., 2026)	62.02	30.62	166.80	–	–	–	–	–	–
ReAL-AD (Lu et al., 2025)	41.17	11.36	–	–	–	–	–	–	–
RAP (Feng et al., 2026)	66.42	37.27	165.47	–	–	–	–	–	–
Trained on Think2Drive expert (full)									
TCP-traj (Wu et al., 2022)	59.90	30.00	76.54	8.89	24.29	51.67	40.00	46.28	34.22
AD-MLP (Zhai et al., 2023)	48.45	0.00	48.45	0.00	0.00	0.00	0.00	4.35	0.87
UniAD-Base (Hu et al., 2023)	45.81	16.36	129.21	14.10	17.78	21.67	10.00	14.21	15.55
VAD (Jiang et al., 2023)	42.35	15.00	157.94	8.11	24.44	18.64	20.00	19.15	18.07
ThinkTwice (Jia et al., 2023b)	62.44	31.23	69.33	27.38	18.42	35.82	50.00	54.23	37.17
DriveAdapter (Jia et al., 2023a)	64.22	33.08	70.22	28.82	26.38	48.76	50.00	56.43	42.08
DriveTrans. (Jia et al., 2025)	63.46	35.01	100.64	17.57	35.00	48.36	40.00	52.10	38.60
ReasonPlan (Liu et al., 2025b)	64.01	34.55	**180.64**	37.50	26.67	33.30	40.00	45.79	36.66
ReCogDrive (Xiong et al., 2026)	71.36	45.45	138.18	29.73	20.00	69.09	20.00	**71.34**	42.03
ORION (Fu et al., 2025)	77.74	54.62	151.48	25.00	71.11	**78.33**	30.00	69.15	54.72
Baseline	73.49	41.82	121.60	20.00	75.56	51.67	**70.00**	21.05	47.65
RISE (Ours)	**85.52 (12.03↑)**	**59.09**↑	128.08↑	41.25↑	**97.78**↑	68.33↑	60.00	30.53↑	**59.58 (11.93↑)**

Table 9: Sensitivity to λ_1/λ_2

λ_1/λ_2	NC↑	DAC↑	TTC↑	Comf.↑	EP↑	PDMS↑
0.005	97.7	98.8	94.7	**100.0**	86.9	91.3
0.05	**97.8**	**98.9**	**94.8**	**100.0**	**87.1**	**91.5**
0.5	**97.8**	**98.9**	94.7	**100.0**	87.0	91.4

Table 10: Sensitivity to α

α	NC↑	DAC↑	TTC↑	Comf.↑	EP↑	PDMS↑
0.005	97.7	**98.9**	94.5	**100.0**	**87.1**	91.3
0.05	**97.8**	**98.9**	**94.8**	**100.0**	**87.1**	**91.5**
0.5	**97.8**	98.7	94.7	**100.0**	86.9	91.3

is activated only in high-risk scenarios, and its influence is adaptively adjusted per scenario by the energy field. As a result, moderate changes in α do not significantly alter the overall behavior.

E Qualitative Visualization

Drivable-area compliance. As shown in Fig. 4 (a) , the baseline model generates trajectories that cut corners or drift out of the curve during sharp turns, which increases the risk of veering out of the drivable area. After applying drivable-area and lane-keeping guidance, RISE shifts the trajectory back toward the lane center and keeps a safer margin to the curb/edge. This qualitative behavior is consistent with the quantitative improvement in DAC, indicating fewer off-road and boundary-violating maneuvers.

Complex urban scenarios. Fig. 4 (b) further demonstrates that in crowded intersections with dense traffic participants, the baseline trajectory may adopt a more aggressive path that passes close to surrounding vehicles. In contrast, RISE can identify potential interaction risks and guide the baseline model to generate



(a) Drivable Area Compliance (b) Complex Urban Scenarios (c) Simple Scenarios

Figure 4: **Qualitative results of RISE on NAVSIM.** Orange dashed: ReCogDrive baseline; red solid: RISE. In sharp turns, rule-based guidance corrects lane departure and maintains a safe margin. In complex interactions where handcrafted rules are insufficient, the risk assessor activates expert guidance to produce a more cautious path. In simple scenarios, it keeps expert guidance inactive to preserve the baseline's efficiency. This shows how RISE overcomes the safety-efficiency trade-off by dynamically integrating both guidance sources.

a trajectory with a larger safety margin and greater caution, while still maintaining forward speed. This demonstrates the significant benefits of integrating the mixed guidance architecture with the risk evaluator: it enables the planner to apply expert-based guidance only when in high-risk scenarios, supplementing the semantically rich energy gradient to address the shortcomings inherent in rule-based guidance alone, thereby ensuring trajectory safety in high-risk scenarios. While in ordinary and simple scenarios, this does not significantly interfere with the baseline model's normal generation process. Ultimately, this method ensures trajectories' safety while minimizing the loss of forward speed.

simple scenarios. (e.g., straight highways with low traffic density, as shown in (a)), our method introduces only minimal modifications to the baseline trajectory. Because RISE activates safer energy guidance only when the risk assessor detects potential hazards, the generated trajectory remains nearly identical to the baseline's high-performance path. Consequently, driving speed and comfort are preserved, demonstrating that our safety-oriented design does not penalize nominal driving performance.

These examples highlight that RISE does not merely smooth trajectories. Instead, it performs risk-aware correction that systematically balances safety constraints and task efficiency, aligning with the observed improvements in both DAC and EP sub-metrics.

E.1 Failure Analysis and Limitations

Despite the overall effectiveness of RISE, we observe two typical failure modes in our experiments.

Irregular trajectory generation. In some scenarios, as shown in Fig. 5 (a), the multi-energy guidance produces distorted or physically implausible trajectories. We attribute this to the complex interaction among the hyperparameters of different energy terms. Our current approach employs a uniform hyperparameter setting across all scenes that need guidance, which may lead to over-correction of certain energy components in specific situations. This over-regularization disrupts the natural balance between safety constraints and trajectory smoothness, resulting in deformed paths.

Uncovered high-risk scenarios. As shown in Fig. 5 (b), our method fails to mitigate certain hazards that are not explicitly encoded in the rule-based guidance. Due to the inherent incompleteness of handcrafted rules and the out-of-distribution (OOD) generalization issues of the learning-based expert model, some dangerous situations remain unaddressed. Moreover, because we do not perform any additional training on the baseline model, the framework lacks the capacity to adapt to these novel risky patterns. Consequently, RISE cannot effectively avoid such hazards.



(a) Irregular Trajectory Generation (b) Uncovered High-Risk Scenarios

Figure 5: Qualitative visualization of failure cases on NAVSIM benchmark. The orange dashed line represents the original trajectory predicted by the ReCogDrive baseline, while the red solid line represents the final trajectory after applying our energy-based mixed guidance. In the (a) irregular trajectory generation examples, overly aggressive hyperparameter settings cause over-correction from multiple energy terms, leading to a distorted and physically implausible trajectory. In the (b) uncovered high-risk scenarios examples, although the scene involves clear risks (*e.g.*, unprotected turns), those risks are not covered by any handcrafted rule and fall outside the expert model's training distribution; as a result, RISE fails to mitigate the hazard, and the final trajectory remains unsafe. These failure modes highlight the sensitivity of hyperparameter balancing and the limitation of training-free guidance when encountering novel risk patterns.