CoReDrive: Actor-Conditioned Planning with Counterfactual Response Learning

Anonymous authors
Paper under double-blind review

Abstract

Autonomous driving requires anticipating the behavior of other road users and adapting ego plans accordingly, but driving logs provide supervision only for observed interactions. We present CoReDrive, which couples actor perception, multimodal motion prediction, motion-derived intent, and ego planning through a structured actor interface. To teach responses beyond the logged interaction, we edit an actor's predicted futures while holding the observation and current scene geometry fixed. A geometric teacher constructs sets of admissible ego responses instead of treating a synthetic trajectory as a unique expert. Response-set targets supervise candidate retention and ranking, while paired factual and edited worlds supervise an interaction-score correction used during velocity admission and final selection. On NAVSIM, CoReDrive achieves 90.8% and 91.4% corrected extended Predictive Driver Model Score with ResNet-34 and ViT-L, respectively. A matched ResNet-34 comparison shows a gain from counterfactual response supervision. Controlled actor-information and future-edit visualizations examine how the planner uses its interaction interface in selected scenes.

1 Introduction

Interactive driving requires an ego vehicle to adapt its actions to how other road users may behave. At an intersection, an approaching vehicle may yield or continue across the ego route; nearly identical current observations can therefore demand substantially different responses. Driving logs record only the joint future that actually occurred, leaving appropriate responses to alternative actor behaviors unobserved. This creates a critical supervision gap: learning a good action in a recorded interaction does not directly teach how that action should change when the interaction changes.

Recent advances provide increasingly capable foundations for addressing this gap. Structured driving systems connect perception, prediction, and planning through explicit scene representations (Hu et al., 2023; Jiang et al., 2023), while candidate-based planners improve action coverage and selection through trajectory generation, metric distillation, and factorized motion vocabularies (Li et al., 2024; Liao et al., 2025; Sun et al., 2026). However, evaluating more ego candidates against recorded actor futures does not itself provide supervision for responses to changed actor behavior. Accurate prediction makes actor behavior available to a planner, but does not ensure that its decisions respond appropriately to changes in that behavior. Contingency planners explicitly reason over alternative futures (Cui et al., 2021; Mustafa et al., 2024). We address the corresponding learning problem: converting alternative actor futures into direct supervision for which ego responses should survive candidate pruning and ultimately be selected.

We introduce CoReDrive, a framework for actor-conditioned planning with counterfactual response learning. Its central insight is to make an actor representation serve both as a planning condition and as an editable source of response supervision. Each actor is represented through detected geometry, multimodal futures, and motion-derived intent, such as turning or braking. The interface exposes both future geometry and motion semantics, allowing the planner to condition its decisions on actor behavior at complementary levels of abstraction. Building on SparseDriveV2's factorized path–velocity representation (Sun et al., 2026), ego planning and vehicle forecasting share motion dictionaries, and ego candidates attend to the actor interface during selection and final scoring. Training first jointly learns actor modeling and factual planning, then adds counterfactual response supervision while retaining factual objectives.

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.

During response learning, we edit one actor's predicted futures and intent while keeping observations, current geometry, and other actors fixed. Edits are filtered for actor reliability, kinematic plausibility, and conflict relevance before response targets are constructed. These edits construct alternative interaction conditions without rendering new sensor observations or requiring additional expert ego responses. Because an edited interaction can admit multiple reasonable behaviors, we formulate supervision as response-set learning. A geometric teacher identifies feasible candidates satisfying interaction-specific temporal constraints and collision checks throughout the planning horizon, including recovery. For a crossing actor, valid responses remain upstream until the conflict clears and can subsequently resume progress. Quality weights account for geometric risk and progress, allowing multiple admissible responses to contribute to supervision.

Learning these responses requires addressing both their availability and their selection: a final scoring objective cannot recover a suitable response that earlier pruning has removed. CoReDrive therefore supervises candidate selectors using response-set targets and trains final selection with a response-set ranking objective. We additionally learn a shared interaction scorer from paired geometric-risk targets for the same ego action under factual and edited actor futures. This pairing directly supervises how action preferences should change when another actor's behavior changes, while an invariance objective constrains score changes for unrelated edits. Applied before velocity pruning and at final selection, the scorer carries learned interaction consequences through both decisions.

On NAVSIM navtest, CoReDrive achieves 90.8% and 91.4% corrected extended Predictive Driver Model Score (EPDMS) with ResNet-34 and ViT-L, respectively. The ResNet-34 model improves over the paper-reported SparseDriveV2 baseline from 90.1% to 90.8%, with gains of 0.7 percentage points in ego progress and 0.6 in extended comfort. Under matched initialization and training budgets, counterfactual response supervision further improves actor-conditioned planning from 90.7% to 90.8%, while increasing both no-at-fault collision and time-to-collision compliance by 0.2 percentage points. Expanded response support is used only during training; inference retains 200 final trajectory candidates.

Our contributions are threefold:

- We introduce CoReDrive, using an editable actor interface to connect actor-conditioned planning with counterfactual response supervision.
- We formulate response-set learning, constructing temporally and geometrically admissible targets for edited interactions without requiring a unique synthetic expert trajectory.
- We develop joint supervision of response retention and selection, combining selector targets, response-set ranking, and paired interaction-risk learning to guide decisions throughout the candidate pipeline.

2 Related Work

Scene representations for planning. UniAD and VAD connect perception and motion modeling to planning through task-specific or vectorized scene representations (Hu et al., 2023; Jiang et al., 2023). These representations expose other road users to the planner; the remaining supervision question is how to train a response when their predicted behavior changes. CoReDrive uses actor geometry, multimodal futures, and motion-derived intent as an explicit planning condition, and edits this condition to create response supervision beyond the logged interaction.

Candidate generation and scoring. Hydra-MDP and Hydra-MDP++ distill planning objectives from demonstrations and rule-based teachers, while DiffusionDrive generates multimodal proposals, and DriveSuprim selects trajectories through a coarse-to-fine procedure (Li et al., 2024; 2025; Liao et al., 2025; Yao et al., 2026). ReCogDrive and DiffusionDriveV2 combine diffusion planning with reinforcement learning (Li et al., 2026; Zou et al., 2025). SparseDriveV2's NAVSIM planner factorizes motion into path and velocity candidates and applies coarse-to-fine scoring (Sun et al., 2026). We build on this candidate representation. Our focus is on the targets used to select actor-dependent actions: edited-world response sets supervise candidate retention and ranking, while paired geometric targets constrain an interaction-score correction.

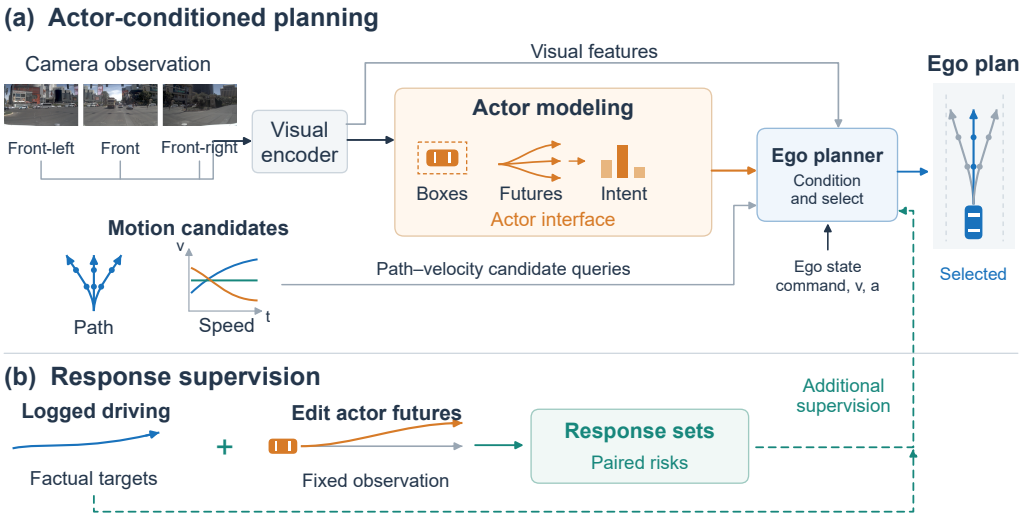


Figure 1: **CoReDrive architecture and supervision.** Visual features support both actor modeling and ego planning. Detected actors, multimodal futures, and derived intent form a structured planning condition. During response learning, factual supervision is augmented with targets constructed from edited actor futures; dashed arrows denote training supervision. The path–speed candidate representation follows SparseDriveV2 (Sun et al., 2026). Camera thumbnails show synchronized front views from one NAVSIM observation; v and a denote ego velocity and acceleration.

Counterfactual interaction supervision. AdvSim perturbs actor behavior with sensor simulation, SAFE-SIM generates controllable safety-critical scenarios, and OmniDrive explores counterfactual reasoning in vision–language driving (Wang et al., 2021; Chang et al., 2024; Wang et al., 2025). CoReDrive intervenes on predicted actor futures inside the planning interface, keeping the current sensor observation fixed. This permits paired supervision of the same ego action under two actor futures without rendering a new sensor sequence. Without an expert ego trajectory for the edited condition, the teacher provides a set of rule-admissible responses. These targets supervise both candidate retention and final ranking; paired geometric risk targets add a consistent score correction across the two conditions.

3 Method

CoReDrive learns a planner conditioned on predicted actors, then adds response supervision for edited interactions. Its three components are the actor interface, a geometric response teacher, and objectives that connect the teacher to candidate admission and selection (Figure 1).

3.1 Actor-conditioned planning

Observations x comprise three front-camera views and ego state: driving command, velocity, and acceleration. A shared visual backbone supports actor modeling and planning. A position embedding transformation (PETR)-style decoder (Liu et al., 2022) predicts actor boxes, classes, and multimodal futures. The interface for actor i is

$$z_i = (b_i, \pi_i, \{w_{im}, \hat{\tau}_{im}, \mu_{im}\}_{m=1}^{M_i}, \iota_i), \tag{1}$$

where b_i is a three-dimensional box, π_i the class distribution, w_{im} the probability of future mode m, $\hat{\tau}_{im}$ its trajectory, μ_{im} its validity mask, and ι_i a motion-intent summary. Actor poses transform local futures into the ego frame.

Motion and intent. We inherit SparseDriveV2's fixed dictionaries of 1,024 paths and 256 velocity profiles for ego and vehicle motion (Sun et al., 2026); forecasting heads remain task-specific.

Pedestrians and cyclists use family-specific dictionaries. Coarse path, speed, and composed-motion categories summarize the corresponding code predictions. For example, the path-intent probability is

$$\iota_i^p(c) = \sum_{u:g_p(u)=c} P_i^p(u), \tag{2}$$

where g_p assigns a path code to a category such as straight, turn, or lane change. Speed intent is aggregated from velocity predictions; composed intent is aggregated from valid path–velocity pairs. These distributions describe the predicted motion at a coarser scale, so geometric and semantic conditions can be edited together. Appendix A.1 gives the categories and mappings.

Conditioning the candidates. Each retained actor–mode pair becomes a token: a fixed projection of its box, class, trajectory, validity, intent, and mode probability is added to the actor query feature and normalized. Ego candidate queries attend to these tokens, with actor presence and mode probability as attention priors. Conditioning enters both path and velocity selection layers and the final trajectory representation; visual features also enter the planner directly. The two selection layers retain $1{,}024 \to 128 \to 20$ paths and $256 \to 64 \to 10$ velocities, forming 200 final trajectories. Eight metric heads produce base scores s_k. A bounded factual reranking head gives s_k^{fact}; an interaction correction then adjusts velocity admission and final selection (Section 3.3). Both matched training variants share these scoring modules and factual objectives; response learning adds edited-world targets. Detection, forecasting, and planning are learned with joint multitask supervision before response learning.

3.2 Edited worlds and response sets

An intervention $\xi = (i, r, h, \kappa)$ selects actor i, edit rule r, severity h, and an ego-response requirement κ. The rule changes all retained future modes of this actor and updates its intent. Images, current boxes, classes, mode weights, and other actors remain fixed. The library covers braking, crossing and cut-in, turn-across, roadside pull-out, blocked lane change, and cancelled yielding. Reliability, kinematic, and conflict checks accept up to two edited worlds z^ξ per scene (Appendix A.2).

The logged ego trajectory need not remain appropriate in z^ξ. Instead, a frozen factual policy supplies a reference route $\bar{\tau} = f_{\bar{\theta}}(x, z)$, and a geometric teacher evaluates candidate responses relative to that route. Candidate τ_k is admissible when

$$a_k = \mathbf{1}[\mathcal{F}(\tau_k)]\mathbf{1}[\mathcal{R}_\kappa(\tau_k; \bar{\tau}, z^\xi)]\mathbf{1}[\neg\mathcal{C}(\tau_k, z^\xi)]\mathbf{1}[\mathcal{V}(z^\xi)], \tag{3}$$

where $\mathcal{F}$ checks feasible motion, $\mathcal{R}_\kappa$ the required response, $\mathcal{C}$ predicted collision, and $\mathcal{V}$ world validity. For a crossing actor, the ego must stay upstream until the last predicted occupancy plus a clearance buffer; it can then recover within a reference-route corridor. Lead braking instead uses a moving rear boundary. Collision checks cover the full horizon, including recovery. Several candidates can satisfy these constraints, forming a response set without a unique synthetic expert (Figure 2). The rule specifies the required behavior, while the set allows alternative trajectories that realize it. Temporal checks distinguish occupying the same location at different times from a simultaneous conflict, and include the subsequent recovery motion.

3.3 Learning counterfactual responses

The teacher expands the 20 selected paths across all 256 velocities, forming a training support $\mathcal{W}_\theta$ of 5,120 candidates. This support expands velocity choices on the retained paths; it does not recover paths removed by earlier pruning. Labels are recomputed on the current support and detached from optimization. Each admissible candidate receives quality $q_k = a_k Q(\rho_k, \eta_k)$, favoring lower geometric risk ρ_k and greater reference-relative progress η_k; the precise bounded weight Q is given in Appendix A.3.

Preserving and ranking responses. For a selector layer with valid input identities U and V, let

$$\widetilde{t}_u^p = \mathbf{1}[u \in U] \sum_{k \in \mathcal{W}_\theta : u_k = u} q_k, \qquad t_u^p = \frac{\widetilde{t}_u^p}{\sum_{j \in U} \widetilde{t}_j^p}. \tag{4}$$

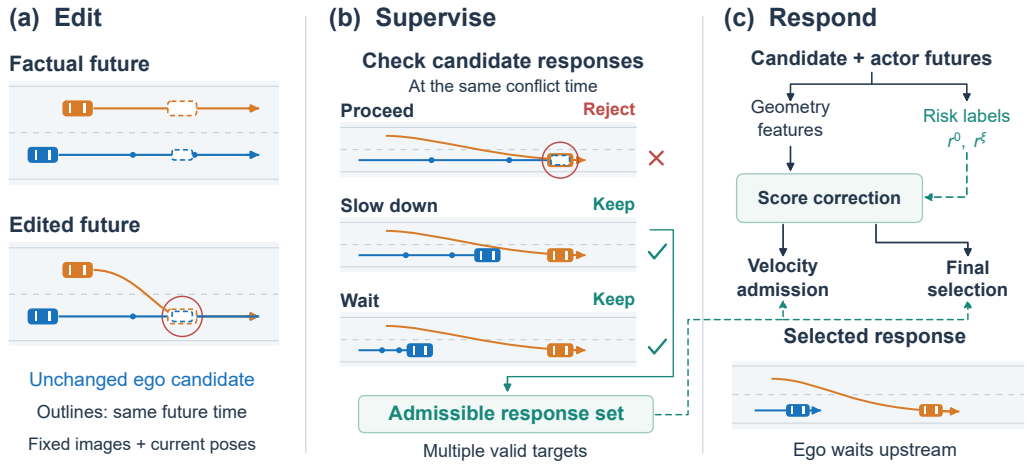


Figure 2: **From an actor edit to response supervision.** (a) Changing the orange actor's future can make an unchanged ego candidate unsafe; outlines mark occupancy at the same future time. (b) Candidate responses are compared at a common conflict time: proceeding overlaps the edited actor, while slowing or waiting keeps the ego upstream. Full-horizon temporal checks retain an admissible response set rather than a single target. (c) Candidate–actor features feed the score correction, while paired geometric risk labels supervise it. The correction acts before velocity pruning and at final selection; response-set targets separately train candidate retention and ranking. Dashed links denote supervision and solid arrows denote computation. Geometry and curves are schematic.

The velocity target t^v is defined analogously. Their cross-entropies with pre-pruning logits form $\mathcal{L}_{\text{sel}}$, teaching the selectors to retain compatible actions within the labeled support. On the final support $\mathcal{A}_\theta$, response-set ranking is

$$\mathcal{L}_{\text{rank}} = -\log \sum_{k:a_k=1} p_k \widehat{q}_k, \quad p_k = \frac{\max(s_k, \epsilon)}{\sum_{j\in\mathcal{A}_\theta} \max(s_j, \epsilon)}, \quad \widehat{q}_k = \frac{q_k}{\max_{j:a_j=1} q_j}. \tag{5}$$

All valid final candidates compete in the denominator. This loss favors probability mass on high-quality admissible candidates without specifying one expert trajectory; it does not require a diverse final distribution. A selector layer contributes only when both target marginals have positive mass; ranking is omitted without a positive final candidate. Selector losses train the discrete selection logits directly; the ranking loss acts through retained candidate representations without differentiating through top-k identities.

Paired interaction-score correction. A bounded score correction $u_\phi(\tau, z) = c \tanh f_\phi(\tau, z)$ consumes detached candidate–actor geometry. Let r^0, r^ξ be conservative clearance-based risk targets for the *same* candidate in factual and edited worlds, and $\widetilde{u} = u_\phi/c$. Its paired objectives are

$$\begin{aligned} \mathcal{L}_{\text{abs}} &= \tfrac{1}{2}[(\widetilde{u}^0 + r^0)^2 + (\widetilde{u}^\xi + r^\xi)^2], \\ \mathcal{L}_{\Delta} &= [(\widetilde{u}^\xi - \widetilde{u}^0) + (r^\xi - r^0)]^2, \\ \mathcal{L}_{\text{inv}} &= \mathbf{1}[\text{unrelated edit}](\widetilde{u}^\xi - \widetilde{u}^0)^2. \end{aligned} \tag{6}$$

The difference term aligns the score change with the negative geometric risk change; it couples the two absolute regressions using the same labels. Invariance keeps the correction stable for a distant, unrelated edit. Both worlds use common valid actor/time support; unknown targets are excluded. These local geometric targets complement the response-set quality proxy (Appendix A.3). Together with factual score supervision, they train a planning correction rather than a calibrated collision probability.

The scorer is reused at two decisions. Before the second velocity pruning, average its output over the selected paths compatible with velocity v, denoted $\mathcal{P}_v$. Let ℓ_v be its pre-pruning logit and $\tau(p, v)$

compose a path and velocity profile. At final selection, score each candidate individually:

$$\widetilde{\ell}_v = \ell_v + \frac{1}{|\mathcal{P}_v|} \sum_{p \in \mathcal{P}_v} u_\phi(\tau(p, v), z), \qquad k^* = \arg\max_{k \in \mathcal{A}_\theta} [s_k^{\text{fact}} + 14 u_\phi(\tau_k, z)]. \tag{7}$$

Here s_k^{fact} is the base score with bounded factual score correction, and 14 is the scale of its metric sum. Thus an interaction preference can affect which velocities remain available as well as which trajectory is chosen.

Training objective. Factual planning supervision is retained during response learning:

$$\mathcal{L} = \mathcal{L}_{\text{fact}}^{\text{all}} + \lambda_{\text{cf}} \mathbb{E}_\xi \left[\mathcal{L}_{\text{sel}} + \mathcal{L}_{\text{rank}} + \mathcal{L}_{\text{abs}} + \mathcal{L}_\Delta + \lambda_{\text{inv}} \mathcal{L}_{\text{inv}} \right]. \tag{8}$$

$\mathcal{L}_{\text{fact}}^{\text{all}}$ includes logged planning targets, reference-policy retention, and factual interaction-score supervision. Visual and actor encoders remain frozen during this phase. Counterfactual gradients update the conditioning, selector, and interaction-scoring parameters; metric and imitation heads receive factual gradients only. The score computation remains differentiable with respect to candidate representations. Appendix A.3 specifies this gradient routing, the regularizers, and all coefficients.

4 Experiments

We evaluate three aspects of the design: factual planning performance, the incremental effect of response supervision under matched training, and conditional behavior when actor information is varied.

4.1 Experimental setup

We use NAVSIM's prescribed 103,288 navtrain and 12,146 navtest scenes (Dauner et al., 2024). Our primary metric is the corrected extended Predictive Driver Model Score (EPDMS). Its nine components are no-at-fault collision (NAC), drivable-area compliance (DAC), driving-direction compliance (DDC), traffic-light compliance (TLC), ego progress (EP), time-to-collision compliance (TTC), lane keeping (LK), history comfort (HC), and two-frame extended comfort (EC) (NAVSIM Contributors, 2026). Scores are percentages, rounded to one decimal place. We evaluate extended scoring on navtest, separately from the navhard pseudo-simulation protocol (Cao et al., 2025).

EPDMS extends the Predictive Driver Model Score (PDMS) with additional traffic-rule and lane-keeping checks and richer comfort measures. Its multiplicative compliance terms gate a weighted combination of progress, collision-time margin, and comfort-related components (NAVSIM Contributors, 2026). EPDMS is therefore not the arithmetic mean of its components; we report all nine to distinguish progress, compliance, and comfort.

Implementation. The planner uses three front-camera views, ego state, and up to 32 predicted actors, and retains 200 path–velocity candidates at inference. Joint actor and planning training precedes response learning. The factual and counterfactual variants share initialization, data order, and adaptation budget; the latter adds the proposed response targets. Architecture, supervision rules, and metric aggregation are specified in Appendix A.

4.2 Planning performance

Table 1 compares planning performance while retaining all nine components. CoReDrive reaches 90.8% corrected EPDMS with ResNet-34, exceeding the reported SparseDriveV2 result of 90.1%. Relative to the reported results, CoReDrive improves ego progress from 91.1% to 91.8%, lane keeping from 96.9% to 97.5%, and extended comfort from 78.4% to 79.0%. These results show improved overall planning quality alongside gains in progress, lane keeping, and comfort. With a ViT-L backbone, CoReDrive reaches 91.4% corrected EPDMS, with 98.9% NAC and 98.7% DAC. It also achieves 91.5% PDMS without fine-tuning with a PDMS scorer.

The table distinguishes original and corrected EPDMS because their implementations differ; cross-method score comparisons use the corrected column. Published component results follow their

Table 1: Planning performance on NAVSIM navtest (%, ↑). † denotes a ViT-L backbone (Vision Transformer, large); unmarked methods use ResNet-34. Original and corrected EPDMS use different scoring implementations. A dash denotes an unavailable result.

Method	Components									EPDMS	
	NAC	DAC	DDC	TLC	EP	TTC	LK	HC	EC	Orig.	Corr.
Hydra-MDP++	97.2	97.5	99.4	99.6	83.1	96.5	94.4	98.2	70.9	81.4	—
Hydra-MDP++†	98.5	98.5	99.5	99.7	87.4	97.9	95.8	98.2	75.7	85.6	—
DriveSuprim	97.5	96.5	99.4	99.6	88.4	96.6	95.5	98.3	77.0	83.1	—
DriveSuprim†	98.4	98.6	99.6	99.8	90.5	97.8	97.0	98.3	78.6	87.1	—
ARTEMIS	98.3	95.1	98.6	99.8	81.5	97.4	96.5	98.3	98.3	83.1	—
DiffusionDriveV2	97.7	96.6	99.2	99.8	88.9	97.2	96.0	97.8	91.0	85.5	87.5
SparseDriveV2	98.1	98.1	99.6	99.8	91.1	97.3	96.9	98.2	78.4	86.7	90.1
CoReDrive	98.3	98.4	99.7	99.8	91.8	97.6	97.5	98.3	79.0	—	**90.8**
CoReDrive†	98.9	98.7	99.7	99.8	90.7	98.3	97.1	98.3	79.0	—	**91.4**

ARTEMIS and DiffusionDriveV2 use LiDAR (light detection and ranging). CoReDrive is evaluated with the corrected scorer. Component definitions are given in Section 4.

Table 2: System progression on navtest (%, ↑). All rows use ResNet-34; EPDMS uses the corrected definition. The last two configurations share initialization and training budget.

Configuration	NAC	DAC	DDC	TLC	EP	TTC	LK	HC	EC	EPDMS
Baseline	98.1	98.1	99.6	99.8	91.1	97.3	96.9	98.2	78.4	90.1
Baseline + Actor	98.1	98.3	99.7	99.8	91.9	97.4	97.5	98.3	79.2	90.7
Baseline + Actor + Response	98.3	98.4	99.7	99.8	91.8	97.6	97.5	98.3	79.0	**90.8**

Baseline: reported SparseDriveV2 results (Sun et al., 2026). Actor: joint actor modeling and conditioning. Response: counterfactual response learning.

respective source protocols, detailed in Appendix A.4. ViT-L (Vision Transformer, large) backbones are marked by a superscript alongside ResNet-34 methods; ARTEMIS and DiffusionDriveV2 also use LiDAR.

4.3 From actor conditioning to response learning

Table 2 presents the progression from the reported SparseDriveV2 baseline to factual CoReDrive and the complete response model. The factual system (Baseline + Actor), which couples actor modeling to planning, reaches 90.7% corrected EPDMS. Adding response supervision increases the score to 90.8% under matched training conditions. NAC rises from 98.1% to 98.3%, DAC from 98.3% to 98.4%, and TTC from 97.4% to 97.6%. The improvement is obtained with the same 200-candidate inference budget: the expanded response support is used only to construct training targets.

The two CoReDrive configurations share their architecture and factual objectives, so this comparison measures the incremental effect of the response-supervision package on factual planning. The unchanged final candidate count shows that the wider teacher support is a training-time choice, rather than an increase in the number of final proposals.

4.4 Actor information and edited responses

We examine how actor information affects a plan and how the plan changes when an actor's predicted future is edited. These examples inspect the interface's conditional behavior. Each comparison fixes model weights and changes only the specified actor information.

Using actor information. With observations and weights fixed, masking actor information causes a collision with the leading vehicle at 3.5 s in native NAVSIM simulation; retaining it avoids contact (Figure 3(a)). This contrast shows that actor information affects planning beyond the visual input.

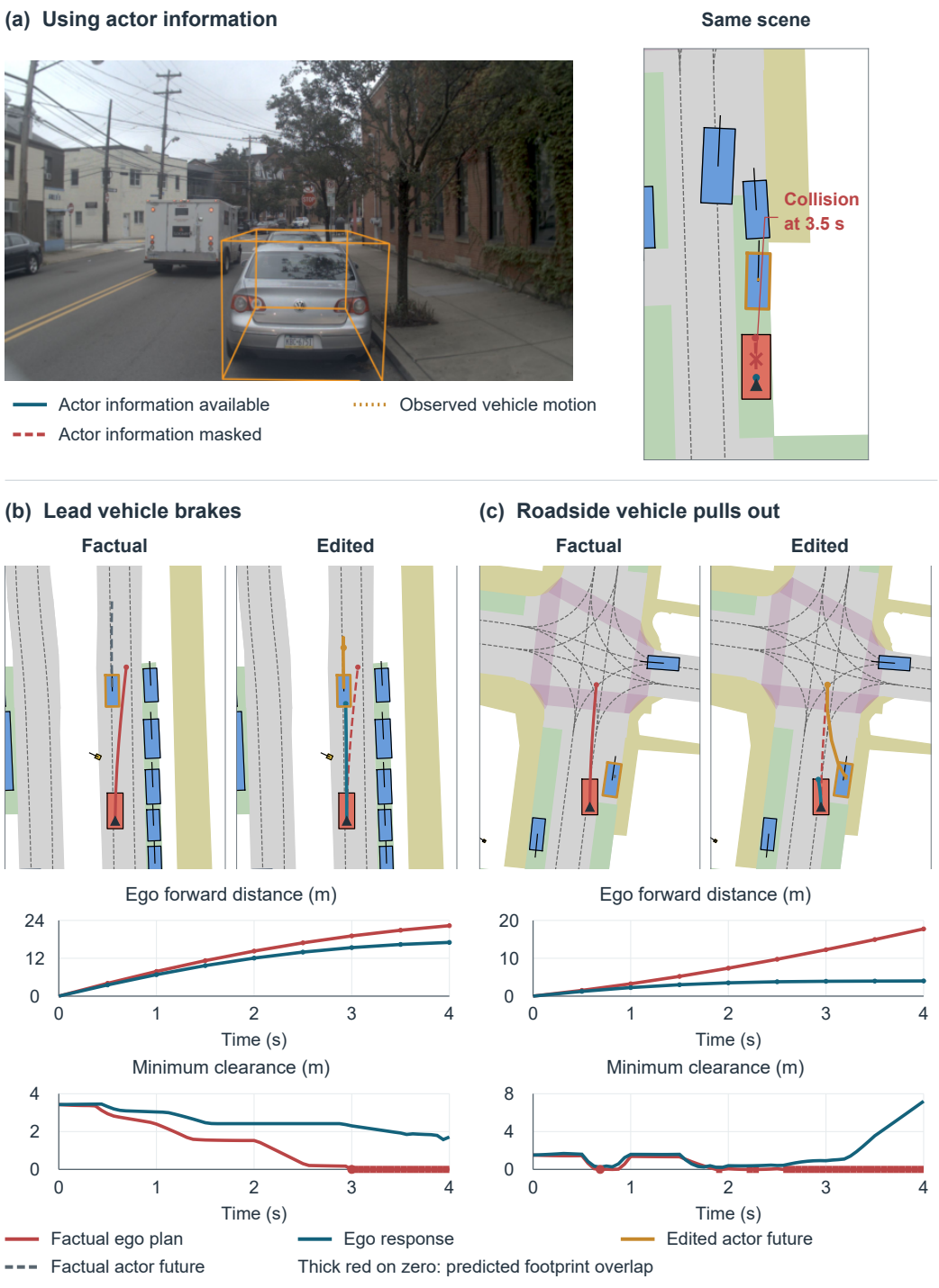


Figure 3: **Using actor information and responding to edited futures.** (a) With fixed observations and ViT-L weights, masking actor information produces an at-fault collision at 3.5 s in native NAVSIM simulation; retaining it avoids contact. Orange identifies the same leading vehicle; the cross marks simulated contact. (b,c) At fixed observations and current geometry, lead braking and roadside pull-out edits elicit different ego responses. Red denotes the factual ego plan, blue the response, and orange the edited actor future. Lower plots show forward distance over time. Clearance plots measure minimum footprint separation in the edited worlds; thick red segments at zero mark overlap. Maps and current boxes provide context. Panels (b,c) are open-loop response diagnostics.

Responding to changed futures. Figure 3(b,c) holds images and current geometry fixed and changes one actor's future and derived intent. Lead braking reduces planned forward distance from 22.3 m to 17.0 m at four seconds. A roadside pull-out elicits stronger braking: planned forward distance at four seconds falls from 17.8 m to 4.0 m. In both edited worlds, the factual ego plan overlaps a predicted actor footprint, whereas the response removes that overlap over the evaluated horizon. The temporal plots make the change visible even when two plans follow nearly the same spatial path. These examples illustrate conditional response behavior under predicted geometry, rather than a closed-loop success rate.

5 Discussion

The central supervision problem is that changing an actor future removes the logged ego trajectory's status as an expert response. CoReDrive replaces that missing target with a set of candidates satisfying explicit temporal and geometric requirements. Projecting the set onto selector targets and ranking retained responses connects this supervision to both decisions in a coarse-to-fine planner. Paired geometric labels complement the set by constraining an interaction-score correction across the two conditions.

The resulting supervision is conditional on predicted actors, the edit rules, and the candidate support. Admissibility covers the specified response requirements over a finite horizon; it does not simulate how other road users react to the ego. The framework therefore provides a way to train responses to alternative predicted futures, with longer, reciprocal interactions remaining a direction for closed-loop evaluation.

6 Conclusion

CoReDrive connects actor-conditioned planning to counterfactual response learning. An editable interface of actor futures and derived intent provides alternative interaction conditions; response sets supply supervision without requiring a unique synthetic expert. These targets supervise candidate retention and ranking, complemented by a paired interaction-score correction at velocity admission and final selection. The resulting planner achieves 90.8% and 91.4% corrected EPDMS on NAVSIM navtest with ResNet-34 and ViT-L, respectively, with a matched ResNet-34 comparison supporting the incremental benefit of counterfactual response supervision.

AI use statement

Generative AI assisted code implementation, experimental planning, artifact inspection, literature search, and manuscript preparation. Conceptual schematics were developed with an OpenAI image tool and refined through local drawing and typesetting. Reported numerical results retain their cited evaluation sources; the authors are responsible for the methods, verification, citations, and presentation.

Ethics statement

This study uses existing driving datasets and model-input interventions, without public-road testing. Rule-based admissibility over predicted actors is not a real-world safety certification. Responsible deployment requires broader validation of perception uncertainty and interactive behavior beyond the evaluation setting studied here.

Reproducibility statement

Appendix A specifies the architecture, intervention rules, response predicates, and evaluation protocol. All reported CoReDrive EPDMS scores use the same evaluation protocol. The response teacher is specified through its edit rules, admissibility predicates, and targets.

REFERENCES

Wei Cao, Marcel Hallgarten, Tianyu Li, Daniel Dauner, Xunjiang Gu, Caojun Wang, Yakov Miron, Marco Aiello, Hongyang Li, Igor Gilitschenski, Boris Ivanovic, Marco Pavone, Andreas Geiger, and Kashyap Chitta. Pseudo-simulation for autonomous driving. In *Conference on Robot Learning*, 2025. URL https://arxiv.org/abs/2506.04218.

Wei-Jer Chang, Francesco Pittaluga, Masayoshi Tomizuka, Wei Zhan, and Manmohan Chandraker. SAFE-SIM: Safety-critical closed-loop traffic simulation with diffusion-controllable adversaries. In *European Conference on Computer Vision*, pp. 242–258, 2024. doi: 10.1007/978-3-031-72664-4_14. URL https://arxiv.org/abs/2401.00391.

Alexander Cui, Sergio Casas, Abbas Sadat, Renjie Liao, and Raquel Urtasun. LookOut: Diverse multi-future prediction and planning for self-driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021. URL https://arxiv.org/abs/2101.06547.

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. In *Advances in Neural Information Processing Systems*, 2024. URL https://arxiv.org/abs/2406.15349.

Renju Feng, Ning Xi, Duanfeng Chu, Rukang Wang, Zejian Deng, Anzheng Wang, Liping Lu, Jinxiang Wang, and Yanjun Huang. ARTEMIS: Autoregressive end-to-end trajectory planning with mixture of experts for autonomous driving. *IEEE Robotics and Automation Letters*, 11(1): 226–233, 2026. doi: 10.1109/LRA.2025.3632616. URL https://doi.org/10.1109/LRA.2025.3632616.

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*, pp. 17853–17862, 2023. URL https://arxiv.org/abs/2212.10156.

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. URL https://arxiv.org/abs/2303.12077.

Kailin Li, Zhenxin Li, Shiyi Lan, Yuan Xie, Zhizhong Zhang, Jiayi Liu, Zuxuan Wu, Zhiding Yu, and Jose M. Alvarez. Hydra-MDP++: Advancing end-to-end driving via expert-guided hydra-distillation, 2025. URL https://arxiv.org/abs/2503.12820.

Yongkang Li, Kaixin Xiong, Xiangyu Guo, Fang Li, Sixu Yan, Gangwei Xu, Lijun Zhou, Long Chen, Haiyang Sun, Bing Wang, Kun Ma, Guang Chen, Hangjun Ye, Wenyu Liu, and Xinggang Wang. ReCogDrive: A reinforced cognitive framework for end-to-end autonomous driving. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/ff7bf6014f7826da531aa50f4538ee19-Abstract-Conference.html.

Zhenxin Li, Kailin Li, Shihao Wang, Shiyi Lan, Zhiding Yu, Yishen Ji, Zhiqi Li, Ziyue Zhu, Jan Kautz, Zuxuan Wu, Yu-Gang Jiang, and Jose M. Alvarez. Hydra-MDP: End-to-end multimodal planning with multi-target hydra-distillation, 2024. URL https://arxiv.org/abs/2406.06978.

Bencheng Liao, Shaoyu Chen, Haoran Yin, Bo Jiang, Cheng Wang, Sixu Yan, Xinbang Zhang, Xiangyu Li, Ying Zhang, Qian Zhang, and Xinggang Wang. DiffusionDrive: Truncated diffusion model for end-to-end autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12037–12047, 2025. URL https://arxiv.org/abs/2411.15139.

Yingfei Liu, Tiancai Wang, Xiangyu Zhang, and Jian Sun. PETR: Position embedding transformation for multi-view 3d object detection. In *European Conference on Computer Vision*, 2022. URL https://arxiv.org/abs/2203.05625.

Khaled A. Mustafa, Daniel Jarne Ornia, Jens Kober, and Javier Alonso-Mora. RACP: Risk-aware contingency planning with multi-modal predictions. *IEEE Transactions on Intelligent Vehicles*, 2024. doi: 10.1109/TIV.2024.3411530. URL https://arxiv.org/abs/2402.17387.

NAVSIM Contributors. NAVSIM: Metrics, 2026. URL https://github.com/autonomousvision/navsim/blob/main/docs/metrics.md. Official metric documentation, accessed September 14, 2026.

Wenchao Sun, Xuewu Lin, Keyu Chen, Zixiang Pei, Xiang Li, Yining Shi, and Sifa Zheng. SparseDriveV2: Scoring is all you need for end-to-end autonomous driving. In *European Conference on Computer Vision*, 2026. URL https://arxiv.org/abs/2603.29163.

Jingkang Wang, Ava Pun, James Tu, Sivabalan Manivasagam, Abbas Sadat, Sergio Casas, Mengye Ren, and Raquel Urtasun. AdvSim: Generating safety-critical scenarios for self-driving vehicles. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9909–9918, 2021. URL https://arxiv.org/abs/2101.06549.

Shihao Wang, Zhiding Yu, Xiaohui Jiang, Shiyi Lan, Min Shi, Nadine Chang, Jan Kautz, Ying Li, and Jose M. Alvarez. OmniDrive: A holistic vision-language dataset for autonomous driving with counterfactual reasoning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 22442–22452, 2025. URL https://arxiv.org/abs/2504.04348.

Wenhao Yao, Zhenxin Li, Shiyi Lan, Zi Wang, Xinglong Sun, Jose M. Alvarez, and Zuxuan Wu. DriveSuprim: Towards precise trajectory selection for end-to-end planning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 11910–11918, 2026. doi: 10.1609/aaai.v40i14.38178. URL https://ojs.aaai.org/index.php/AAAI/article/view/38178.

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. URL https://arxiv.org/abs/2512.07745.

A Implementation and Supervision Details

A.1 Model configuration

Table A1 gives the shared configuration of the factual and counterfactual variants; only the latter uses response targets.

Table A1: Model configuration.

Input	Latest synchronized front-left, front, front-right views; driving command, ego velocity and acceleration
Hidden representation	Width 256; eight attention heads
Actor decoder	144 queries; four geometry decoder layers
Planning interface	Up to 32 actors, five modes per actor; eight future samples over four seconds
Candidate selectors	Two layers; paths $1024 \to 128 \to 20$; velocities $256 \to 64 \to 10$
Frozen during response learning	Visual and actor encoders, motion dictionaries, reference policy

Training. Joint actor and planning training precedes response learning. We use Adam with cosine learning-rate decay. The matched ResNet-34 comparison shares initialization, data order, and training budget; only the response-supervision terms differ. During response learning, the encoders and motion dictionaries remain fixed.

Actor tokens. For actor i and retained mode m, concatenate three foreground class probabilities, a seven-dimensional box, an 8×3 trajectory, eight validity indicators, the $7+5+8$ intent probabilities, and the mode weight into $a_{im} \in \mathbb{R}^{63}$. Its token is $h_{im} = \mathrm{LN}(q_i + a_{im}W_0)$, where q_i is the actor query and $W_0[r, j] = \cos(0.017(r+1)(j+1))/\sqrt{63}$ is a fixed projection. Invalid tokens are masked. Candidate-to-actor attention adds $\log \max(p_i, 10^{-6}) + \log \max(w_{im}, 10^{-6})$ to its logits, where p_i is foreground presence. Attention also includes a relation bias; the final trajectory layer derives the relation from candidate–actor geometry. The two path and two velocity selector layers use the default relation. The attended feature is added residually to the candidate representation.

Derived intent. Path and speed distributions are aggregated from their respective code logits; composed motion uses the valid pair distribution:

$$\iota_i^p(c) = \sum_{u:g_p(u)=c} \mathrm{softmax}(\ell_i^p)_u, \qquad \iota_i^v(c) = \sum_{v:g_v(v)=c} \mathrm{softmax}(\ell_i^v)_v,$$
$$\iota_i^m(c) = \sum_{(u,v)\in\mathcal{S}_i} P_i(u,v)\mathbf{1}[g_m(g_p(u), g_v(v)) = c]. \tag{9}$$

For a valid pair $(u, v) \in \mathcal{S}_i$, its logit is $L_{iuv} = \ell_i^p(u) + \ell_i^v(v) + h_i(u, v)$, where h_i is a learned pair score from the actor query and the two code embeddings. We use $P_i(u, v) = \exp L_{iuv} / \sum_{(u',v')\in\mathcal{S}_i} \exp L_{iu'v'}$; invalid pairs have zero mass. Thus composed intent aggregates the normalized valid-pair distribution, not independent intent predictions. Path categories follow ordered actor-local rules, later rules taking priority: straight by default; left/right change for $y_T > 1.25$ or $y_T < -1.25$ m; left/right turn for $\theta_T > 0.55$ or $\theta_T < -0.55$ rad and $\max_t |y_t| > 1$ m; reverse for $x_T < -0.5$ m; stationary for $\max_t \|(x_t, y_t)\| < 0.6$ m. Speed categories similarly start with constant speed, assign acceleration/deceleration for $v_T - v_1 > 0.6$ or < -0.6 m/s, then braking-to-stop for $v_T < 0.4$ and $v_1 > 1$ m/s, and finally stopped/crawling for mean speed below 0.4 m/s. Composed categories merge left/right changes and turns, retain reverse, and split straight into constant, accelerating, or decelerating (stopped/crawling joins constant; braking-to-stop joins decelerating). A stationary path maps to stationary with stopped/crawling speed, otherwise to other motion: eight composed categories in total.

A.2 COUNTERFACTUAL RULE DESIGN

Table A2 specifies the rules. Severity $\in \{0.33, 0.66, 1.0\}$ is fixed by scene, actor, and rule identities. Intent is rule-assigned, except cancelled yielding, which reclassifies the dominant edited mode. The token encoder re-encodes the edited condition.

Table A2: Intervention families and the interaction they expose.

Rule	Actor change	Ego-response constraint
Lead braking	Reduce forward motion of a leading vehicle	Remain behind its moving rear boundary
Crossing / cut-in	Move an actor into the reference corridor	Yield/stop for crossing; keep lane for vehicle cut-in
Oncoming turn-across	Bend an approaching vehicle's future across the route	Delay entry until the conflict clears
Roadside pull-out	Move a roadside actor into the route	Yield upstream
Blocked lane change	Occupy the intended lateral maneuver	Keep lane or abort the maneuver
Cancelled yielding	Replace yielding motion with continued entry	Accommodate the renewed conflict

Eligible actors require a valid detection matched to a same-family annotation within 3 m and at least four valid annotated future samples. Edits must alter displacement by at least 0.75 m and pass kinematic, reference-conflict, and dominant-mode actor–actor consistency checks. These checks reject unsupported or ineffective edits before response targets are constructed.

A.3 RESPONSE TARGETS AND OPTIMIZATION

Temporal admissibility and quality. Admissibility is checked for four seconds at 16 Hz, including current poses and all valid actor modes. Each 0.5 s interval is subdivided into eight steps: positions are linearly interpolated and headings follow the shortest wrapped angular difference. Actor segments require both future endpoints to be valid; the first interval uses the current pose and the first future validity bit. These checks use 65 poses including the current one. Speed is capped at 25 m/s, longitudinal acceleration at 8 m/s^2, and lateral acceleration at 6 m/s^2. Crossing responses use a 0.5 s clearance buffer; stopping requires speed below 0.5 m/s for at least 0.5 s within the conflict window. The route corridor spans half an ego width, with 0.5 m pre-conflict margin and terminal overshoot tolerance; backward progress is limited to 0.05 m per dense sample. Collision checks also cover recovery.

Response quality uses

$$\rho_k = 0.7d_k + 0.2e_k + 0.1v_k, \qquad q_k = a_k[0.25 + 0.75(1 - \rho_k)][0.5 + 0.5\eta_k], \tag{10}$$

where d_k is collision-time fraction, e_k emphasizes earlier first collisions, and v_k is soft proximity over a 3 m scale. $\eta_k \in [0, 1]$ is the clipped ratio of candidate to reference terminal forward progress; temporal rules measure progress along the reference route. Selector marginals are projected onto each layer's valid global code identities and normalized separately. We average cross-entropies over contributing path/velocity selectors and layers. Targets and quality weights are detached. Both path and velocity targets must have positive mass for a selector layer to contribute; ranking requires a positive final candidate. Otherwise the term is zero. Known geometric risk targets remain usable. Equation (5) uses $\epsilon = 10^{-8}$.

Interaction-risk targets. For the interaction scorer, oriented footprints are conservatively enclosed by axis-aligned boxes at the eight predicted times. Let (e_x, e_y) and (o_x, o_y) be their half extents. Define

$$d_{k,i,m,t} = \max\{|\Delta x| - e_x - o_x, |\Delta y| - e_y - o_y\}, \qquad r_k = \exp[-\max(\min_{i,m,t} d_{k,i,m,t}, 0)/2]. \tag{11}$$

The ego footprint has half length 2.4 m and half width 1 m; its center is shifted 1.4 m forward from the trajectory point. Risk uses the minimum gap over valid modes and times, treating each retained valid mode as a possible occupancy constraint. This is a conservative geometric target rather than a

probability-weighted collision estimate. This gap proxy is distinct from the dense temporal quality ρ_k above. The 12 scorer features are normalized minimum gap; weighted means of longitudinal and lateral actor offsets, relative-heading cosine and sine, proximity indicator ($d \leq 6$ m), time-weighted proximity, and overlap indicator; ego-valid and joint-valid fractions; and mean ego longitudinal and lateral positions. Gap, longitudinal, and lateral distances use scales 20, 40, and 20 m and are clipped to $[-1, 1]$. Means over actor modes and time use presence times mode probability on valid support; normalized time takes values $1/8, \ldots, 1$. Ego-position means use ego-valid times. These features are detached and passed through a $12 \to 32 \to 1$ multilayer perceptron with SiLU activation and $u = 0.03 \tanh(f)$; it returns zero without valid actor support. An edit is unrelated to a candidate when the changed actor remains more than 6 m away under this gap measure in both worlds. Paired losses use common known support and are averaged over known candidates; the invariance mask does not change that denominator.

Base scoring and factual correction. For predicted metric components $\hat{m}_{k,c}$, the inherited score is

$$s_k = \left(\prod_{c \in \{\mathrm{NAC,DAC,DDC,TLC}\}} \hat{m}_{k,c} \right) (5\hat{m}_{k,\mathrm{TTC}} + 5\hat{m}_{k,\mathrm{EP}} + 2\hat{m}_{k,\mathrm{LK}} + 2\hat{m}_{k,\mathrm{HC}}) \,. \tag{12}$$

EC is evaluated by the benchmark and has no prediction head. Let $k_0 = \arg\max_k s_k$. The factual reranking head predicts a_k^{gain} from detached candidate embeddings. Its bounded correction is

$$s_k^{\mathrm{fact}} = s_k + 14 G_k \max\{0, 0.03 \tanh(a_k^{\mathrm{gain}} - a_{k_0}^{\mathrm{gain}})\}, \tag{13}$$

where G_k admits valid candidates within 0.02 of the best normalized score $s_{k_0}/14$ whose predicted NAC, DAC, DDC, TLC, TTC, LK, and HC are no lower than those of k_0. The gate is detached. The head is trained on factual candidate-score gains relative to the original winner, with weight 0.1. The interaction correction in Eq. (7) is applied afterward. For admission it is added directly to the second velocity layer's logits before pruning $64 \to 10$; empty compatible-path support gives zero correction.

Factual regularizers. Reference-policy retention matches current and frozen-reference candidates by global path–velocity identity. On their common valid support I, with at least two candidates,

$$\mathcal{L}_{\mathrm{retain}} = D_{\mathrm{KL}}(\bar{\pi} \| \pi), \qquad \pi_k = \mathrm{softmax}_I(s_k/(14T)), \quad T = 0.1. \tag{14}$$

The reference distribution is detached. Factual interaction training uses absolute risk on the wide support and two final-support terms. For detached factual score label $y_k \in [0, 1]$, detached normalized score $b_k = s_k^{\mathrm{fact}}/14$, $c = 0.03$, and $k_b = \arg\max y_k$,

$$\begin{aligned} t_k &= \mathrm{clip}(y_k - b_k, -c, c), & \mathcal{L}_{\mathrm{residual}} &= \mathrm{mean}_k\big[((u_k - t_k)/c)^2\big], \\ \Delta_k &= \mathrm{clip}(y_{k_b} - y_k, 0, c), & \mathcal{L}_{\mathrm{guard}} &= \mathrm{mean}_k \frac{[\Delta_k - (b_{k_b} + u_{k_b} - b_k - u_k)]_+}{c}. \end{aligned} \tag{15}$$

The factual residual and geometric targets constrain the same bounded correction; their weighted sum trades off benchmark-score fidelity and local clearance preferences. Thus $\mathcal{L}_{\mathrm{fact}}^{\mathrm{all}}$ combines factual planning supervision, $0.1\mathcal{L}_{\mathrm{retain}}$, and weight 0.1 for each of factual absolute risk, residual, and ranking-guard terms. Equation (8) uses $\lambda_{\mathrm{cf}} = 0.25$ and $\lambda_{\mathrm{inv}} = 0.1$. Losses are averaged over valid candidates within a scene and then over contributing scenes or accepted variants.

Gradient routing. Let $\mathcal{L}_{\mathrm{cf}} = \mathcal{L}_{\mathrm{sel}} + \mathcal{L}_{\mathrm{rank}} + \mathcal{L}_{\mathrm{abs}} + \mathcal{L}_{\Delta} + \lambda_{\mathrm{inv}}\mathcal{L}_{\mathrm{inv}}$ denote the bracketed counterfactual objective in Eq. (8). Let Θ_C, Θ_S, Θ_M, and ϕ denote conditioning, selectors, factual metric/imitation/reranking heads, and the interaction scorer. For each trainable parameter group θ, optimization uses

$$g_\theta = \nabla_\theta \mathcal{L}_{\mathrm{fact}}^{\mathrm{all}} + \lambda_{\mathrm{cf}} M_\theta \nabla_\theta \mathbb{E}_\xi \mathcal{L}_{\mathrm{cf}}, \qquad M_\theta = \begin{cases} 1, & \theta \in \{\Theta_C, \Theta_S, \phi\}, \\ 0, & \theta = \Theta_M. \end{cases} \tag{16}$$

The teacher, reference policy, actor attributes, and interaction-geometry inputs are detached; score computation still passes gradients to conditioned representations.

A.4 Evaluation protocol and comparison sources

Navtest contains 12,146 scenes, with extended comfort applicable to 10,040; missing values are not imputed. Overall scores are averaged per scene. Published ResNet-34 and ViT-L results follow Table 3 of Sun et al. (2026) and Yao et al. (2026), respectively; NC is renamed NAC. The Hydra-MDP++ rows refer to the extension by Li et al. (2025); ARTEMIS is due to Feng et al. (2026). Method references identify the approaches, while the above comparison tables supply the reported backbone-specific scores. Corrected EPDMS adds human-behavior filtering omitted by the original implementation. DiffusionDriveV2's corrected 87.5% comes from Sun et al. (2026). SparseDriveV2 results in both tables are taken directly from the paper.