# CCFM: Collision-Constrained Flow Matching for Safety-Critical Scenario Generation

Ke Li[1], Kaidi Liang[1], Yuxin Ding[2], Debojyoti Biswas[2], Xianbiao Hu[2], and Ruwen Qin[1*]

[1] Stony Brook University, Stony Brook, NY 11794, USA
[2] Pennsylvania State University, University Park, PA 16802, USA
{ke.li.1, kaidi.liang, ruwen.qin}@stonybrook.edu
{ymd5170, dbb5917, xbhu}@psu.edu
[*] Corresponding author

**Abstract.** Evaluation of autonomous vehicle (AV) planners in safety-critical closed-loop simulation is essential for real-world deployment. However, generating controllable safety-critical scenarios remains challenging. Existing approaches use soft guidance that provides only probabilistic preferences and cannot guarantee the satisfaction of geometric and severity constraints associated with specific collision types. We introduce Collision-Constrained Flow Matching (CCFM), a novel framework that guarantees precise collision control through hard physical constraints. CCFM consists of three key components: (i) a heuristic collision selector that optimally identifies an adversarial agent and collision type via composite scoring; (ii) structured hard constraints that explicitly define four collision types (rear-end, side, cut-in, head-on) through contact point, heading, and severity requirements; and (iii) a collision-constrained flow matching sampler that enforces the constraints via Gauss-Newton manifold projection. CCFM achieves collision rate up to 46.4% on nuScenes and 83.1% on nuPlan, significantly outperforming baselines while preserving realistic driving behavior. By enabling controllable collision characteristics in safety-critical scenario generation, CCFM provides a reliable foundation for AV safety evaluation and sim-to-real crash data generation. The code and implementation details are available at https://github.com/KELISBU/CCFM.

## 1 Introduction

Safety-critical scenario generation is essential for evaluating the robustness of autonomous vehicles (AV) [7], as such scenarios are rare, lying in the long tail of the driving-event distribution [18] and often lead to highly consequential interactions. These events are difficult to collect at scale from naturalistic driving data due to their low frequency and high risk. Therefore, closed-loop simulation provides a critical alternative by enabling repeatable and scalable interactive AV testing. In this context, explicitly generating realistic and controllable scenarios, in which an adversarial agent is designated to challenge the ego vehicle, is essential to expose the vulnerabilities of AV planners before real-world deployment.

While recent advancements have explored safety-critical scenario generation and closed-loop simulation, significant gaps remain in balancing *controllability*, precise *collision constraints*, and *behavior realism*. First, traditional optimization-based [10] or RL-based [14] methods often induce overly aggressive maneuvers that deviate from human-like driving behavior, making it difficult to preserve realistic multi-agent interaction patterns. More recently, diffusion-based generative approaches [3, 16, 33] improve behavior realism by learning from traffic priors. However, their controllability typically relies on cost-based *soft guidance*, which cannot rigorously ensure the satisfaction of spatial-geometric and severity constraints. Moreover, under dense and dynamic traffic context, pre-selecting an adversarial agent and its most feasible collision mode remains an underexplored problem.

In this work, we reformulate the safety-critical scenario generation as a collision-constrained sampling problem rather than unconstrained guided sampling or adversarial optimization. Our key idea is to preserve the learned prior of realistic traffic scenarios, while enforcing structured collision constraints that specify the desired event type. Built upon flow matching [17,28], which generates samples by integrating an ordinary differential equation (ODE), our framework naturally allows intermediate projection steps to correct the evolving sample toward a collision-constrained feasible set.

In summary, we propose a Collision-Constrained Flow Matching (**CCFM**) framework. The main contributions of this paper are as follows:

- Introduction of a Heuristic Collision Selector (HCS) that scores candidate agent-collision pairs based on topological reachability, geometry, and road legality, enabling dynamic identification of the most threatening yet physically viable interaction.
- Formulation of a comprehensive collision-type-specific constraint set that mathematically formulates the physical constraints governing the contact point, relative heading, and severity metrics for different collision types in adversarial trajectory generation.
- Development of a Gauss-Newton manifold projection method that projects intermediate action sequences onto the feasible set while preserving plausible driving behavior.

Extensive experiments on nuScenes [1] and nuPlan [2] closed-loop simulation illustrate that CCFM generates more safety-critical scenarios than prior methods, while achieving interpretable collision-type controllability, diverse collision outcomes, and competitive realism.

## 2 Related Work

### 2.1 Generative Models for Traffic Scenario Generation

Early data-driven approaches established the foundation for multimodal interactive trajectory generation using diverse paradigms, including conditional variational autoencoders (CVAEs) [25], generative adversarial networks (GANs) [9], energy-based models [22], and autoregressive transformers [8, 39]. To achieve better distributional expressivity and controllability, diffusion models [12, 26] and Flow Matching [17, 19, 28] methods have recently emerged as the dominant baseline. By framing generation as a progressive denoising process, frameworks such as Diffuser [13] unify generative priors with planning objectives through classifier-guided sampling. However, Flow Matching provides a more efficient continuous-time formulation for distribution transport. For instance, Flow Planner [27] and GoalFlow [32] successfully leverage it to model interactive multi-agent behaviors and goal-driven multimodal trajectories, demonstrating significant inference speedups and robust performance on large-scale planning benchmarks.

### 2.2 Safety-critical Scenario Generation with Controllability

Rigorous validation of autonomous vehicles necessitates closed-loop simulation to generate rare, but safety-critical scenarios. Existing generation paradigms can be broadly grouped into (i) optimization-based methods, such as AdvSim [31], ReGentS [35], and KING [10], which iteratively perturb trajectories toward risk; (ii) hybrid approaches combining learned priors with optimization and rule-based rollouts, like STRIVE [24] and SLEDGE [5]. More recently, frameworks such as DiffScene [33] and SAFE-SIM [3] exemplify the diffusion-based paradigm, employing guided cost functions to steer realistic multi-agent distributions toward collision-prone outcomes during closed-loop rollouts. Recent studies further improve controllability from different perspectives, including closed-loop adversarial resampling [37], feasibility-guided adversarial policy learning [4], inverse motion prediction for inserted adversarial agents [20], and preference-aligned steerable scenario generation [21]. To further enhance high-level controllability, recent studies integrate large language models (LLMs). For instance, LD-Scene [23] translates natural language intents into executable adversarial loss functions for latent diffusion, whereas ChatScene [36] acts as an autonomous agent that bridges unstructured text into structured configurations of CARLA simulator.

Diffusion models steer generative distributions toward risk via soft guidance [3, 16, 23], rather than enforcing hard constraints, thereby limiting controllability. To overcome this limitation, recent works such as PCFM [30] and HardFlow [15] enforce hard constraints at inference time by projecting intermediate flow states onto the constraint manifold without retraining. However, both methods have been developed and evaluated exclusively in the context of physics-governed PDE systems. Whether such hard-constraint enforcement transfers to interactive scenario generation remains an open question, since traffic constraints are dynamics-mediated and defined relative to an evolving ego trajectory.

# 3 Methodology

We propose Collision-Constrained Flow Matching (CCFM), a framework for safety-critical scenario generation, illustrated in Fig. 1. CCFM is formulated within a closed-loop simulation environment. The Heuristic Collision Selector (HCS) dynamically identifies the most threatening yet physically viable collision type and the associated adversarial vehicle. For each collision type, a set of collision-type-specific constraints are defined, resulting in the collision-constrained Flow Matching problem. CCFM is solved via a damped Gauss-Newton projection for generating the desired safety-critical scenario. Details of the methodology are presented as follows.

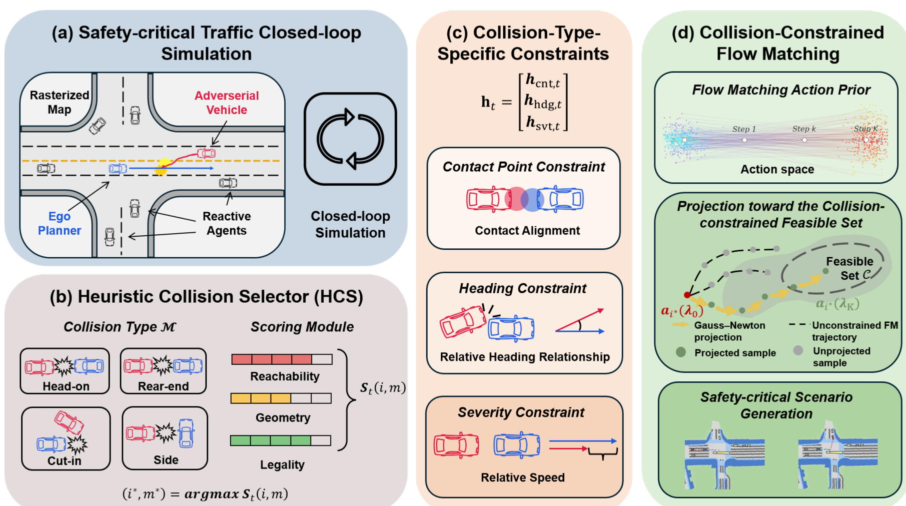


**Fig. 1:** Overview of the proposed safety-critical scenario generation by CCFM.

## 3.1 Safety-critical Traffic Generation in Closed-loop Simulation

In the closed-loop simulation, we consider a traffic scene with $N$ vehicles: an *ego vehicle* ($i =$ ego) and $N-1$ non-ego vehicles. One non-ego vehicle is selected as the *adversarial agent* and tasked with colliding with the ego vehicle, while the others are *reactive agents* that respond to traffic to preserve realism. Our goal is to generate safety-critical scenarios that satisfy controllable conditions for any specific collision type.

$\mathbf{s}_{i,t} = [x_{i,t}, y_{i,t}, \vartheta_{i,t}, v_{i,t}]$ denotes the state of vehicle $i$ at timestep $t$, $\forall\, i \in \{1, \ldots, N-1\} \cup \{\text{ego}\}$. $\mathbf{p}_{i,t} = [x_{i,t}, y_{i,t}]$ represents vehicle $i$'s 2D position, while $\vartheta_{i,t}$ and $v_{i,t}$ denote its heading angle and linear velocity, respectively. At timestep $t$, the action of vehicle $i$, $\mathbf{a}_{i,t} = [\dot{v}_{i,t}, \dot{\vartheta}_{i,t}]$, consists of its acceleration and yaw rate. The selection of this action is conditioned on the vehicle's context, $\mathbf{c}_{i,t}$, defined in the supplementary material. The state evolution is governed by a one-step deterministic unicycle dynamics model $f$:

$$\mathbf{s}_{i,t+1} = f(\mathbf{s}_{i,t}, \mathbf{a}_{i,t}). \tag{1}$$

The ego vehicle drives autonomously, with its future trajectory $\boldsymbol{\tau}_{\text{ego},t:t+T}$ controlled by an AV planner $\pi$:

$$\boldsymbol{\tau}_{\text{ego},t:t+T} = \pi(\mathbf{c}_{\text{ego},t}). \tag{2}$$

Non-ego vehicles are governed by a generative prior $p_\theta(\cdot \mid \mathbf{c}_t)$ learned via Flow Matching, which captures the distribution of realistic driving trajectories. We generate the action sequence for each non-ego vehicle, $\mathbf{a}_{i,t:t+T-1}$, from this learned prior:

$$\mathbf{a}_{i,t:t+T-1} \sim p_\theta(\cdot \mid \mathbf{c}_{i,t}), \tag{3}$$

and obtain the corresponding trajectory, $\boldsymbol{\tau}_{i,t:t+T}$, using a forward $T$-step rollout function $\mathcal{F}$:

$$\boldsymbol{\tau}_{i,t:t+T} = \mathcal{F}(\mathbf{s}_{i,t}, \mathbf{a}_{i,t:t+T-1}). \tag{4}$$

The generation of trajectories for reactive agents is *unconstrained*, whereas the sampled action sequence for the adversarial agent must yield a trajectory satisfying the hard physical constraints for a given collision type, to be introduced in Sec. 3.3.

In a receding-horizon closed-loop simulation, all vehicles update their contexts from the latest states, re-plan their horizon-$T$ action sequences, execute only the first action, and repeat.

### 3.2 Heuristic Collision Selector

At each timestep $t$, the selection of the most plausible safety-critical scenario to generate is formulated as an optimization problem that jointly identifies the adversarial agent and its specific collision type with the ego vehicle. Directly optimizing over all candidate agents and collision types when solving the constrained Flow Matching problem for trajectory generation is computationally inefficient and often physically implausible. Therefore, we decouple this selection process from the constrained Flow Matching formulation by introducing a Heuristic Collision Selector (HCS) to identify an optimal pair of adversarial agent and collision type, $(i^*, m^*)$, that maximizes the collision score function:

$$S_t(i, m) = w_{\text{rch}} S_{\text{rch},t}(i) + w_{\text{geo}} S_{\text{geo},t}(i, m) + w_{\text{lgt}} S_{\text{lgt},t}(i, m), \tag{5}$$

where $i \in \{1, \ldots, N-1\}$ and $m \in \mathcal{M} = \{\text{Rear-End}, \text{Side}, \text{Cut-In}, \text{Head-On}\}$. The collision score in Eq. (5) is defined as a weighted sum of the reachability score ($S_{\text{rch},t}$), the geometry score ($S_{\text{geo},t}$), and the legality score ($S_{\text{lgt},t}$), with weights $w_{(\cdot)}$ summing to one. These similarity-based scores are outlined below, and their modeling details are provided in the supplementary material.

The reachability score for agent $i$ at timestep $t$ is calculated from its position relative to the ego vehicle:

$$S_{\mathrm{rch},t}(i) = \psi_{\mathrm{rch}}(\mathbf{p}_{i,t}, \mathbf{p}_{\mathrm{ego},t}), \tag{6}$$

where $\psi_{\mathrm{rch}}(\cdot)$ measures the proximity-based similarity between the two vehicles, reflecting the higher collision likelihood of nearby agents.

The geometry score for agent $i$ under collision type $m$ is computed by comparing the agent's relative pose with respect to the ego vehicle at time $t$, $\mathbf{pos}_{i,t}$, against the typical pose associated with that collision type, $\mathbf{pos}_m$:

$$S_{\mathrm{geo},t}(i,m) = \psi_{\mathrm{geo}}(\mathbf{pos}_{i,t}, \mathbf{pos}_m), \tag{7}$$

where $\mathbf{pos}_{i,t}$ is derived from the positions and heading angles of the two vehicles, and $\psi_{\mathrm{geo}}(\cdot)$ measures the similarity between $\mathbf{pos}_{i,t}$ and the reference pose $\mathbf{pos}_m$.

The legality score is computed by comparing the topological relationship between agent $i$ and the ego vehicle at timestep $t$, $\mathbf{tpl}_{i,t}$, against the typical relationship for collision type $m$, $\mathbf{tpl}_m$:

$$S_{\mathrm{lgt},t}(i,m) = \psi_{\mathrm{lgt}}(\mathbf{tpl}_{i,t}, \mathbf{tpl}_m), \tag{8}$$

where $\mathbf{tpl}_{i,t}$ is determined based on the contexts, positions, and heading angles of the two vehicles. $\psi_{\mathrm{lgt}}(\cdot)$ measures the similarity between $\mathbf{tpl}_{i,t}$ and $\mathbf{tpl}_m$, ensuring that the generated adversarial behavior remains map-compliant and realistic.

We note that HCS is implemented as a plug-in function that can be overridden by users' manual selections, providing flexibility to generate user-specified scenarios.

### 3.3 Collision-Type-Specific Constraints

At timestep $t$, after HCS identifies the adversarial agent $i^*$ and the collision type $m^*$ for the generation of safety-critical scenario, a set of constraints is imposed to ensure that the adversarial agent's trajectory will lead to the desired collision type at the future timestep $t + T_{\mathrm{col}}$. Here, $T_{\mathrm{col}}$ is the target time-to-collision, with details of its selection provided in the supplementary material.

We formulate these constraints through residual measures:

$$\mathbf{h}_t = \begin{bmatrix} h_{\mathrm{cnt},t} \\ h_{\mathrm{hdg},t} \\ h_{\mathrm{svt},t} \end{bmatrix}, \tag{9}$$

where $h_{\mathrm{cnt},t}$, $h_{\mathrm{hdg},t}$, and $h_{\mathrm{svt},t}$ denote the residual functions related to contact-point (cnt), relative-heading (hdg), and impact severity (svt) constraints, respectively:

$$h_{j,t} = \phi_j(l_{j,t}, \bar{l}_j), \quad \text{for } j \in \{\mathrm{cnt}, \mathrm{hdg}, \mathrm{svt}\}. \tag{10}$$

In eq. (10), $l_{j,t}$ denotes the $j$-measure computed from the predicted states of the adversarial agent and the ego vehicle at the target time of impact $t + T_{\mathrm{col}}$, and

$\bar{l}_j$ represents the corresponding target value for collision type $m^*$. The function $\phi_j(\cdot)$ maps $l_{j,t}$ and $\bar{l}_j$ to a dissimilarity score, enabling the related constraints to be imposed in action sequence generation. Details of the derivations for $l_{j,t}$, $\bar{l}_j$, and $\phi_j$ for all $j$ are provided in the supplementary material.

### 3.4 Collision-Constrained Flow Matching

**Unconstrained Action Sequence Sampling.** Following the conditional Flow Matching method [28], a velocity field $\mathbf{v}_\theta(\lambda, \mathbf{a}(\lambda), \mathbf{c})$ is learned, where $\lambda \in [0, 1]$ is the flow time variable. At inference time, to sample an action sequence for agent $i$, $\mathbf{a}_{i,t:t+T-1}$, we draw an initial action sequence from a standard Gaussian distribution: $\mathbf{a}_i(\lambda_0) \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, and then transport it to the learned conditional action prior $p_\theta(\cdot|\mathbf{c}_{i,t})$ by solving the ordinary differential equation (ODE):

$$d\mathbf{a}_i(\lambda)/d\lambda = \mathbf{v}_\theta(\lambda, \mathbf{a}_i(\lambda), \mathbf{c}_{i,t}), \tag{11}$$

from $\lambda$=0 to $\lambda$=1. The terminal point $\mathbf{a}_i(1)$ is the planned action sequence $\mathbf{a}_{i,t:t+T-1}$ in Eq. (3), which is subsequently propagated through the dynamics model $\mathcal{F}$ in Eq. (4) to produce the future state trajectory $\boldsymbol{\tau}_{i,t:t+T}$.

We numerically integrate the unconstrained ODE in Eq. (11) using a discrete Euler solver with $K$ steps. The intermediate sample is updated at each integration step $k = 0, 1, \ldots, K-1$:

$$\mathbf{a}_i(\lambda_{k+1}) = \mathbf{a}_i(\lambda_k) + \Delta\lambda \cdot \mathbf{v}_\theta(\lambda_k, \mathbf{a}_i(\lambda_k), \mathbf{c}_{i,t}), \tag{12}$$

where $\Delta\lambda = 1/K$ and $\lambda_k = k\Delta\lambda$. However, this unconstrained integration alone does not guarantee that the adversarial trajectory satisfies the desired collision conditions.

**Manifold Projection in Flow Matching.** With the set of three residual functions $\mathbf{h}_t(\cdot)$ defined in Sec.3.3, CCFM enforces collision conditions when sampling adversarial action sequences. Following the step-wise projection strategy of PCFM [30], we project intermediate action sequences onto the collision-constrained feasible set at every ODE step, and propagate the projected sequence back to the current flow time via an optimal-transport (OT) reverse update.

*Collision-constrained Feasible Set.* At each integration step, the intermediate action sequence is projected onto the collision-constrained feasible set:

$$\mathcal{C} = \{\mathbf{a}_{i^*,t:t+T-1} \mid \mathbf{h}_t = \mathbf{0}\}, \tag{13}$$

where $\mathbf{h}_t$ is the set of residual functions defined in Eq.(9). We denote the corresponding projection operator by $\Pi_\mathcal{C} : \mathbb{R}^{T\times 2} \to \mathcal{C}$, which maps an action sequence to the closest sequence satisfying $\mathbf{h}_t = \mathbf{0}$ at the dynamic time-to-collision $T_{\text{col}}$. Here, $T_{\text{col}}$ is updated once per closed-loop re-planning step. In practice we implement $\Pi_\mathcal{C}$ as a damped Gauss-Newton (GN) projection with Jacobians obtained by automatic differentiation through the forward-dynamics rollout $\mathcal{F}$. Full algorithmic details of the dynamic time-to-collision and GN projection are provided in the supplemental material.

**Algorithm 1** Collision-Constrained Flow Matching (CCFM) Sampling

**Input:** Adversarial agent's context $\mathbf{c}_{i^*,t}$ and state $\mathbf{s}_{i^*,t}$, ego trajectory $\boldsymbol{\tau}_{\text{ego},t:t+T}$, target collision type $m^*$, ODE integration steps $K$
**Output:** Adversarial action sequence $\mathbf{a}_{i^*,t:t+T-1}$
Sample $\mathbf{a}_{i^*}(\lambda_0) \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$; $\quad \Delta\lambda \leftarrow 1/K$
Compute $T_{\text{col}}$ from current ego–adversary kinematics ▷ See supplementary
**for** $k = 0, 1, \ldots, K-1$ **do**
  $\lambda_k \leftarrow k \cdot \Delta\lambda$
  $\mathbf{v} \leftarrow \mathbf{v}_\theta(\lambda_k, \mathbf{a}_{i^*}(\lambda_k), \mathbf{c}_{i^*,t})$ ▷ Predict flow velocity
  $\hat{\mathbf{a}}_{i^*}(\lambda_{k+1}) \leftarrow \mathbf{a}_{i^*}(\lambda_k) + \Delta\lambda \cdot \mathbf{v}$ ▷ Euler step
  $\mathbf{a}_{\text{proj}} \leftarrow \Pi_{\mathcal{C}}(\hat{\mathbf{a}}_{i^*}(\lambda_{k+1}))$ ▷ Damped GN projection; see supplementary
  $\mathbf{a}_{i^*}(\lambda_{k+1}) \leftarrow \lambda_{k+1}\,\mathbf{a}_{\text{proj}} + (1-\lambda_{k+1})\,\mathbf{a}_{i^*}(\lambda_0)$ ▷ OT reverse update
**end for**
**return** $\mathbf{a}_{i^*}(\lambda_K)$

*OT Reverse Update.* Applying $\Pi_{\mathcal{C}}$ directly at the current flow time $\lambda_{k+1}$ produces a large discontinuity when $\lambda_{k+1}$ is small, since the unprojected sample is still far from the data manifold. Similar to PCFM [30], we blend the projected sequence with the initial noise $\mathbf{a}_{i^*}(\lambda_0)$ along the optimal-transport interpolant:

$$\mathbf{a}_{i^*}(\lambda_{k+1}) \;=\; \lambda_{k+1}\,\Pi_{\mathcal{C}}(\hat{\mathbf{a}}_{i^*}(\lambda_{k+1})) + (1-\lambda_{k+1})\,\mathbf{a}_{i^*}(\lambda_0), \tag{14}$$

where $\hat{\mathbf{a}}_{i^*}(\lambda_{k+1}) = \mathbf{a}_{i^*}(\lambda_k) + \Delta\lambda\,\mathbf{v}_\theta(\lambda_k, \mathbf{a}_{i^*}(\lambda_k), \mathbf{c}_{i^*,t})$ is the unprojected Euler step. Eq. (14) recovers the unprojected sample as $\lambda_{k+1} \to 0$ and the fully-projected sample at $\lambda_{k+1} = 1$, so the constraint is satisfied exactly at the terminal step while early-step iterates remain close to the learned flow. The sampling procedure is summarized in Algorithm 1.

# 4 Experimental Setup

## 4.1 Datasets

We conduct extensive experiments on the real-world driving dataset nuScenes [1], and nuPlan [2]. Our CCFM model is trained on the nuScenes train split and evaluated in the closed-loop simulation on the nuScenes validation split and nuPlan mini validation split. In addition to the standard planning setting with an 8-second horizon, we further perform the closed-loop simulation over the duration of 20 seconds.

## 4.2 Implementation Details

**Closed-loop Simulation Setup.** Our closed-loop traffic simulation is built upon TBSim [34] with a re-planning frequency of 2 Hz. We consider three planners for the ego vehicle: (i) a rule-based planner operating on the lane graph following STRIVE [24], (ii) PDM [6], a deterministic closed-loop planner, and

(iii) the Intelligent Driver Model (IDM) [29]. To evaluate the effectiveness of our approach, we compare CCFM against three baselines for closed-loop simulation: STRIVE [24], CCDiff [16], and SAFE-SIM [3].

**Flow Matching Model.** Following previous studies [3, 38], we encode the rasterized map and contextual trajectories using a ResNet [11] backbone, and adopt a U-Net architecture with several 1D convolutional blocks to model temporal consistency. During training, we use 10 historical frames (i.e., $T_{\text{hist}}$=10) as context and predict the 32 future frames (i.e., $T$=32). Additional details about training and inferring the Flow Matching model are provided in the supplementary material.

### 4.3 Evaluation Metrics

To generate controllable and diverse safety-critical scenarios with realistic behaviors, we validate the proposed CCFM method on three dimensions: scenario criticality, behavior realism, and scenario controllability and diversity.

**Scenario Criticality.** We quantify the criticality of generated scenarios using (i) the Collision Rate (CR) and (ii) the collision severity. Specifically, CR is defined as the percentage of simulated scenes in which at least one collision involving the ego vehicle and the adversarial agent. For collision severity, we compute the value of relative speed (MS) between the ego and the adversarial vehicles at the collision time, and report the mean value over all collision events.

**Behavioral Realism.** Adopting the approach from SAFE-SIM [3] and CTG [38], we evaluate realism (RM) by computing the Wasserstein distance between ground truth and generated scene on normalized histograms of the driving longitudinal acceleration, lateral acceleration, and jerk. In addition, the Off-Road (OR) rate reports the ratio of reactive agents driving off-road across the scene level.

**Scenario Controllability and Diversity.** We assess controllability by measuring the ratio of generated safety-critical events that match the desired event specification. Specifically, we compute the collision Type-Match (TM) rate between the generated event and the target type selected from HCS. To quantify diversity among generated safety-critical outcomes, we measure the distribution of collision contact regions on the ego vehicle at impact, reported as the Front, Rear, and Side rates by entropy score (EN). Meanwhile, we report the standard deviation of (i) the relative speed (SD) and (ii) the relative heading (HD) at first impact.

We also report the wall-clock runtime of a full closed-loop simulation (TIME), and provide detailed formulations of all metrics and composite scores, including the Criticality Score (CS), Diversity Score (DS), Realism Score (RS), and the overall Composite Weighted Score (CWS), in the supplementary material.

## 5 Evaluation and Analysis

In this section, we evaluate the proposed CCFM framework for controllable safety-critical scenario generation in closed-loop traffic simulation on nuScenes

**Table 1:** Closed-loop simulation results under different rollout horizons. CR is the collision rate for adversarial agent and OR is the off-road rate for reactive agents. † denotes relative composite scores normalized within this comparison table.

| Dataset | Model | CR (%) ↑ | MS ($m/s$) ↑ | TM (%) ↑ | CS† ↑ | OR (%) ↓ | RM ↓ | RS† ↑ | Time (s) ↓ |
|---|---|---|---|---|---|---|---|---|---|
| **Horizon = 80** | | | | | | | | | |
| nuScenes [1] | CCDiff [16] | 4.0 | 0.3 | – | 0.10 | **4.4** | **0.49** | **1.00** | 180.5 |
| | STRIVE [24] | 20.8 | 1.9 | – | 0.56 | 5.8 | 0.83 | 0.67 | 466.3 |
| | SAFE-SIM [3] | 25.8 | 0.8 | – | 0.42 | 5.0 | 0.63 | 0.83 | 128.9 |
| | CCFM (Ours) | **46.4** | **2.8** | **84.3** | **1.00** | 5.4 | 0.78 | 0.72 | **123.0** |
| nuPlan [2] | SAFE-SIM [3] | 42.1 | 2.4 | – | 0.48 | 3.6 | 0.33 | 0.74 | **271.4** |
| | CCFM (Ours) | **83.1** | **5.2** | **85.6** | **1.00** | **1.7** | **0.33** | **1.00** | 356.3 |
| **Horizon = 200** | | | | | | | | | |
| nuScenes [1] | CCDiff [16] | 11.3 | 0.7 | – | 0.18 | **9.1** | 0.56 | **0.90** | 533.6 |
| | SAFE-SIM [3] | 26.1 | 1.5 | – | 0.41 | 11.4 | **0.45** | 0.90 | 322.5 |
| | CCFM (Ours) | **59.9** | **4.0** | **85.4** | **1.00** | 11.4 | 0.71 | 0.72 | **316.3** |
| nuPlan [2] | SAFE-SIM [3] | 50.6 | 3.0 | – | 0.51 | 8.7 | **0.27** | 0.61 | **728.4** |
| | CCFM (Ours) | **95.2** | **6.1** | **84.7** | **1.00** | **1.9** | 0.30 | **0.95** | 970.5 |

and nuPlan. Unless otherwise specified, all experiments are conducted on nuScenes. The results show that our method can generate safety-critical scenarios with explicit control over desired collision types while preserving physical plausibility.

### 5.1 Comparison with SOTA Models

First, we compare CCFM with STRIVE, CCDiff, and SAFE-SIM on nuScenes and nuPlan under different rollout horizons using the same Lane-graph ego planner [24]. As shown in Table 1, under the practical setting of an 80-frame rollout horizon, CCFM achieves the highest collision rate on both datasets, reaching 46.4% on nuScenes and 83.1% on nuPlan. CCFM also achieves the highest severity of collision with the highest relative speed, demonstrating stronger safety-critical scenario generation capability. Extending the rollout horizon to 200 frames further improves the collision rate to 59.9% on nuScenes, and 95.2% on nuPlan, suggesting that a longer generation horizon evolves more safety-critical events with controllability. Despite some reduction in realism metrics, the substantial improvements in collision rate and severity make the trade-off acceptable.

To further explore the observed trade-off between realism and criticality, Table 2 details the realism evaluation of generated safety-critical scenarios at the scene level and the adversary agent level, respectively. The results show that CCFM achieves the highest collision rate, but maintains competitive behavior realism at both levels. Critically, the deviations are kinematically regulated,

**Table 2:** Trade-off between criticality and realism. ADV.ADE and ADV.FDE denote the average displacement error and final displacement error of the adversarial agent, respectively. ADE, and FDE report the corresponding metrics over all generated trajectories. ADV.OR and OR report the off-road rate of the adversarial agent and reactive agents respectively. Metric definitions are detailed in the supplementary material.

| Model | CR (%) ↑ | ADV.OR (%) ↓ | ADV.ADE (m) ↓ | ADV.FDE (m) ↓ | OR (%) ↓ | ADE (m) ↓ | FDE (m) ↓ |
|---|---|---|---|---|---|---|---|
| STRIVE | 20.8 | 11.2 | 15.9 | 27.4 | 5.8 | 6.8 | 14.5 |
| SAFE-SIM | 25.8 | **8.7** | **8.1** | **18.7** | **5.0** | **4.6** | **11.1** |
| CCFM | **46.4** | 10.0 | 12.0 | 24.9 | 5.4 | 5.9 | 13.3 |

reflecting aggressive yet physically plausible maneuvers rather than unrealistic artifacts.

### 5.2 Evaluation on Controllability and Diversity

**Controllability.** Quantitative results in Table 1 illustrate that only CCFM provides explicit collision-type controllability. CCFM achieves 84.3% type match under the 80-frame rollout horizon and 85.4% under the 200-frame rollout horizon, whereas STRIVE, CCDiff, and SAFE-SIM do not provide explicit type-level control. The consistent performance across the two rollout horizons indicates that CCFM can reliably steer generated scenarios toward the desired collision outcome, rather than merely increasing the overall collision rate.

Furthermore, Fig. 2 qualitatively illustrates the controllability of CCFM. Given identical initial scene contexts and a fixed ego-adversary pair, CCFM can generate distinct safety-critical interactions by altering the target constraint. In particular, the *Cut-in*, *Head-on*, *Side*, and *Rear-end* examples are generated from the same scene context. These visualizations demonstrate that CCFM provides fine-grained, interpretable control over the final collision outcomes.

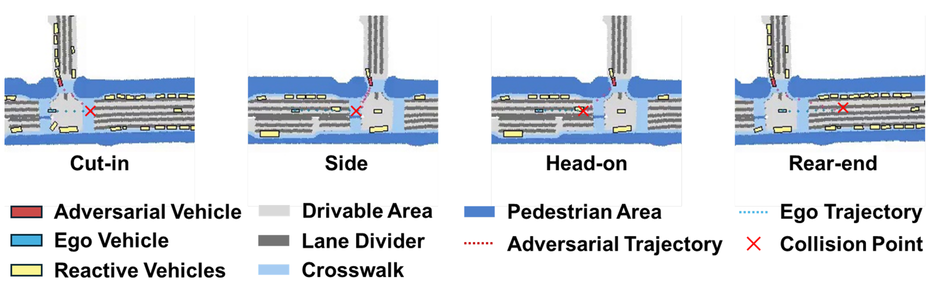


**Fig. 2:** Qualitative visualization of collision-type controllability. Given an assigned ego–adversary pair and scene context, CCFM generates distinct scenarios by specifying target collision types.

**Table 3:** Comparison of diversity results with baseline models. We evaluate the distribution of three contact regions at collision by EN. Meanwhile, the standard deviation SD and HD reflect the diversity of relative speed and heading angle. † denotes relative composite scores normalized within this comparison table.

| Model | Rear (%) | Front (%) | Side (%) | EN ↑ | SD ($m/s$) ↑ | HD (°) ↑ | DS† ↑ |
|---|---|---|---|---|---|---|---|
| STRIVE [24] | 18.4 | 24.2 | 57.4 | 0.86 | $1.9 \pm 2.8$ | $2.2 \pm 12.9$ | 0.62 |
| CCDiff [16] | 26.6 | 28.2 | 45.2 | **0.97** | $0.3 \pm 1.7$ | $1.5 \pm 14.1$ | 0.59 |
| SAFE-SIM [3] | 10.2 | 28.6 | 61.2 | 0.81 | $0.8 \pm 2.5$ | $2.4 \pm 15.8$ | 0.61 |
| CCFM (Ours) | 50.5 | 14.2 | 35.3 | 0.89 | $\mathbf{2.8 \pm 4.9}$ | $\mathbf{12.3 \pm 30.9}$ | **0.96** |

**Diversity.** Compared to STRIVE, CCDiff, and SAFE-SIM, our method achieves the second-best contact-region entropy with EN=0.89, indicating a relatively diverse distribution over Rear, Front, and Side ego contact regions, as shown in Table 3. Although CCDiff obtains the highest entropy score of EN=0.97, its collision rate remains much lower at only 4.0%. In addition, CCFM achieves a larger standard deviation in relative speed (SD = 4.9 m/s) and relative angle (HD = 30.9°), indicating richer diversity in collision kinematics and interaction geometry. The highest DS = 0.96 suggests that CCFM does not collapse to a single collision configuration, but is able to generate safety-critical scenarios with various impact forms and motion characteristics.

### 5.3 Assessment Across Ego Planners

We evaluate the proposed CCFM framework with three ego planners: IDM [29], PDM [6], and the lane-graph method used in STRIVE [24]. As shown in Table 4, CCFM consistently achieves high collision rate, and strong type match performance across different ego planners, indicating planner-agnostic controllability and robust effectiveness. Although CCFM achieves a relatively lower collision rate with the PDM planner, this is expected because PDM's structured motion planning and explicit collision-avoidance mechanism allow the ego vehicle to execute more robust evasive maneuvers.

Meanwhile, the variation across ego planners further suggests that stronger planning algorithms can avoid a subset of the generated safety-critical interactions. However, our current evaluation does not explicitly distinguish avoidable

**Table 4:** Closed-loop simulation results using different ego planners for assessing CCFM planner-agnostic controllability and effectiveness.

| Planner | CR (%) ↑ | MS ($m/s$) ↑ | TM(%) ↑ | OR(%) ↓ | RM↓ |
|---|---|---|---|---|---|
| IDM [29] | **51.6** | **3.3** | 82.7 | **5.1** | 0.81 |
| PDM [6] | 36.1 | 2.6 | 81.2 | 8.5 | 0.85 |
| Lane-graph [24] | 46.4 | 2.8 | **84.3** | 5.4 | **0.78** |

**Table 5:** Ablation study on the Heuristic Collision Selector (HCS). We compare fixed single-type selectors with the HCS, which adaptively assigns ego-adversary pair and feasible collision type. † denotes relative composite scores normalized within this comparison table.

| HCS | **CR** (%) ↑ | **TM** (%) ↑ | **MS** ($m/s$) ↑ | **CS**† ↑ | **EN** ↑ | **DS**† ↑ | **OR** (%) ↓ | **RM** ↓ | **RS**† ↑ | **CWS**† ↑ |
|---|---|---|---|---|---|---|---|---|---|---|
| Rear-end only | 30.2 | 70.7 | 2.2 | 0.51 | **0.96** | 0.69 | 7.0 | 0.86 | 0.84 | 0.68 |
| Head-on only | 27.7 | 16.4 | 2.7 | 0.36 | 0.72 | 0.83 | 7.5 | 0.85 | 0.82 | 0.67 |
| Side only | 31.4 | 14.5 | 3.1 | 0.41 | 0.64 | 0.75 | 6.7 | 0.87 | 0.85 | 0.67 |
| Cut-in only | 40.0 | 4.6 | **4.4** | 0.52 | 0.79 | 0.82 | 5.5 | 0.80 | 0.98 | 0.77 |
| HCS | **46.4** | **84.3** | 2.8 | **0.82** | 0.89 | **0.84** | **5.4** | **0.78** | **1.00** | **0.89** |

collisions from physically unavoidable ones. Future work will incorporate avoidability analysis to better quantify the planner relevance of generated scenarios, and explore their use as targeted training data to enhance the robustness and generalization of downstream planning models.

### 5.4 Ablation Study

We conduct a series of ablation studies to evaluate the effectiveness of three key components: Heuristic Collision Selector, Collision-Type-specific Constraints, and Collision-Constrained Gauss–Newton Projection of our CCFM framework in safety-critical closed-loop simulation.

**Effectiveness of Heuristic Collision Selector.** Table 5 shows that HCS achieves the best overall performance compared to fixed single-type selectors. Although individual fixed strategies may perform well on a single metric, such as higher EN for *Rear-end only* or higher MS for *Cut-in only*, they suffer from limited controllability or realism. In contrast, HCS provides a more effective trade-off among criticality, diversity, controllability, and realism. These results demonstrate that adaptively selecting the adversarial agent and collision type based on reachability, geometric alignment, and lane-graph legality provides a feasible basis for satisfying the subsequent collision-type-specific constraints.

**Effectiveness of Collision Type-specific Constraint Components.** As shown in Table 6, each constraint component contributes differently to controllability, diversity, realism, and overall generation quality. Using only the contact point constraint already produces safety-critical scenarios with a collision rate of 51.6% and a type match of 82.1%, showing that $h_{\mathrm{cnt},t}$ is effective for triggering collisions at the desired contact region. Adding the relative heading constraint slightly reduces the collision rate to 48.4%, but improves the type match to 83.8% and diversity score to 0.92, indicating that $h_{\mathrm{hdg},t}$ provides stronger

**Table 6:** Ablation study on different components in collision-type-specific constraint. We evaluate the contribution of contact point constraint $h_{\mathrm{cnt}}$, relative heading constraint $h_{\mathrm{hdg}}$, and severity constraint $h_{\mathrm{svt}}$ to safety-critical scenario generation in closed-loop simulation. † denotes relative composite scores normalized within this comparison table.

| $h_{\mathrm{cnt},t}$ | $h_{\mathrm{hdg},t}$ | $h_{\mathrm{svt},t}$ | **CR** (%) ↑ | **TM** (%) ↑ | **MS** ($m/s$) ↑ | **CS**† ↑ | **EN** ↑ | **DS**† ↑ | **RS**† ↑ | **CWS**† ↑ |
|---|---|---|---|---|---|---|---|---|---|---|
| ✓ | | | **51.6** | 82.1 | 2.3 | 0.91 | 0.85 | 0.90 | 0.94 | 0.92 |
| ✓ | ✓ | | 48.4 | 83.8 | 2.6 | 0.94 | 0.86 | 0.92 | 0.94 | 0.93 |
| ✓ | ✓ | ✓ | 46.4 | **84.3** | **2.8** | **0.96** | **0.89** | **0.96** | **0.96** | **0.96** |

geometric controllability over the target collision type and greater diversity respectively. When the full set of constraints is applied, the type match further increases to 84.3%, and the relative speed increases to 2.8 m/s, suggesting that the severity constraint $h_{\mathrm{svt},t}$ encourages more intense safety-critical interactions. Meanwhile, the full model achieves the highest criticality score, entropy score, diversity score, RS, and composite weighted score, demonstrating the best overall balance among criticality, diversity, realism, and controllability.

**Effectiveness of Manifold Projection versus Soft Guidance.** As shown in Table 7, the soft collision guidance $J_{\mathrm{col}}$, implemented following SAFE-SIM [3], only marginally improves the collision rate from 8.2% to 10.2% compared to the baseline Flow Matching model. In contrast, the proposed GN manifold projection substantially increases the collision rate to 46.4% with fewer denoising steps. Beyond collision rate, GN projection also leads to more critical and diverse collision outcomes, achieving the highest relative speed of 2.8 m/s, and the highest diversity score of 0.97. Although its realism score is lower than the FM baseline, the overall CWS reaches 0.89, significantly outperforming both FM and soft guidance. Overall, the results suggest that hard projection onto the collision-constrained feasible set is significantly more effective than soft guidance for structured safety-critical generation. However, the final outcome still

**Table 7:** Ablation study comparing soft collision guidance and manifold projection. Starting from the baseline Flow Matching (FM) model, we evaluate the effect of adding soft collision guidance $J_{\mathrm{col}}$ and the proposed Gauss-Newton (GN) manifold projection. † denotes relative composite scores normalized within this comparison table.

| **Model** | **CR** (%) ↑ | **MS** ($m/s$) ↑ | **CS**† ↑ | **EN** ↑ | **DS**† ↑ | **OR** (%) ↓ | **RM** ↓ | **RS**† ↑ | **CWS**† ↑ |
|---|---|---|---|---|---|---|---|---|---|
| FM | 8.2 | 0.7 | 0.21 | 0.63 | 0.48 | **2.3** | **0.77** | **1.00** | 0.56 |
| FM $+J_{\mathrm{col}}$ | 10.2 | 0.8 | 0.25 | 0.70 | 0.43 | 6.2 | 0.92 | 0.60 | 0.43 |
| FM +GN | **46.4** | **2.8** | **1.00** | **0.89** | **0.97** | 5.4 | 0.78 | 0.71 | **0.89** |

depends on the road topology, traffic context, projection feasibility, and the ego planner's closed-loop response. Therefore, CCFM explicitly targets collision configurations, but a collision is realized only when the constraint can be satisfied within the evolving interactive scenario.

## 6 Conclusion

In this paper, we developed a Collision-Constrained Flow Matching (CCFM) method for controllable safety-critical scenario generation. Extensive experiments on nuScenes closed-loop simulation demonstrated that CCFM generates more safety-critical scenarios than prior methods, while achieving interpretable collision-type controllability, diverse collision outcomes, and competitive realism. In particular, our results showed that hard projection is considerably more effective than soft guidance for enforcing structured collision constraints in generation. Future work will focus on using CCFM-generated safety-critical scenarios to stress-test and improve end-to-end AV systems, as well as integrating it into sim-to-real safety-critical scenario generation for improving real-world validation.

## Acknowledgments

Qin receives funding from the Rural Safe Efficient Advanced Transportation (R-SEAT) Center, a Tier-1 University Transportation Center (UTC) funded by the United States Department of Transportation (USDOT), through agreement number 69A3552348321. The contents of this paper reflect the views of the authors. USDOT assumes no liability for the contents or use thereof.

## References

1. Caesar, H., Bankiti, V., Lang, A.H., Vora, S., Liong, V.E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., Beijbom, O.: nuScenes: A multimodal dataset for autonomous driving. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2020)
2. Caesar, H., Kabzan, J., Tan, K.S., Fong, W.K., Wolff, E., Lang, A., Fletcher, L., Beijbom, O., Omari, S.: nuplan: A closed-loop ml-based planning benchmark for autonomous vehicles. arXiv preprint arXiv:2106.11810 (2021)
3. Chang, W.J., Pittaluga, F., Tomizuka, M., Zhan, W., Chandraker, M.: SAFE-SIM: Safety-critical closed-loop traffic simulation with diffusion-controllable adversaries. In: Computer Vision – ECCV 2024. pp. 242–258 (2025)
4. Chen, K., Lei, Y., Cheng, H., Wu, H., Sun, W., Zheng, S.: FREA: Feasibility-guided generation of safety-critical scenarios with reasonable adversariality. In: 8th Annual Conference on Robot Learning (2024)
5. Chitta, K., Dauner, D., Geiger, A.: SLEDGE: Synthesizing driving environments with generative models and rule-based traffic. In: Computer Vision – ECCV 2024. pp. 57–74 (2025)

6. Dauner, D., Hallgarten, M., Geiger, A., Chitta, K.: Parting with misconceptions about learning-based vehicle motion planning. In: Conference on Robot Learning (CoRL) (2023)
7. Ding, W., Xu, C., Arief, M., Lin, H., Li, B., Zhao, D.: A survey on safety-critical driving scenario generation—a methodological perspective. IEEE Transactions on Intelligent Transportation Systems **24**(7), 6971–6988 (2023)
8. Feng, L., Li, Q., Peng, Z., Tan, S., Zhou, B.: Trafficgen: Learning to generate diverse and realistic traffic scenarios. In: 2023 IEEE International Conference on Robotics and Automation (ICRA). pp. 3567–3575. IEEE (2023)
9. Gupta, A., Johnson, J., Fei-Fei, L., Savarese, S., Alahi, A.: Social GAN: Socially acceptable trajectories with generative adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2018)
10. Hanselmann, N., Renz, K., Chitta, K., Bhattacharyya, A., Geiger, A.: KING: Generating safety-critical driving scenarios for robust imitation via kinematics gradients. In: Computer Vision – ECCV 2022. pp. 335–352 (2022)
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2016)
12. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: Advances in Neural Information Processing Systems. vol. 33, pp. 6840–6851 (2020)
13. Janner, M., Du, Y., Tenenbaum, J.B., Levine, S.: Planning with diffusion for flexible behavior synthesis. In: International Conference on Machine Learning (2022)
14. Koren, M., Alsaif, S., Lee, R., Kochenderfer, M.J.: Adaptive stress testing for autonomous vehicles. In: 2018 IEEE Intelligent Vehicles Symposium (IV). pp. 1–7 (2018)
15. Li, Z., Alim, K., Azizan, N.: HardFlow: Hard-constrained sampling for flow-matching models via trajectory optimization. arXiv preprint arXiv:2511.08425 (2025)
16. Lin, H., Huang, X., Phan, T., Hayden, D., Zhang, H., Zhao, D., Srinivasa, S., Wolff, E., Chen, H.: Causal composition diffusion model for closed-loop traffic generation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 27542–27552 (2025)
17. Lipman, Y., Chen, R.T.Q., Ben-Hamu, H., Nickel, M., Le, M.: Flow matching for generative modeling. In: The Eleventh International Conference on Learning Representations (2023)
18. Liu, H.X., Feng, S.: Curse of rarity for autonomous vehicles. Nature Communications **15**(1), 4808 (2024)
19. Liu, X., Gong, C., qiang liu: Flow straight and fast: Learning to generate and transfer data with rectified flow. In: The Eleventh International Conference on Learning Representations (2023)
20. Liu, Y., Peng, Z.M., Cui, X., Zhou, B.: Adv-BMT: Bidirectional motion transformer for safety-critical traffic scenario generation. In: Advances in Neural Information Processing Systems. vol. 38, pp. 55310–55335 (2025)
21. Nie, T., Mei, Y., Tang, Y., He, J., Sun, J., Shi, H., Ma, W., Sun, J.: Steerable adversarial scenario generation through test-time preference alignment. arXiv preprint arXiv:2509.20102 (2025)
22. Pang, B., Zhao, T., Xie, X., Wu, Y.N.: Trajectory prediction with latent belief energy-based model. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 11814–11824 (2021)

23. Peng, M., Xie, Y., Guo, X., Yao, R., Yang, H., Ma, J.: Ld-scene: Llm-guided diffusion for controllable generation of adversarial safety-critical driving scenarios. arXiv preprint arXiv:2505.11247 (2025)
24. Rempe, D., Philion, J., Guibas, L.J., Fidler, S., Litany, O.: Generating useful accident-prone driving scenarios via a learned traffic prior. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 17305–17315 (2022)
25. Salzmann, T., Ivanovic, B., Chakravarty, P., Pavone, M.: Trajectron++: Dynamically-feasible trajectory forecasting with heterogeneous data. In: Computer Vision – ECCV 2020. pp. 683–700 (2020)
26. Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B.: Score-based generative modeling through stochastic differential equations. In: International Conference on Learning Representations (2021)
27. Tan, T., Zheng, Y., Liang, R., Wang, Z., Zheng, K., Zheng, J., Li, J., Zhan, X., Liu, J.: Flow matching-based autonomous driving planning with advanced interactive behavior modeling. In: Advances in Neural Information Processing Systems. vol. 38, pp. 38310–38335 (2025)
28. Tong, A., Malkin, N., Huguet, G., Zhang, Y., Rector-Brooks, J., Fatras, K., Wolf, G., Bengio, Y.: Conditional flow matching: Simulation-free dynamic optimal transport. arXiv preprint arXiv:2302.00482 (2023)
29. Treiber, M., Hennecke, A., Helbing, D.: Congested traffic states in empirical observations and microscopic simulations. Physical review E **62**(2), 1805 (2000)
30. Utkarsh, U., Cai, P., Edelman, A., Gomez-Bombarelli, R., Rackauckas, C.V.: Physics-constrained flow matching: Sampling generative models with hard constraints. In: The Thirty-ninth Annual Conference on Neural Information Processing Systems (2025)
31. Wang, J., Pun, A., Tu, J., Manivasagam, S., Sadat, A., Casas, S., Ren, M., Urtasun, R.: AdvSim: Generating safety-critical scenarios for self-driving vehicles. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 9909–9918 (2021)
32. Xing, Z., Zhang, X., Hu, Y., Jiang, B., He, T., Zhang, Q., Long, X., Yin, W.: GoalFlow: Goal-driven flow matching for multimodal trajectories generation in end-to-end autonomous driving. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1602–1611 (2025)
33. Xu, C., Petiushko, A., Zhao, D., Li, B.: DiffScene: Diffusion-based safety-critical scenario generation for autonomous vehicles. Proceedings of the AAAI Conference on Artificial Intelligence **39**(8), 8797–8805 (2025)
34. Xu, D., Chen, Y., Ivanovic, B., Pavone, M.: Bits: Bi-level imitation for traffic simulation. In: 2023 IEEE International Conference on Robotics and Automation (ICRA). pp. 2929–2936. IEEE (2023)
35. Yin, Y., Khayatan, P., Zablocki, É., Boulch, A., Cord, M.: ReGentS: Real-world safety-critical driving scenario generation made stable. In: Computer Vision – ECCV 2024 Workshops. pp. 262–276 (2025)
36. Zhang, J., Xu, C., Li, B.: ChatScene: Knowledge-enabled safety-critical scenario generation for autonomous vehicles. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 15459–15469 (2024)
37. Zhang, L., Peng, Z., Li, Q., Zhou, B.: CAT: Closed-loop adversarial training for safe end-to-end driving. In: Proceedings of The 7th Conference on Robot Learning. Proceedings of Machine Learning Research, vol. 229, pp. 2357–2372 (2023)

38. Zhong, Z., Rempe, D., Xu, D., Chen, Y., Veer, S., Che, T., Ray, B., Pavone, M.: Guided conditional diffusion for controllable traffic simulation. In: 2023 IEEE International Conference on Robotics and Automation (ICRA). pp. 3560–3566 (2023)
39. Zhou, Z., Hu, H., Chen, X., Wang, J., Guan, N., Wu, K., Li, Y.H., Huang, Y.K., Xue, C.J.: BehaviorGPT: Smart agent simulation for autonomous driving with next-patch prediction. In: Advances in Neural Information Processing Systems. vol. 37, pp. 79597–79617 (2024)